Delegated Updates in Epistemic Graphs for Opponent Modelling

Anthony Hunter and Sylwia Polberg*

University College London, London, United Kingdom

Nico Potyka

University of Osnabrück, Germany

Abstract

In an epistemic graph, belief in arguments is represented by probability distributions. Furthermore, the influence that belief in arguments can have on the belief in other arguments is represented by constraints on the probability distributions. Different agents may choose different constraints to describe their reasoning, thus making epistemic graphs extremely flexible tools. A key application for epistemic graphs is modelling participants in persuasion dialogues, with the aim of modelling the change in beliefs as each move in the dialogue is made. This requires mechanisms for updating the model throughout the dialogue. In this paper, we introduce the class of delegated update methods, which harness existing, simpler update methods in order to produce more realistic outputs. In particular, we focus on hypothesized updates, which capture agent's reluctance or susceptibility to belief updates that can be caused by certain factors, such as time of the day, fatigue, dialogue length, and more. We provide a comprehensive range of options for modelling different kinds of agents and we explore a range of properties for categorising the options.

Keywords— probabilistic argumentation, epistemic argumentation, epistemic updates

*Corresponding author. E–mail: sylwia.polberg at gmail.com. University College London, Department of Computer Science, 66-72 Gower Street, London WC1E 6EA, United Kingdom.

The published version of this manuscript is available at https://doi.org/10.1016/j.ijar.2019.07.006.

Contents

1 Introduction

Persuasion is an activity that involves one party trying to get another party to do (or not do) some action or to believe (or not believe) certain information. It is an important, widespread, and multifaceted human facility. Consider, for example, a doctor persuading a patient to drink less, a road safety expert persuading drivers to not text while driving, or an online safety expert persuading users of social media sites against revealing too much personal information.

In computational persuasion, the role of the persuader is played by a system that can engage in dialogues with users to convince them to accept or reject a given persuasion goal. The focus is therefore on the study of formal models of dialogues involving arguments and counterarguments, persuadee models, and strategies for such systems.

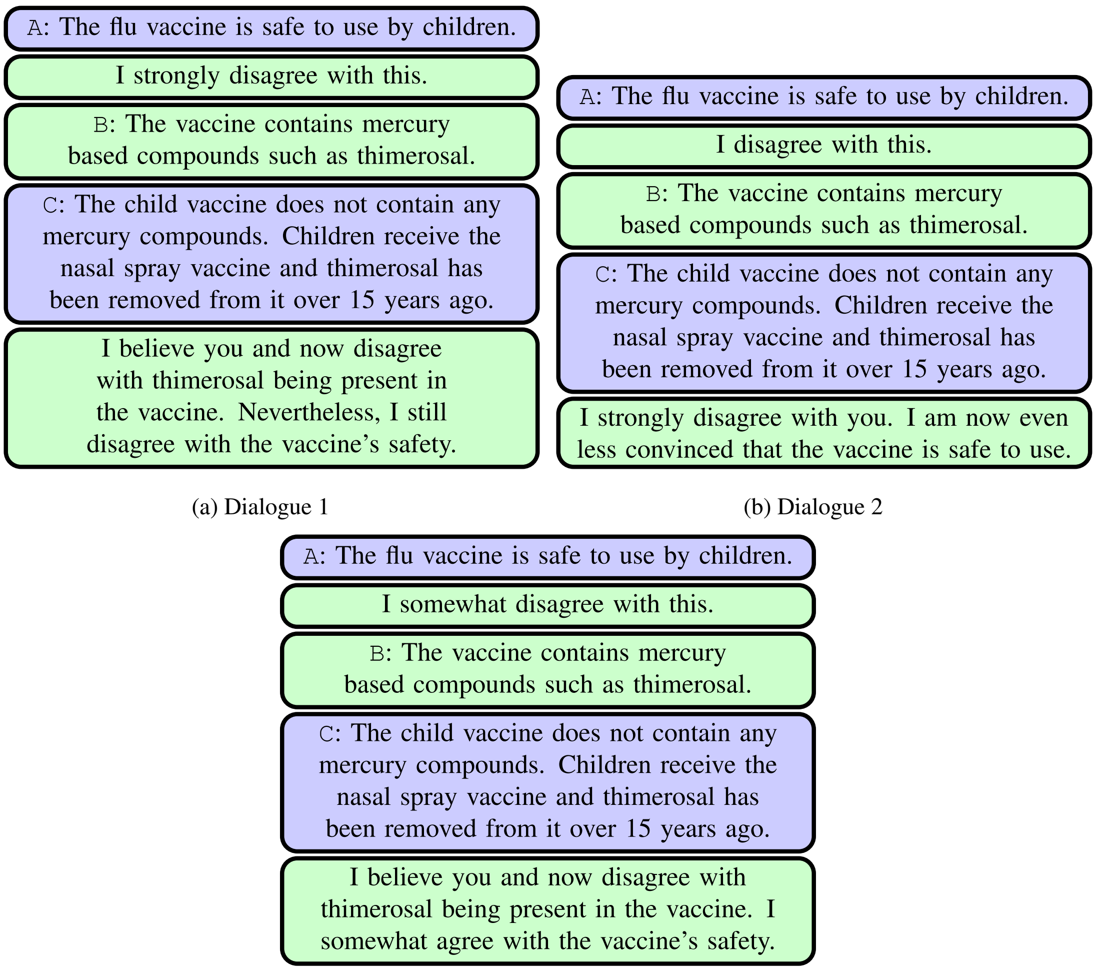
Depending on the application, our expectations of computational persuasion formalisms can change. For example, convincing the judge in a court case that the defendant is guilty follows different rules than convincing a person to stop smoking. Particularly, in the former, a given persuasion system can take a more normative approach in modelling the persuadee, as a judge is expected to follow the rule of the law and attempt to remain objective. However, in the latter, the persuadee need not be rational or obey any rules in particular. Thus, a more descriptive approach may be more suitable, as we want to find out how the agent reasons and which arguments may work best, independently of what the dialectical argumentation semantics would consider appropriate.

Human agents tend to exhibit what is often referred to as “biases” or “fallacies” [50, 52]. This can include various phenomena, such as belief bias, confirmation bias, conservatism, backfire effect, and more. Many of those can be roughly described as tendencies to interpret, prefer and select information that is aligned with one’s privately held beliefs, and to revise one’s beliefs insufficiently when faced with new information, particularly when it contradicts that person’s opinions. For example, the supporters of a given politician may downplay any evidence of the politician’s misdeeds, or a skilled artist may strongly believe they lack talent in spite of all the compliments and rewards received.

There is also only so much information that an agent can hold and process at a given point in time. Human cognitive capacity and attention span are finite, which, particularly in the case of longer dialogues, may lead to diminishing belief changes independently of how “logical” or “reasonable” a given agent is. Furthermore, standard argumentation semantics often fail to account for the fact how distance between the arguments can affect the impact they have on each other [14].

These behaviours pose challenges to computational persuasion, perhaps even more so as not taking them into account can lead to persuasion attempts causing more harm than good. It has been observed that persuasion attempts relying purely on provision of information and resolving agent’s misconceptions can have undesirable effects on the persuasion goal [49, 51, 55]. Consequently, being able to account for these non-normative behaviours is important for the efficiency and success of applications of argumentation. It is therefore worthwhile to investigate the methods for modelling situations in which the agent may become more resistant or more susceptible to belief changes and the causes for such behaviours, such as their attitudes or personality [60, 9], chosen strategies [42], or length of a discussion [71]. To illustrate some of such cases, we consider the following examples.

Example 1 (Adapted from [58])**.** Consider a doctor trying to persuade three concerned parents to give their children the flu vaccine this season. Two of the parents are anti-vaxxers who are strongly convinced of the vaccine’s harmfulness, while the third parent is mostly confused by the conflicting pieces of information they keep receiving concerning vaccines. The dialogues are presented in Figure 1, and we assume a seven–point Likert scale ⟨*Strongly Disagree, Disagree, Somewhat Disagree, Neither Agree nor Disagree, Somewhat Agree, Agree, Strongly Agree*⟩. We can observe that the arguments posited by the persuader and persuadees are relatively similar, however, the reactions as to how much the persuadees agree or disagree with the arguments differ. At the very start of the dialogue, all parents disagree with the safety of the vaccine, but to a different degree. At the end of the dialogue, one of the parents is now somewhat convinced of the safety, while the others still disagree. In addition, one of the parents agrees with the safety even less than before the dialogue due to his disbelief in the statements made by the doctor. We can observe that on the presented scale, it is not only the initial agreement levels of the parents that differ, but also the relative difference between their initial and end agreement levels is not the same. This example shows people with stronger opinions cling to them more forcefully and thus can insufficiently revise their opinions when presented with new arguments. In particular, they may still disbelieve certain information even when the original reasons for their disbelief are addressed.

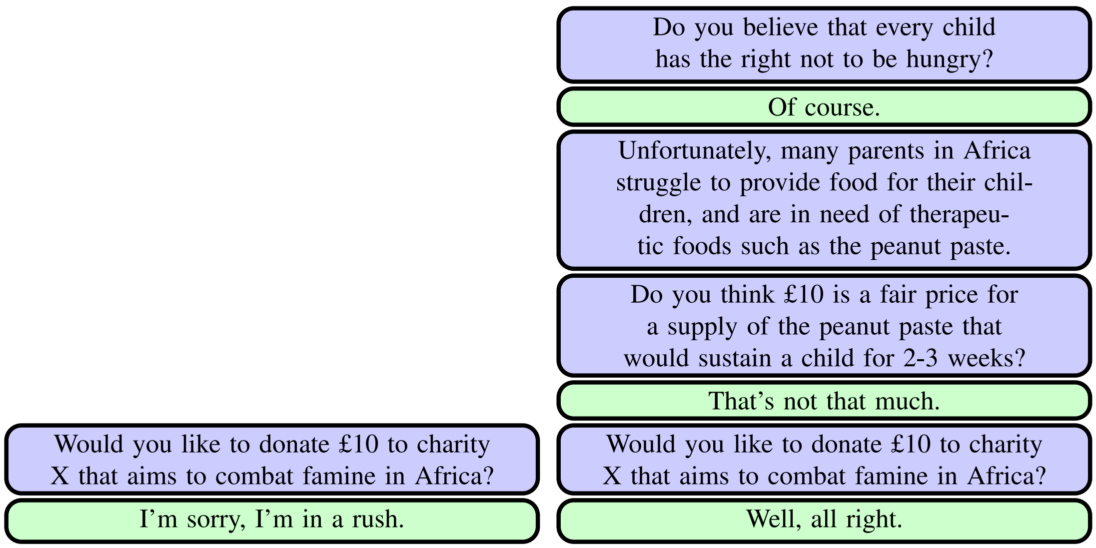


(a) Dialogue 1 (b) Dialogue 2 (c) Dialogue 3

Figure 1: Flu dialogues. The blue boxes represent statements made by the persuader and the green boxes stand for statements made by the persuadee.

Example 2 (Adapted from [24])**.** Consider the dialogues presented in Figure 2, in which a volunteer tries to stop people on the street and make them sign up and donate to charity X. We can observe that the direct approach in Dialogue 1 fails to stop the passerby and grab their attention, while appealing to a person's morality in Dialogue 2 and getting them to agree with a few statements first has increased their susceptibility and the dialogue ends in a desirable way.

Example 3. Consider a dialogue concerning student fees, an extract of which is visible in Figure 3. We can observe that the further we are from the opening statement, the more the focus of the dialogue is shifting. For example, while the argument concerning the value of menial jobs is indirectly connected to fees increasing the quality of education, it would not be entirely intuitive to demand that believing one should imply believing the other, which is what standard dialectical semantics would advise us to do. Therefore, independently of whether the persuader or the persuadee actually has the "final" argument, its impact on the arguments stated at the very beginning might be marginal. We can therefore observe that the



(a) Dialogue 1

(b) Dialogue 2

Figure 2: Charity dialogues. The blue boxes represent statements made by the persuader and the green boxes stand for statements made by the persuadee.

effect one argument has on another may diminish with distance, be it due to the focus shift as presented in this dialogue or other human factors.

Consequently, there is a need for considering argumentation formalisms that are highly flexible and allow us to model how agents reason without prejudice or enforcing any requirements as to what is "rational" or not. In [38], epistemic graphs were introduced, which through the use of epistemic constraints allowed for representing how beliefs in one argument can depend on and affect beliefs in other arguments. The way other arguments influence a given argument is expressed by epistemic constraints, and the freedom in defining these constraints allows for epistemic graphs to handle a variety of relations between the arguments. They can be specified based on a given agent's perspective and without enforcing any logic-based restrictions that may arise depending on the actual content of these arguments. Furthermore, the graphs also allow us not to specify the nature of connections between arguments at all, thus allowing for modelling of incomplete situations. These features are important in predicting and explaining how certain real-world agents reason, modelling agents which might be unable or unwilling to provide their counterarguments, and dealing with enthymeme arguments that can be decoded differently by the agents.

The versatility of epistemic graphs makes them a valuable tool for user modelling in computational persuasion, particularly in applications that call for more user–tailored approaches. Furthermore, the methods used for crowdsourcing data for probabilistic and/or bipolar argumentation that we considered in [36, 58] can be generalized for epistemic graphs and thus there is the possibility of learning epistemic graphs from participant data.

In order to provide the means for harnessing the graphs in dynamic scenarios, a method for updating user's beliefs with respect to a given epistemic graph has been presented in [37]. This method produced a new probability distribution that minimized a given notion of distance with respect to the prior distribution, and still satisfied the constraints and respected the information it was updated with. As a proof of concept of the usefulness of this approach we have considered a straightforward scenario on how user's reasoning can be modelled with the use of epistemic graphs and how this knowledge can be harnessed in a persuasion dialogue.

In this paper, we take a step further and show how epistemic graphs and update functions can be used in modelling scenarios in which, due to various possible factors, dialogue participants are more resistant or more susceptible to changing their opinion. We introduce the notion of a *delegated update method* for epistemic graphs which can harness existing, simpler methods in order to produce more advanced

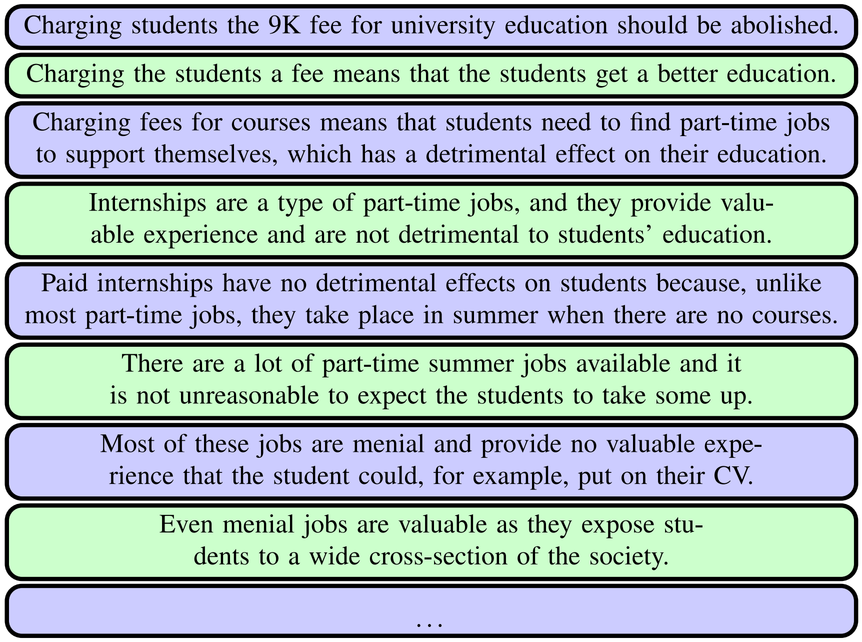


Figure 3: Student fees dialogue. The blue boxes represent statements made by the persuader and the green boxes stand for statements made by the persuadee.

solutions that can provide us with more realistic belief distributions. In particular, we focus on the notion of a *hypothesized belief update*, which through employing the effect functions for modulating the degree to which the belief in an argument should change, and hypothesized distributions that represent updated belief distribution that an agent would have in an idealized setting without external factors, aim to produce more realistic belief distributions.

This paper is organized as follows. Sections 2 to 3 recall and introduce epistemic graphs, distance measures, and updates that are the foundation for this work. Section 4 introduces the notion of a delegated update and related properties. We then propose the concept of a hypothesized belief update, analyze its components and their properties. We compare our approach to other relevant works in Section 5 and close the paper with a discussion on how we can take our proposal forward in Section 6.

2 Linear Epistemic Graphs

An argument graph specifies the arguments and relations between them. It can be seen as a directed graph in which nodes represent arguments and arcs represent relations. The nature of these relations can be denoted by edge types, labels, or additional formulae. In this section we recall epistemic graphs, which through the use of epistemic constraints tell us about the dependencies between beliefs in arguments. In order to do so, we first recall the epistemic language of the constraints.

2.1 Epistemic Language

The epistemic language introduced in [38] consists of Boolean combinations of inequalities involving statements about probabilities of formulae built out of arguments. In this work we recall a simple refinement of this approach from [37]. Throughout the section, we will assume that we have a directed graph $\mathcal{G} = (V, R)$, where V is a set of nodes (arguments) and $R \subseteq V \times V$ is a set of arcs (relations). $\mathsf{Nodes}(\mathcal{G}) = V$ denotes the nodes and $\mathsf{Arcs}(\mathcal{G}) = R$ denotes the arcs in $\mathcal{G}$.

Definition 2.1. The epistemic language based on $\mathcal{G}$ is defined as follows:

- a **term** is a Boolean combination of arguments. We use $\vee$, $\wedge$ and $\neg$ as connectives in the usual way, and can derive secondary connectives, such as implication $\rightarrow$, as usual. $\mathsf{Terms}(\mathcal{G})$ denotes all the terms that can be formed from the arguments in $\mathcal{G}$.

- a **linear operational formula** is of the form $\sum_{i=1}^{k} c_i \cdot p(\alpha_i)$ where all $\alpha_i \in \mathsf{Terms}(\mathcal{G})$ are terms, $c_i \in \mathbb{Q}$ are rational coefficients, and $p(\alpha)$ is read as "probability of α". $\mathsf{LOFormulae}(\mathcal{G})$ denotes all possible linear operational formulae of $\mathcal{G}$.

- a **linear epistemic atom** is of the form $\alpha\#x$ where $\# \in \{=, \neq, \geq, \leq, >, <\}$, $x \in \mathbb{Q}$ and $\alpha \in \mathsf{LOFormulae}(\mathcal{G})$.

- a **linear epistemic formula** is a Boolean combination of linear epistemic atoms. $\mathsf{LFormulae}(\mathcal{G})$ denotes the set of all possible linear epistemic formulae of $\mathcal{G}$.

For $\alpha \in \mathsf{Terms}(\mathcal{G})$, $\mathsf{Args}(\alpha)$ denotes the set of all arguments appearing in α and for a set of terms $\Gamma \subseteq \mathsf{Terms}(\mathcal{G})$, $\mathsf{Args}(\Gamma)$ denotes the set of all arguments appearing in Γ. Given a formula $\psi \in \mathsf{LFormulae}(\mathcal{G})$, let $\mathsf{FTerms}(\psi)$ denote the set of terms appearing in ψ and let $\mathsf{FArgs}(\psi) = \mathsf{Args}(\mathsf{FTerms}(\psi))$ be the set of arguments appearing in ψ. For a set of formulae $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$, $\mathsf{FArgs}(\Psi) = \mathsf{FArgs}(\wedge_{\psi \in \Psi} \psi)$, where $\wedge_{\psi \in \Psi} \psi$ represents the conjunction of all formulae in Ψ (or $\top$ in case Ψ is empty).

Example 4. For $\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D} \in \mathsf{Nodes}(\mathcal{G})$, $\psi : p(\mathtt{A} \wedge \mathtt{B}) - p(\mathtt{C}) - p(\mathtt{D}) > 0$ is an example of a linear epistemic formula. The terms of that formula are $\mathsf{FTerms}(\psi) = \{\mathtt{A} \wedge \mathtt{B}, \mathtt{C}, \mathtt{D}\}$, the arguments appearing in them are $\mathsf{FArgs}(\psi) = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$.

Having defined the syntax of our language, let us now focus on its semantics. For this, we require belief distributions (also called probability functions).

Definition 2.2. A **belief distribution** on the set of arguments of a graph $\mathcal{G}$ is a function $P : 2^{\mathsf{Nodes}(\mathcal{G})} \rightarrow [0,1]$ s.t. $\sum_{\Gamma \subseteq \mathsf{Nodes}(\mathcal{G})} P(\Gamma) = 1$. With $\mathsf{Dist}(\mathcal{G})$ we denote the set of all belief distributions on $\mathsf{Nodes}(\mathcal{G})$.

We assume a belief distribution conforms to the usual Kolmogorov axioms for probability theory, and that it can capture subjective probabilities.

Each $\Gamma \subseteq \mathsf{Nodes}(\mathcal{G})$ corresponds to an interpretation of arguments. We say that Γ *satisfies an argument* $\mathtt{A}$ and write $\Gamma \models \mathtt{A}$ iff $\mathtt{A} \in \Gamma$. The satisfaction relation is extended to complex terms as usual. For instance, $\Gamma \models \neg\alpha$ iff $\Gamma \not\models \alpha$ and $\Gamma \models \alpha \wedge \beta$ iff $\Gamma \models \alpha$ and $\Gamma \models \beta$. From a given belief distribution, which states probabilities of sets of arguments, we can derive probabilities of terms. They are defined as the sum of the probabilities (beliefs) of its models.

Definition 2.3. The probability of a term $\alpha \in \mathsf{Terms}(\mathcal{G})$ in a belief distribution $P \in \mathsf{Dist}(\mathcal{G})$ is denoted $P(\alpha)$ and defined as:

$$P(\alpha) = \sum_{\Gamma \subseteq \mathsf{Nodes}(\mathcal{G}) \text{ s.t. } \Gamma \models \alpha} P(\Gamma).$$

It is important to highlight that in this notation, $P(\mathtt{A})$ is the probability of a term $\mathtt{A}$, while $P(\{\mathtt{A}\})$ is the probability assigned to set $\{\mathtt{A}\}$, and these two values can be distinct.

We say that an agent believes a term α to some degree if $P(\alpha) > 0.5$, disbelieves α to some degree if $P(\alpha) < 0.5$, and neither believes nor disbelieves α when $P(\alpha) = 0.5$.

From the general set of distributions we can distinguish those that satisfy epistemic formulae.

Definition 2.4. Let φ be a linear epistemic atom $\sum_{i=1}^{k} c_i \cdot p(\alpha_i)\#b$. The **satisfying distributions** of φ are defined as $\mathsf{Sat}(\varphi) = \{P' \in \mathsf{Dist}(\mathcal{G}) \mid \sum_{i=1}^{k} c_i \cdot P'(\alpha_i) \# b\}$.

The set of satisfying distributions for a linear epistemic formula is as follows where ϕ and ψ are linear epistemic formulae:

- $\mathsf{Sat}(\phi \wedge \psi) = \mathsf{Sat}(\phi) \cap \mathsf{Sat}(\psi)$;

- $\mathsf{Sat}(\phi \vee \psi) = \mathsf{Sat}(\phi) \cup \mathsf{Sat}(\psi)$; and

- $\mathsf{Sat}(\neg\phi) = \mathsf{Sat}(\top) \setminus \mathsf{Sat}(\phi)$.

For a set of linear epistemic formulae $\Phi = \{\phi_1, \ldots, \phi_n\}$, the set of satisfying distributions is $\mathsf{Sat}(\Phi) = \mathsf{Sat}(\phi_1) \cap \ldots \cap \mathsf{Sat}(\phi_n)$.

Example 5. Consider a graph with nodes $\{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$ and the constraint $\psi : p(\mathtt{A} \wedge \mathtt{B}) - p(\mathtt{C}) - p(\mathtt{D}) > 0 \wedge p(\mathtt{D}) > 0$. A probability distribution P_1 with $P_1(\mathtt{A} \wedge \mathtt{B}) = 0.7$, $P_1(\mathtt{C}) = 0.1$ and $P_1(\mathtt{D}) = 0.1$ is in $\mathsf{Sat}(\psi)$. However, a distribution P_2 with $P_2(\mathtt{A} \wedge \mathtt{B}) = 0$ cannot satisfy ψ and so $P_2 \notin \mathsf{Sat}(\psi)$.

2.2 Epistemic Graphs

In this section we recall epistemic graphs and some of their semantics. Epistemic graphs are based on labelled graphs, which are defined as follows.

Definition 2.5. Let $\mathcal{G} = (V, R)$ be a directed graph. A **labelled graph** is a tuple $X = (\mathcal{G}, \mathcal{L})$ where $\mathcal{L} : R \to 2^{\Omega}$ is a labelling function and Ω is a set of possible labels. X is **fully labelled** iff for every $\alpha \in R$, $\mathcal{L}(\alpha) \neq \emptyset$.

The label associated with a given arc is meant to represent its general nature. We use a positive label to denote a positive influence (i.e. support), a negative label to denote a negative influence (i.e. attack), and a star label to denote an influence that is neither strictly positive nor negative (i.e. dependency). Thus, we assume that $\Omega = \{+, -, *\}$ and that the graph is fully labelled. Note, the labellings are not required for our purposes in this paper. We just present them here for completeness.

Epistemic constraints are epistemic formulae that contain at least one argument. Epistemic graphs are labelled graphs equipped with a set of such constraints.

Definition 2.6. A **linear epistemic constraint** is a linear epistemic formula $\psi \in \mathsf{LFormulae}(\mathcal{G})$ s.t. $\mathsf{FArgs}(\psi) \neq \emptyset$. An **epistemic graph** is a tuple $(\mathcal{G}, \mathcal{L}, \mathcal{C})$ where $(\mathcal{G}, \mathcal{L})$ is a labelled graph, and $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ is a set of linear epistemic constraints associated with the graph.

The semantics of epistemic graphs are given in terms of probability distributions. A range of semantics has been proposed in [38]. In the context of this work it suffices to focus on the simplest one, demanding that the constraints of the graph are satisfied.

Definition 2.7. Let $X = (\mathcal{G}, \mathcal{L}, \mathcal{C})$ be an epistemic graph. An **epistemic semantics** associates X with a set $\mathcal{R} \subseteq \mathsf{Dist}(\mathcal{G})$. A distribution $P \in \mathsf{Dist}(\mathcal{G})$ meets the **satisfaction semantics** iff $P \in \mathsf{Sat}(\mathcal{C})$.

We say that a framework is constraint consistent iff $\mathsf{Sat}(\mathcal{C}) \neq \emptyset$, i.e. the satisfaction semantics produces at least one distribution for this graph.

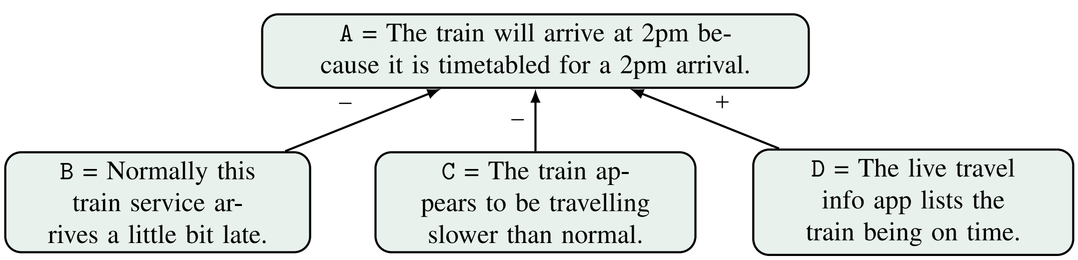


Figure 4: Labelled graph for Example 6. Edges labelled − denote attack and edges labelled + denote support.

Example 6. Consider the labelled graph in Figure 4 and imagine a passenger named Terry. We model Terry's opinions on how the arguments interact in the following manner. Since A is attacked by B and C, we want the belief in A bounded from above by the average belief in B and C. This can be formalized by the formula $p(\mathtt{A}) + 0.5 \cdot p(\mathtt{B}) + 0.5 \cdot p(\mathtt{C}) \leq 1$. Since D supports A, we also want to impose a lower bound on the belief in A. This lower bound can be decreased by the average belief in B and C. We capture this intuition with the formula $p(\mathtt{A}) + 0.5 \cdot p(\mathtt{B}) + 0.5 \cdot p(\mathtt{C}) - p(\mathtt{D}) \geq 0$. Finally, Terry is a regular on this line and believes that the train normally arrives late. We model this by the formula $p(\mathtt{B}) \geq 0.65$. Probability distributions P s.t. $P(\mathtt{A}) = 0.45$, $P(\mathtt{B}) = 0.65$, $P(\mathtt{C}) = 0.2$ and $P(\mathtt{D}) = 0.5$, and P' s.t. $P'(\mathtt{A}) = 0.425$, $P'(\mathtt{B}) = 0.65$, $P'(\mathtt{C}) = 0.5$ and $P'(\mathtt{D}) = 0.5$, are examples of satisfying distributions of this graph.

The above example also illustrates how it is the constraints rather than the graph that dictate the influence of any argument on the other arguments. However, this should not be understood as the graph or its labeling being in any way redundant. [38] introduces a series of notions for contrasting the information found in the constraints with the information found in the graph. This can be used to highlight incompleteness or imperfections of the scenarios we intend to model and can, for instance, be harnessed by argumentation-based applications, such as dialogue systems, for strategical purposes. Nevertheless, in the context of this work, we will rely primarily on the constraints in our analysis.

2.3 Normal Forms and Language Fragments

The epistemic language is quite expressive and there is a lot of freedom in how constraints can be defined. However, having a more restricted representation can offer computational benefits or simplify certain kinds of analyses. In this section, we will discuss some of the normal forms for epistemic formulae and identify fragments of the language that will be used later on.

Epistemic formulae can be represented by a variety of normal forms. For instance, by treating epistemic atoms as propositional variables, we can rewrite them as propositional conjunctive, disjunctive, and negation normal forms.

Definition 2.8. Consider a general epistemic formula ϕ. We say that ϕ is in:

- conjunctive normal form (CNF) if it is a conjunction of one or more disjunctive clauses (i.e. disjunctions of epistemic literals)
- disjunctive normal form (DNF) if it is a disjunction of one or more conjunctive clauses (i.e. conjunctions of epistemic literals)
- negation normal form (NNF) if negation occurs only directly in front of atoms and the only used operators are $\neg$, $\wedge$ and $\vee$

We can observe that in the epistemic language, any negative literal can be turned positive by changing the inequality in the atom. For example, $\neg p(\mathtt{A}) = 0$ is equivalent to $p(\mathtt{A}) \neq 0$. The ability to "pull in" the negation into the epistemic atom by changing the internal inequality and vice versa allows us to control the number of positive literals or strict inequalities in a formula. We therefore can distinguish the negation elimination normal form which contains no negations. Based on it and on DNF, we also introduce the relaxed disjunctive normal form, which will prove useful in Section 3.

Definition 2.9. Consider an epistemic formula ϕ in negation normal form. It is brought to negation eliminated normal form (NENF) by turning every negative literal into a positive one by changing the inequality in the following way:

- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) = x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) \neq x$
- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) \neq x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) = x$
- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) \geq x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) < x$
- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) \leq x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) > x$
- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) > x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) \leq x$
- $\neg \sum_{i=1}^{k} c_i \cdot p(\alpha_i) < x$ becomes $\sum_{i=1}^{k} c_i \cdot p(\alpha_i) \geq x$

Definition 2.10. Consider a general epistemic formula ϕ in DNF where all negations have been eliminated. The relaxed DNF (RDNF) of ϕ is obtained by first deleting all contradictory conjunctions in ϕ and then replacing all appearances of $<$ with $\leq$, $>$ with $\geq$ and replacing all atoms containing $\neq$ with the tautology $p(\top) = 1$ afterwards.

Each replacement results in a weaker constraints. For instance, replacing $p(\mathtt{A}) \neq 0.7$ with $p(\top) = 1$ means that $p(\mathtt{A})$ can take any value instead of taking any value except 0.7. We also observe that $p(\mathtt{A}) \neq 0.7$ is equivalent to $p(\mathtt{A}) > 0.7 \vee p(\mathtt{A}) < 0.7$. The relaxed DNF of that formula would be $p(\mathtt{A}) \geq 0.7 \vee p(\mathtt{A}) \leq 0.7$, which is a tautology and hence equivalent to $p(\top) = 1$. Consequently, directly replacing $\neq$ occurrences with tautologies can be seen as a form of a shortcut.

We will denote the set of all arguments appearing in minimal normal forms of a given kind of a given formula with $\mathsf{MinFArgs}^x(\varphi) = \bigcup\{\mathsf{FArgs}(\psi) \mid \psi \text{ is a minimal-length formula in } x \text{ s.t. } \mathsf{Sat}(\psi) = \mathsf{Sat}(\varphi)\}$, where $x \in \{\mathsf{CNF}, \mathsf{NNF}, \mathsf{DNF}, \mathsf{NENF}, \mathsf{RDNF}\}$. For a set of formulae $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$, by $\mathsf{MinFArgs}^x(\Psi)$ we understand $\mathsf{MinFArgs}^x(\wedge_{\psi\in\Psi}\psi)$.

Example 7. Consider the following epistemic formula:

$$(p(\mathtt{A}) > 0.5 \wedge \neg p(\mathtt{B}) \leq 0.5) \vee (\neg p(\mathtt{A}) > 0.5 \wedge p(\mathtt{B}) \leq 0.5) \vee (p(\mathtt{B}) \leq 0.5 \wedge \neg p(\mathtt{C}) = 0.5) \vee (\neg p(\mathtt{B}) \leq 0.5 \wedge p(\mathtt{C}) = 0.5)$$

It is in DNF and NNF, but not in CNF, NENF, or RDNF. An equivalent CNF formula would be:

$$(\neg p(\mathtt{A}) > 0.5 \vee \neg p(\mathtt{B}) \leq 0.5 \vee \neg p(\mathtt{C}) = 0.5) \wedge (p(\mathtt{A}) > 0.5 \vee p(\mathtt{B}) \leq 0.5 \vee p(\mathtt{C}) = 0.5)$$

The negation eliminated version of the original formula would be:

$$(p(\mathtt{A}) > 0.5 \wedge p(\mathtt{B}) > 0.5) \vee (p(\mathtt{A}) \leq 0.5 \wedge p(\mathtt{B}) \leq 0.5) \vee (p(\mathtt{B}) \leq 0.5 \wedge p(\mathtt{C}) \neq 0.5) \vee (p(\mathtt{B}) > 0.5 \wedge p(\mathtt{C}) = 0.5)$$

Bringing it to relaxed disjunctive normal form would yield:

$$(p(\mathtt{A}) \geq 0.5 \wedge p(\mathtt{B}) \geq 0.5) \vee (p(\mathtt{A}) \leq 0.5 \wedge p(\mathtt{B}) \leq 0.5) \vee (p(\mathtt{B}) \leq 0.5 \wedge p(\top) = 1) \vee (p(\mathtt{B}) \geq 0.5 \wedge p(\mathtt{C}) = 0.5)$$

We observe that the original formula is DNF, but is not minimal. Examples of its minimal DNFs include

$$(p(\mathtt{A}) > 0.5 \wedge \neg p(\mathtt{C}) = 0.5) \vee (\neg p(\mathtt{A}) > 0.5 \wedge p(\mathtt{B}) \leq 0.5) \vee (\neg p(\mathtt{B}) \leq 0.5 \wedge p(\mathtt{C}) = 0.5)$$

and

$$(p(\mathtt{A}) > 0.5 \wedge \neg p(\mathtt{B}) \leq 0.5) \vee (\neg p(\mathtt{A}) > 0.5 \wedge p(\mathtt{C}) = 0.5) \vee (\neg p(\mathtt{C}) = 0.5 \wedge p(\mathtt{B}) \leq 0.5)$$

Thus, the set $\mathsf{MinFArgs}^{\mathsf{DNF}}$ of the original formula is $\{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$, and is the same as the FArgs of this formula.

Certain problems related to epistemic graphs can be more closely related to linear optimization problems. However, they can be compounded by the presence of strict inequalities, which leads us to distinguishing the following fragment of our language.

Definition 2.11 (Non-strict Epistemic Atom)**.** A **non-strict epistemic atom** is a linear epistemic atom $\sum_{i=1}^{n} c_i \cdot p(\alpha_i) \# b$ where $\# \in \{\leq, =, \geq\}$.

Definition 2.12 (Non-strict Epistemic Formulae)**.** The set of *non-strict epistemic formulae* is the closure of non-strict epistemic atoms under the logical connectives $\wedge$ and $\vee$.

We observe that the negation is not included in the above fragment due to the relations between inequalities highlighted in the negation eliminated normal form.

3 Updating Probability Distributions with Epistemic Constraints

Whether we consider monological or dialogical argumentation, receiving new information calls for an update in our beliefs. For instance, during a persuasion dialogue, the involved parties may update their models of each others' beliefs as the discussion progresses in order to make better choices of moves to make. It is therefore important to develop epistemic update functions, that take our current epistemic state (i.e. current belief distribution), epistemic constraints representing existing information and epistemic formulae representing new information, and return a set of candidates for the next epistemic state. We adopt the principles for update functions that have been set out in [37]:

Definition 3.1 (Update Function)**.** An **update function** for a graph $\mathcal{G}$ is a function $\mathsf{U} : \mathsf{Dist}(\mathcal{G}) \times 2^{\mathsf{LFormulae}(\mathcal{G})} \times 2^{\mathsf{LFormulae}(\mathcal{G})} \to 2^{\mathsf{Dist}(\mathcal{G})}$. With $\mathcal{U}(\mathcal{G})$ we denote the universe of update functions for $\mathcal{G}$.

In order to guarantee meaningful update functions, we consider the properties previously set out in [37][1]

[1]Due to a modified syntax of the update functions the *Representation Invariance* property now comes in three types.

where P is the current belief distribution, $\mathcal{C}$ is a set of constraints, and Ψ is the set of update formulae:

- **Uniqueness:** $|\mathsf{U}(P, \mathcal{C}, \Psi)| \leq 1$.
- **Completeness:** If $\mathsf{Sat}(\mathcal{C} \cup \Psi) \neq \emptyset$, then $|\mathsf{U}(P, \mathcal{C}, \Psi)| \geq 1$.
- **Epistemic Consistency:** $\mathsf{U}(P, \mathcal{C}, \Psi) \subseteq \mathsf{Sat}(\mathcal{C})$.
- **Success:** $\mathsf{U}(P, \mathcal{C}, \Psi) \subseteq \mathsf{Sat}(\Psi)$.
- **Tautology:** If $P \in \mathsf{Sat}(\mathcal{C})$ and $\mathsf{Sat}(\Psi) = \mathsf{Sat}(\top)$, then $\mathsf{U}(P, \mathcal{C}, \Psi) = \{P\}$.
- **Contradiction:** If $\mathsf{Sat}(\mathcal{C} \cup \Psi) = \emptyset$, then $\mathsf{U}(P, \mathcal{C}, \Psi) = \emptyset$.
- **Update Representation Invariance:** If Ψ_1, Ψ_2 are equivalent, i.e., $\mathsf{Sat}(\Psi_1) = \mathsf{Sat}(\Psi_2)$, then $\mathsf{U}(P, \mathcal{C}, \Psi_1) = \mathsf{U}(P, \mathcal{C}, \Psi_2)$.
- **Constraint Representation Invariance:** If $\mathcal{C}_1$ and $\mathcal{C}_2$ are equivalent, i.e., $\mathsf{Sat}(\mathcal{C}_1) = \mathsf{Sat}(\mathcal{C}_2)$, then $\mathsf{U}(P, \mathcal{C}_1, \Psi) = \mathsf{U}(P, \mathcal{C}_2, \Psi)$.
- **Complete Representation Invariance:** *Update* and *Constrain Representation Invariance*
- **Idempotence:** If $\mathsf{U}(P, \mathcal{C}, \Psi) = \{P^*\}$, then $\mathsf{U}(P^*, \mathcal{C}, \Psi) = \{P^*\}$.

Uniqueness guarantees that there is at most one candidate for the next epistemic state, and *Completeness* states that there is at least one if the update is consistent. If both properties are satisfied, the next epistemic state is uniquely defined whenever the update is consistent. *Epistemic Consistency* demands that the constraints in our graph are maintained and *Success* demands that the next state satisfies the beliefs that we updated with. *Tautology* states that updating with a tautological set of constraints should not change a distribution otherwise meeting the constraints and *Contradiction* states that an inconsistent update should yield the empty set. *Representation Invariance* guarantees that changing the syntactic representation of updates and/or constraints does not change the outcome of the update. Finally, *Idempotence* demands that a repeated update does not change beliefs.

In addition to the aforementioned properties, we can also consider the following optional postulates:

- **Indiscrimination**: $\mathsf{U}(P, \mathcal{C}, \Psi) = \mathsf{U}(P, \mathcal{C} \cup \Psi, \emptyset) = \mathsf{U}(P, \emptyset, \mathcal{C} \cup \Psi)$
- **Conservatism:** If $P \in \mathsf{Sat}(\mathcal{C} \cup \Psi)$ then $\mathsf{U}(P, \mathcal{C}, \Psi) = \{P\}$

Indiscrimination states that new and existing information is treated in the same manner, which may or may not be a desirable property depending on the setting and on whether the constraints and updating formulae are jointly consistent. *Conservatism* states that performing an update with a piece of information already consistent with our constraints should lead to no change.

3.1 Epistemic Distance Measures

The ability to compare the answers produced by various abstract argumentation approaches, be they extensions, labelings or distributions, is valuable in many applications. This can be done, for example, through subset relations or information orderings. In the context of this work, we will consider measuring distances between probability distributions, which will be particularly relevant in guiding the distribution update process. Our distance functions may not necessarily be metrics, but we assume that they satisfy the properties explained below [37].

Continuity and convexity are defined as usual [69]. Formally, a function $f : X \to Y$ is called *continuous* at a point $c \in X$ if $\lim_{x \to c} f(x) = f(c)$. Intuitively, this means that if x is close to c, $f(x)$ must be close to $f(c)$ (the function graph does not make any jumps at c). f is called continuous if it is continuous on the whole domain X. f is called *convex* if $f(\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1 - \lambda) \cdot f(x_2)$ for all $\lambda \in [0, 1]$. Intuitively, convex functions have a non-decreasing curvature (the function graph of f has a convex shape). In particular, there can be no non-global local minima. Therefore, convex functions are particularly interesting for minimization purposes.

Definition 3.2 (Epistemic Distance Function)**.** An **epistemic distance function** is a function $d : \mathsf{Dist}(\mathcal{G}) \times \mathsf{Dist}(\mathcal{G}) \to \mathbb{R}$ that satisfies

1. Positive Definiteness: $d(P, P') \geq 0$ and $d(P, P') = 0$ iff $P = P'$.
2. Continuity: d is continuous in the second argument.
3. Strict Convexity: d is strictly convex in the second argument.

Intuitively, the continuity guarantees that probability distributions that assign similar probabilities to subsets of arguments have a low distance value. Strict convexity guarantees that there often is a unique solution and no non-global local minima when we minimize the distance. The popular examples of epistemic distance functions are the Least Squares distance and the Kullback-Leibler divergence.

- **Least Squares Distance:**
  $d_2(P, P') = \sum_{E \subseteq \mathsf{Nodes}(\mathcal{G})} (P(E) - P'(E))^2$.
- **Kullback-Leibler divergence:**
  $d_{KL}(P, P') = \sum_{E \subseteq \mathsf{Nodes}(\mathcal{G})} P(E) \cdot \log \frac{P(E)}{P'(E)}$.

The KL-divergence is an example of an epistemic distance function that is not a metric (it does not satisfy symmetry and the triangle-inequality). However, it still has some intuitive geometric properties and is a popular measure to compare probability distributions [27]. In the following sections we will focus on update functions that minimize some epistemic distance function to a prior belief state. In the definition of optimization problems, "$\min f(x)$" denotes the minimum function value that f takes over the feasible region and "$\arg\min f(x)$" denotes the set of points where f takes this value. For instance, for a function $f(x) = x^2 + 0.5$, $\min f(x) = 0.5$ and $\arg\min f(x) = \{0\}$.

3.2 Distance-minimizing Update Functions

We will now recall the distance-based update functions, i.e. functions that aim to minimize a certain notion of distance between the old and the updated distribution(s) [37].

Definition 3.3 (Distance-minimizing Update Function)**.** Given some epistemic distance function d, the **distance-minimizing update function** with respect to d is defined as follows where P is the current belief distribution, $\mathcal{C}$ is a set of constraints, and Ψ is the set of update formulae.

$$\mathsf{U}_d(P, \mathcal{C}, \Psi) = \arg \min_{P' \in \mathsf{Sat}(\mathcal{C} \cup \Psi)} d(P, P')$$

for all finite sets of constraints $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$, finite sets of formulae $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ and probability distributions $P \in \mathsf{Dist}(\mathcal{G})$.

We observe that the distance minimizing update function is not well-defined for all kinds of formulae. The analysis carried out in [37] considers appropriate fragments of the epistemic language and studies the conditions under which at least one (and ideally, exactly one) updated distribution is produced, as well as methods one may use in fragments that pose difficulties. The results can be summarized as follows.

Proposition 3.4. *[Extended version of [37]] Every distance-minimizing update functions satisfies* Epistemic Consistency*,* Success*,* Tautology*,* Contradiction*,* Update Representation Invariance*,* Constraint Representation Invariance*,* Complete Representation Invariance*,* Idempotence*,* Indiscrimination*, and* Conservatism*.*

Theorem 3.5. *[Taken from [37]] In the fragment of non-strict epistemic axioms, every distance-minimizing update function satisfies* Uniqueness *and* Completeness*. In the fragment of non-strict epistemic formulae, every distance-minimizing update function satisfies* Completeness*.* Uniqueness *can be violated, but $|\mathsf{U}_d(P, \Psi)| \leq k^*$, where k^* is the smallest number of conjunctions in all DNF representations of $\mathcal{C} \cup \Psi$.*

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_1	0.45	0.65	0.2	0.5
$P_2 = U_{d_2}(P_1, \Psi_1)$	0.67	0.65	0	1
$P_3 = U_{d_2}(P_2, \Psi_2)$	0.175	0.65	1	0.5

Table 1: Returning to Example 8, beliefs in arguments before and after updating the epistemic state with new knowledge $\Psi_1 = \{p(\mathtt{D}) = 1, p(\mathtt{C}) = 0\}$ and $\Psi_2 = \{p(\mathtt{C}) = 1\}$.

Proposition 3.6. *[Taken from [37]] Suppose that $\mathcal{C} \cup \Psi$ is from the fragment of non-strict epistemic formulae. Let $\bigvee_{i=1}^{k} \Gamma_i$ denote a DNF representation of $\mathcal{C} \cup \Psi$. Let P_i^* be the unique solution of $\mathsf{U}_d(P, \Gamma_i)$ and let $m^* = \min\{d(P, P_i^*) \mid 1 \le i \le k\}$. Then $\mathsf{U}_d(P, \Psi) = \{P_i^* \mid d(P, P_i^*) = m^*\}$ and $|\mathsf{U}_d(P, \Psi)| \le k$.*

Proposition 3.7. *[Taken from [37]] Let $\mathcal{C} \cup \Psi$ contain arbitrary epistemic formulae. Let $\bigvee_{i=1}^{k} \Gamma_i$ denote a DNF of $\mathcal{C} \cup \Psi$ where negations have been eliminated. Let $\bigvee_{i=1}^{k'} \Gamma'_i$ denote the corresponding relaxed DNF. For $i = 1, \ldots, k'$, let P_i^* be the unique solution of the optimization problem*

$$\min_{P' \in \mathsf{Sat}(\Gamma'_i)} d(P, P')$$

corresponding to the i-th conjunction Γ'_i in the relaxed DNF. Let $m^* = \min\{d(P, P_i^*) \mid 1 \le i \le k'\}$ be the minimum distance obtained among all P_i^*. Then $\mathsf{U}_d(P, \Psi)$ equals

$$\{P_i^* \mid 1 \le i \le k',\ d(P, P_i^*) = m^* \text{ and } P_i^* \in \mathsf{Sat}(\Gamma_i)\}.$$

Example 8. Let us continue Example 6. The first row in Table 1 shows the beliefs in arguments for our initial epistemic state P_1. Suppose that Terry can access the info app and learns that the train is indeed on time. Furthermore, he also strongly believes that the train is travelling at its usual speed. We can model this by an update with $\Psi_1 = \{p(\mathtt{D}) = 1, p(\mathtt{C}) = 0\}$. The second row in Table 1 shows the beliefs in arguments after updating P_1 to $P_2 = U_{d_2}(P_1, \Psi_1)$. Assume that a little bit later, the train has to slow down because of bad weather conditions. The third row in Table 1 shows the beliefs in arguments after updating P_2 with $\Psi_2 = \{p(\mathtt{C}) = 1\}$ to $P_3 = U_{d_2}(P_2, \Psi_2)$. Note that the fact that the train slows down not only decreases Terry's belief in A, but also indirectly leads to a decrease in D, which can be seen as Terry no longer being sure that the app showing the train on time indeed means the train will arrive on time.

Note, the above example also illustrates how an update changing the belief in one argument can cause a change in belief in other arguments that are indirectly connected to (or even disconnected from) the other arguments. There are three reasons for this situation.

First, we observe that the graph structure is not used to determine the influence of an argument on the other arguments in the distance minimizing updates as defined above. Only the constraints are used to determine the influence of an argument on the other arguments. Among others, Section 4 (and in particular, Section 4.5.2) will provide examples of how the graph structure can be accounted for in an update. In the next section we also propose a modification of the distance minimizing approach that limits the amount of arguments that undergo belief change, though following a different principle.

Second, we argue that it is not only directly connected arguments that should affect each other. One of the core concepts of bipolar argumentation is that the interactions between arguments, particularly mixtures of attack and support, can give rise to new, indirect relations which do not always follow the directionality of the edges in the graph (see [20, 57] for further analysis). Further analysis of relations between arguments and how constraints can expose more interactions can be found in [38].

Last, but not least, one of the purposes of epistemic graphs is to model incomplete or imperfect scenarios. Consequently, it is possible that there is a disparity between the constraints and the graph. One of the examples when such situations may arise is in dialogical argumentation, where as a starting point we take an expert constructed argumentation graph. During the dialogue it can happen that the constraints describing an agent's reasoning pattern point to a different graph structure due to the fact that they process information differently due to, for instance, various types of bias. Updates performed using such constraints can cause changes that we could see as clashing with the changes we would expect from the graph.

We thus observe that there can be good reasons for allowing (or not) changes in belief in a given argument and that various methods have their merits. Our proposal in Section 4 will offer more ways of controlling how beliefs of an agent are modified in contrast to the standard update methods.

3.3 Atomic Distance Minimization

Minimizing the distance between the sets of arguments to which probabilities are assigned is not the same as minimizing the differences between the probabilities assigned to arguments themselves. Consequently, [39, 37] has also considered atomic distance, i.e. ones that measure distance in terms of probability mass assigned to arguments rather than sets. Although they do not satisfy epistemic distance requirements, they can be used as a preprocessing step for other update functions in order to filter the candidate set.

Definition 3.8 (Weighted Atomic Distance)**.** Let $S \subseteq \mathsf{Nodes}(\mathcal{G})$ be a set of arguments and let $w : S \to \mathbb{R}_0^+$ be a weight function assigning a non-negative weight to each argument in S. The **weighted atomic distance** with respect to w is defined as $d^w_{\mathrm{At}}(P, P') = \sum_{\mathtt{A} \in S} w(\mathtt{A}) \cdot |P(\mathtt{A}) - P'(\mathtt{A})|$.

Lemma 3.9. *[Taken from [37]] If Γ is a conjunction of non-strict epistemic atoms, then the solutions of*

$$\arg \min_{P' \in \mathsf{Sat}(\Gamma)} d^w_{\mathrm{At}}(P, P')$$

correspond to the solution of the linear program

$$\begin{aligned} \min \quad & \sum_{\mathtt{A} \in S} \left(\delta^+_{\mathtt{A}} + \delta^-_{\mathtt{A}}\right) \qquad (1) \\ \textit{s.t.} \quad & P' \in \mathsf{Sat}(\Gamma) \\ & w(\mathtt{A}) \cdot \left(P(\mathtt{A}) - P'(\mathtt{A})\right) = \delta^+_{\mathtt{A}} - \delta^-_{\mathtt{A}} \quad \textit{for all } \mathtt{A} \in S \\ & \delta^+_{\mathtt{A}}, \delta^-_{\mathtt{A}} \in \mathbb{Q}_0^+ \quad \textit{for all } \mathtt{A} \in S \end{aligned}$$

and form a compact and convex set.

Definition 3.10 (Atomic Distance-minimizing Update Function)**.** Given some epistemic distance function d and some weight function $w : S \to \mathbb{R}_0^+$ over a subset of arguments $S \subseteq \mathsf{Nodes}(\mathcal{G})$, the **atomic distance-minimizing update function** $\mathsf{U}^w_d(P, \mathcal{C}, \Psi)$ with respect to d is defined by the set of minimal solutions of the optimization problem

$$\begin{aligned} \min \quad & d(P, P') \\ \textit{s.t.} \quad & P' \in \mathsf{Sat}(\mathcal{C} \cup \Psi) \\ & \sum_{\mathtt{A} \in S} \left(\delta^+_{\mathtt{A}} + \delta^-_{\mathtt{A}}\right) = m^* \\ & w(\mathtt{A}) \cdot \left(P(\mathtt{A}) - P'(\mathtt{A})\right) = \delta^+_{\mathtt{A}} - \delta^-_{\mathtt{A}} \quad \textit{for all } \mathtt{A} \in S \\ & \delta^+_{\mathtt{A}}, \delta^-_{\mathtt{A}} \in \mathbb{Q}_0^+ \quad \textit{for all } \mathtt{A} \in S, \end{aligned}$$

where m^* is the minimum of all minima of (1) computed for all conjunctions of a relaxed DNF of $\mathcal{C} \cup \Psi$.

With U^1_d we will denote a distance minimizing function in which the weight function assigns value 1 to every argument. This means that all atomic values are treated equally.

Theorem 3.11. *[Extended version of [37]] Every atomic distance-minimizing update functions satisfies* Epistemic Consistency*,* Success*,* Tautology*,* Contradiction*,* Update Representation Invariance*,* Constraint Representation Invariance*,* Complete Representation Invariance*,* Idempotence*,* Indiscrimination*, and* Conservatism*. In the fragment of non-strict epistemic formulae,* Completeness *is satisfied as well and $\mathsf{U}^w_d(P, \Psi)$ is guaranteed to be finite. In the fragment of non-strict epistemic atoms,* Uniqueness *is also satisfied.*

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_1	0.45	0.65	0.2	0.5
$P_2 = U_{d_2}(P_1, \Psi_1)$	0.5	0.65	0.36	1
$P'_2 = U^1_{d_2}(P_1, \Psi_1)$	0.575	0.65	0.2	1

Table 2: Beliefs in arguments before and after updating the epistemic state with new knowledge $\Psi_1 = \{p(\mathtt{D}) = 1\}$ in Example 9.

Example 9. Let us continue Example 8 and let us consider a different update to the initial epistemic state. Suppose another traveller tells Terry that the info app states that the train is indeed on time and that Terry trusts his statement. We describe this with $\Psi_1 = \{p(\mathtt{D}) = 1\}$. By performing a least-squares update, we obtain the distribution P_2 and its beliefs in arguments are visible in Table 2. Although this distribution satisfies our constraints and minimizes the overall change in probability mass, the increase in belief in C might not be considered justified.

Let us now carry out this update using the atomic distance-minimizing update approach s.t. for every argument X, $w(\mathtt{X}) = 1$ (i.e. all atomic distances are treated equally). By performing an update with the formula $\Psi_1 = \{p(\mathtt{D}) = 1\}$, we obtain the probability function P'_2 visible in the third row of Table 2. The new probability assignments to arguments are now more intuitive than in the case of P_2.

While definitely interesting, the work on distance minimizing update functions only gives a limited way of constructing a new distribution. As we discussed in Section 1, constructing better models for human reasoning requires richer update formalisms, and this is what we will consider in the next section.

4 Delegated Updates of Probability Distributions

In this section, we present our solution to the requirements raised in Section 1. For this, we introduce the notion of delegated updates. These take the context of the update into account (as we explain in Section 4.1), and they delegate part of the process of finding an updated belief distribution to other functions (as we explain in Section 4.2)

4.1 Contextual Information in Updates

Epistemic constraints can be used in order to model the reasoning patterns of the users, and such user models can be later harnessed by automated persuasion systems, applications supporting decision making, agent simulations and more. However, while particular patterns may hold for a given user in general, their actual reactions may be subject to factors associated with the actual application. For instance, in a system that engages in a discussion with a participant, we can consider the length of the dialogue (e.g. the user may become exhausted), how much the system is trusted, emotional reactions to the arguments etc. Elements unrelated to the system itself also have some impact. For example, we could expect that a person who had a bad day so far will be less malleable than a well–rested participant in a good mood. Access to such additional information could be harnessed by the system in order to, for example, decide whether to engage in a dialogue and to better estimate the actual belief changes that a given dialogue move might lead to, thus affecting the overall strategy.

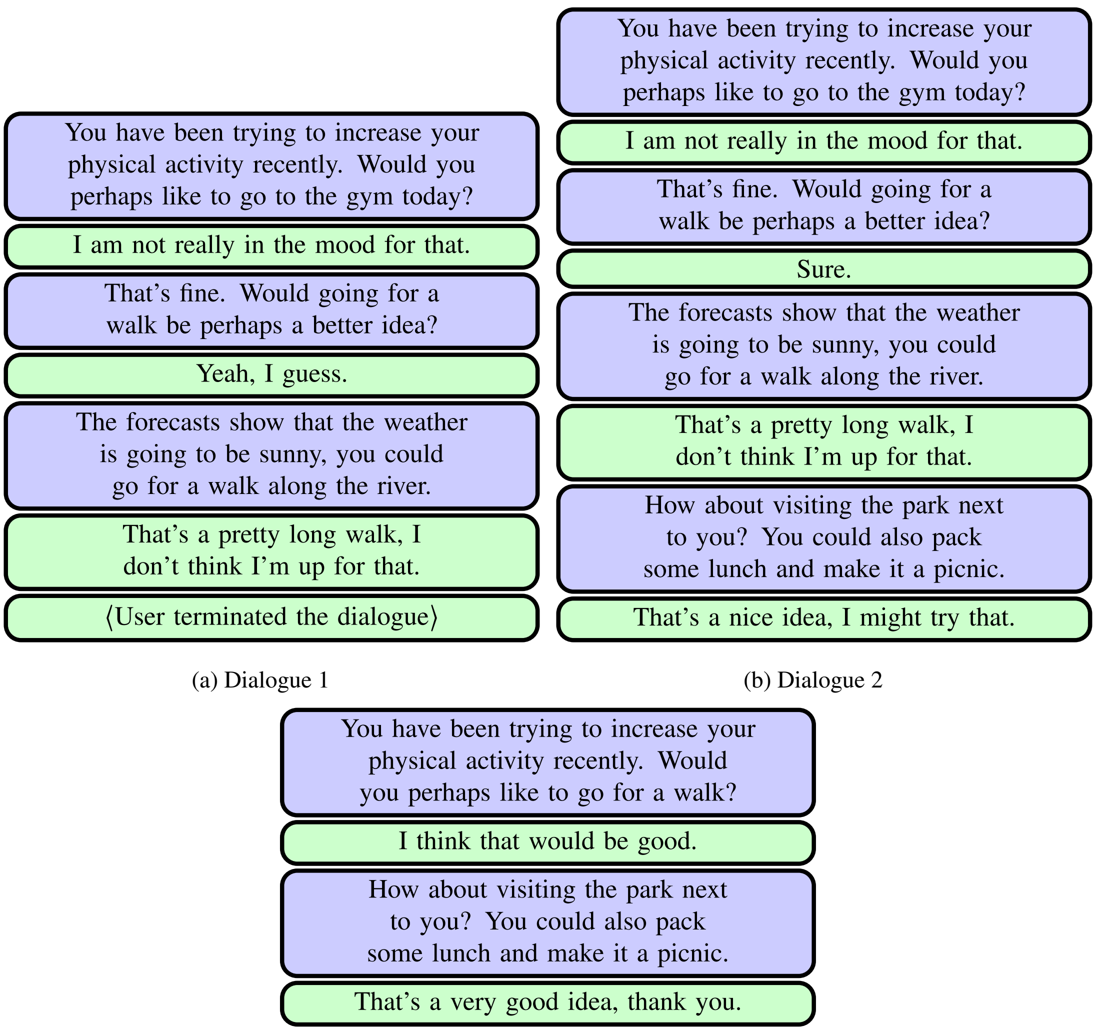
Example 10. Consider three possible dialogues presented in Figure 5 between a user and three different systems. In each case, the systems have access to weather forecasts.

Dialogue 1 Here, the system tries to engage in a dialogue with the user on Thursday afternoon. The user is tired and somewhat impatient, thus ending the dialogue before an agreeable solution is found.

Dialogue 2 In this case we have a system that is aware of time and date and knows that during the weekends (outside of evenings), the user does not mind engaging in longer dialogues as much as on evenings or other days of the week. The system therefore tries to engage in a dialogue on Saturday afternoon and a satisfactory solution is found.

Dialogue 3 This system can measure the fatigue and emotional state of the user through video analysis. The system realizes that the user is tired and that presenting more straining options is going to be met with resistance or rejection. Thus, the system goes straight for the easier options, which is met with more enthusiastic agreement than in Dialogue 2.

These are the kinds of dialogues that we might want to support in a computational persuasion system for behaviour change. In the modelling of the argumentation, we want to represent the arguments as arising in the dialogues, and we do not want to introduce arguments that cannot be exchanged. Therefore, we do not want to introduce auxiliary arguments for the purposes of directly or indirectly defeating arguments that are disbelieved by the agent, particularly that their effect can change throughout the dialogue and thus necessitate updates to the graph as well as to the beliefs.



(a) Dialogue 1 (b) Dialogue 2

(c) Dialogue 3

Figure 5: Physical activity dialogues. The blue boxes represent statements made by the persuader and the green boxes stand for statements made by the persuadee.

These additional requirements or pieces of information affect the way distributions are updated throughout the discussion and a possible way to handle them is to consider an appropriate reasoning layer on top of the methodology. In the next section we will therefore focus on the notion of a delegated update method, which can be seen as an update method that, through harnessing possible additional information and other existing update methods, aims to produce a more refined output that would better reflect the actual behaviour of the person being modelled. In other words, by using delegated updates we are aiming for more accurate prediction of the reasoning of a given individual.

However, In order to provide a definition of a delegated update, we need a way to describe the extra input for the update that would take account of the internal features of the agent being modelled (e.g. their emotional state, their attitude to new information, personality, and more) and the external features concerning the agent being modelled (e.g. time of the day, date, or weather), as seen in Example 10. The range of possible features is enormous and the exact choice of features to consider in practice depends on the application. Nevertheless, it should be possible to represent them with some combinations of mathematical objects (i.e. objects with a mathematical definition), such as functions, tuples, formulae, and more. For instance, if we want to model an agent who resists updates, then we may use a function that decreases the magnitude of the updates based on the information about the update. We will refer to these mathematical objects as update parameters.

Notation 4.1. With Λ we denote a set of mathematical objects which we refer to as **update parameters**.

Update parameters can depend on the application, environment, sensory modules of the agent, and more. For instance, we can consider statements about the weather $weather(sunny)$, about the length of the dialogue combined with the fatigue level of the user $[length(1), fatigue(0.7)]$, or the epistemic formulae corresponding to the exchanges made during a dialogue, such as $\varphi : p(\mathtt{A}) = 1 \to p(\mathtt{B}) = 0$. In the rest of this paper, we will consider some specific mathematical objects and explain how they can reflect various internal and external features of agents being modelled.

4.2 Towards Delegated Update

We now present our proposal for belief updates that meets the requirements we have set out so far in the paper. In particular, in order to improve the produced results, it takes into account the context of the update into account by using update parameters. We call it the delegated update method, as it can delegate to multiple existing standard update functions in order to perform its tasks. We start by considering a simple notation to represent sequences of elements of a given kind and length.

Definition 4.2. For a natural number $j \geq 1$ and a set of elements T, $\mathsf{Sequence}(T, j)$ denotes the collection of all j-tuples over T.

So, for example, given a set $T = \{t_1, t_2, t_3\}$, $\mathsf{Sequence}(T, 2) = \{\langle t_1, t_1\rangle, \langle t_1, t_2\rangle, \langle t_1, t_3\rangle, \langle t_2, t_1\rangle, \langle t_2, t_2\rangle, \langle t_2, t_3\rangle, \langle t_3, t_1\rangle, \langle t_3, t_2\rangle, \langle t_3, t_3\rangle\}$.

We can now propose the following definition of a delegated update function, which given a probability distribution, existing constraints, an updating formula, and possible additional information, can adjust how updates should be performed and delegates their actual execution to the underlying sequence of simpler update methods.

Definition 4.3. Let $\mathcal{G}$ be a graph, $\mathcal{U}(\mathcal{G})$ the universe of update methods for $\mathcal{G}$. Let Λ be the set of update parameters. A **delegated update function** of degree $n \geq 1$ for $\mathcal{G}$ is a function $\mathsf{DU} : \mathsf{Dist}(\mathcal{G}) \times \mathsf{Sequence}(\mathcal{U}(\mathcal{G}), n) \times 2^{\mathsf{LFormulae}(\mathcal{G})} \times 2^{\mathsf{LFormulae}(\mathcal{G})} \times \Lambda \to 2^{\mathsf{Dist}(\mathcal{G})}$.

For particular instances of DU, in order to improve readability instead of writing $\langle \mathsf{U} \rangle$ for a single element sequence, we just write U.

Note, we do not consider in this paper how we would select a set of update parameters to model a specific agent. However, we envisage that based on the information known about the agent (e.g. moves made in a dialogue, answers to questions, behaviour of similar agents in the past, and more) or the environment (e.g. day of the week, time, or weather), an appropriate set of update parameters can be selected. In future work, we intend to explore how such selection methods can be efficiently constructed from the available data.

In order to illustrate why we need more sophisticated update methods than proposed in our previous works, we examine the following scenario.

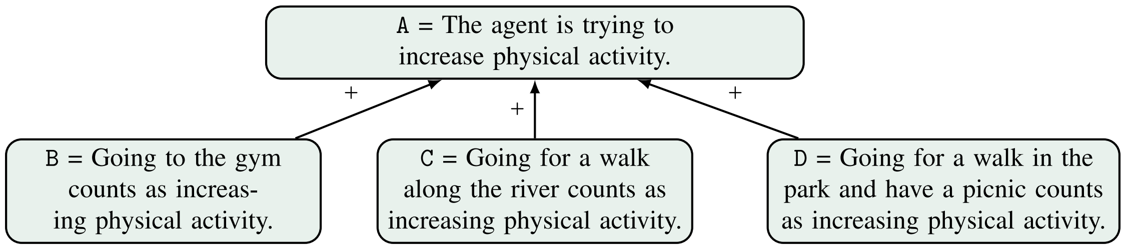


Figure 6: Labelled graph for Example 11. Edges labelled + denote support.

Example 11. Let us consider a possible formalization of the dialogues presented in Example 10. Assume that the underlying labelled graph is as presented in Figure 6. We consider the following constraints $\mathcal{C}$ in order to form an epistemic graph, which simply state that A is believed at least as well as any of its supporters and that believing A implies believing at least one of the supporters.

- $\psi_1 : p(\mathtt{A}) - p(\mathtt{B}) \geq 0$
- $\psi_2 : p(\mathtt{A}) - p(\mathtt{C}) \geq 0$
- $\psi_3 : p(\mathtt{A}) - p(\mathtt{D}) \geq 0$
- $\psi_4 : p(\mathtt{A}) > 0.5 \rightarrow (p(\mathtt{B}) \geq 0.51 \vee p(\mathtt{C}) \geq 0.51 \vee p(\mathtt{D}) \geq 0.51)$

Let us assume that we want to perform updates based on the system's moves. The beliefs in arguments in the resulting distribution can be then compared with user's responses in order to verify the accuracy of the approach. The epistemic constraints corresponding to the moves could, for example, be transformed into the following updating formulae: $\Phi_1 = \{p(\mathtt{B}) = 1\}$, $\Phi_2 = \{p(\mathtt{C}) \geq 0.51 \vee p(\mathtt{D}) \geq 0.51\}$, $\Phi_3 = \{p(\mathtt{C}) = 1\}$ and $\Phi_4 = \{p(\mathtt{D}) = 1\}$. They can be seen as the system expecting the user to trust and agree with the made proposals based on prior agreement to increase physical activity.

We can now consider starting with a uniform distribution and performing standard updates using the above formulae. We observe that Dialogue 1 would corresponds to updates with Φ_1, Φ_2 and Φ_3. Dialogue 2 extends this with an update with Φ_4. Dialogue 3 corresponds to updates with Φ_2 and Φ_4. Consecutive applications of these updates lead to types of distributions listed in Table 3. We observe that in some cases, there can be more than one outcome, and we investigate possible distribution patterns in parallel.

We observe that both types of final distributions for Dialogue 1 (P_3 and P'_3) are highly inaccurate. By this we understand that if we contrasted the updated distributions with the actual user distribution, there would be a high discrepancy in the beliefs in arguments. Based on the reactions of the user as expressed in the dialogue, we can expect both B and C to be disbelieved, which is the opposite of what the final distributions would tell us. A similar pattern occurs in Dialogue 2 (P_4 and P'_4), despite the fact that the dialogue ends successfully. Independently, based on the user's answers, we can also observe that the degrees of (dis)belief are more polarized than expected. Only in the case of Dialogue 3, the obtained final distributions somehow reflect the user's opinions on the arguments that were used in the exchange.

In order to improve this, we could consider examples of delegated updates that, based on additional information, decrease or increase the beliefs in arguments in the updating formulae. For example, we can imagine a delegated update method Y that, given the length of the dialogue, state of the user, time, and weather, would transform the set Φ_1 into a different one and delegate the execution to the standard update function U_{d_2}:

$$Y(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Phi_1, \lambda_1) = \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \{p(\mathtt{B}) = 0.2\})$$
$$\text{where } \lambda_1 = [length(1), fatigue(0.7), date(16:00 - 04AUG2018), weather(sunny)]$$

Instead of P_2, this would produce the distribution P' s.t. $P'(\mathtt{A}) = P'(\mathtt{C}) = P'(\mathtt{D}) = 0.5$ and $P'(\mathtt{B}) = 0.2$, which can be seen as better reflecting the reluctance of the user expressed in Dialogues 1 and 2, and makes it more justified to avoid performing this dialogue move.

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_0	0.5	0.5	0.5	0.5
$P_1 \in \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Phi_1)$	1	1	0.5	0.5
$P_2 \in \mathsf{U}^1_{d_2}(P_1, \mathcal{C}, \Phi_2)$	1	1	0.51	0.5
$P'_2 \in \mathsf{U}^1_{d_2}(P_1, \mathcal{C}, \Phi_2)$	1	1	0.5	0.51
$P_3 \in \mathsf{U}^1_{d_2}(P_2, \mathcal{C}, \Phi_3)$	1	1	1	0.5
$P'_3 \in \mathsf{U}^1_{d_2}(P'_2, \mathcal{C}, \Phi_3)$	1	1	1	0.51
$P_4 \in \mathsf{U}^1_{d_2}(P_3, \mathcal{C}, \Phi_4)$	1	1	1	1
$P'_4 \in \mathsf{U}^1_{d_2}(P'_3, \mathcal{C}, \Phi_4)$	1	1	1	1
$P^*_2 \in \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Phi_2)$	0.51	0.5	0.51	0.5
$P^{**}_2 \in \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Phi_2)$	0.51	0.5	0.5	0.51
$P^*_4 \in \mathsf{U}^1_{d_2}(P^*_2, \mathcal{C}, \Phi_4)$	1	0.5	0.51	1
$P^{**}_4 \in \mathsf{U}^1_{d_2}(P^{**}_2, \mathcal{C}, \Phi_4)$	1	0.5	0.5	1

Table 3: Consecutive applications of updates from Example 11. In some cases, the updates may produce more than one distribution. The table presents the resulting beliefs in arguments in these distributions and not their full descriptions.

Please note that at this point we do not fully define how λ_1 would give us this update - this is the subject of the rest of this section. We also do not consider how update parameters such as fatigue can be monitored and revised throughout the dialogue - we discuss this further in Section 6.

We observe that the accuracy issue of the distributions in the above example could be addressed by performing updates after both system and user moves. While it would adjust the inaccurate distributions after a move is made, it would not stop them from being flawed in the first place. This means that the analysis of effects of possible dialogues moves by the system would still be inaccurate and could lead the system to the select sub-optimal dialogue moves during the discussion.

4.3 Properties of Delegated Updates

Section 3 has introduced a number of postulates for epistemic update methods and the existing definitions that can be easily adapted to delegated updates. However, while the idea behind the postulates does not change in this new setting, the view on their desirability does. Although introduced for good reasons and valuable from the analytical point of view, not all of these properties are realistic in the case of delegated updates. Delegated updates aim to provide more life-like solutions, even if it means not meeting certain theoretical requirements. In this section we will discuss which of the original postulates may or may not be applicable to delegated updates and why.

Let us start with the properties that delegated updates should meet. Out of the previously proposed postulates, *Uniqueness* and *Completeness* are perhaps the most desirable, as they guarantee unique answers whenever possible.

- **Uniqueness:** $|\mathsf{DU}(P, \langle \mathsf{U}_1, \ldots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda)| \leq 1$.
- **Completeness:** If $\mathsf{Sat}(\mathcal{C} \cup \Psi) \neq \emptyset$, then $|\mathsf{DU}(P, \langle \mathsf{U}_1, \ldots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda)| \geq 1$.

Unfortunately, depending on the situation, the remaining properties may not always be realistic. To start with, while the *Contradiction* property is reasonable for update methods, it should not be enforced in delegated updates. A delegated method is meant to improve upon the results of the standard methods, which may include providing an answer when no good one exists. Attempting to resolve the inconsistencies can lead to favouring existing formulae over the new ones and vice versa, thus easily violating the *Indiscrimination* property as well.

- **Contradiction:** If $\mathsf{Sat}(\mathcal{C} \cup \Psi) = \varnothing$ then $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \varnothing$.
- **Indiscrimination**: $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C} \cup \Psi, \varnothing, \lambda) = \mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \varnothing, \mathcal{C} \cup \Psi, \lambda)$

Other properties that may be easily violated are the *Epistemic Consistency* and *Success*. The cognitive capacity of human agents is not infinite and their resistance or susceptibility to changes can mean that some beliefs in arguments may be revised to a lesser or greater extent than expected. For example, in a long discussion involving a chain of hundred arguments (and counterarguments, countercounterarguments ...), it is unrealistic to expect that the change in the latest argument propagates back completely to the first argument stated. Thus, while constraints can accurately describe how a given person sees relations between arguments in general, the effects of a dynamic setting can lead to some of them being omitted or ignored.

- **Epistemic Consistency:** $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) \subseteq \mathsf{Sat}(\mathcal{C})$.
- **Success:** $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) \subseteq \mathsf{Sat}(\Psi)$.

The same human-centered perspective as with the previous properties can make the *Tautology* and *Idempotence* less important. Consider a dialogue in which a persuadee tries to convince the persuader to have a flu shot and after a while starts answering the queries of the persuadee with tautologies resembling excluded middle. So, for example, a query "*Is it true that the vaccine is safe to use?*" is answered with "*It is..or it isn't.*". After a series of such uninformative answers it is possible that the persuadee will actually develop doubts and will be less likely to have the flu shot, which can violate both *Tautology* and *Idempotence* properties. In a similar fashion, repetitive presentation of information that an agent already agrees with may lead to some form of reinforcement of their beliefs, which contradicts the *Conservatism* property.

- **Tautology:** If $P \in \mathsf{Sat}(\mathcal{C})$ and $\mathsf{Sat}(\Psi) = \mathsf{Sat}(\top)$ then $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P\}$.
- **Idempotence:** If $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P^*\}$ then $\mathsf{DU}(P^*, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P^*\}$.
- **Conservatism:** If $P \in \mathsf{Sat}(\mathcal{C} \cup \Psi)$ then $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P\}$

Finally, while in a purely logical setting, the *Representation Invariance* properties are important, in real life situations it may not necessarily be the case that two logically equivalent, but syntactically different formulae will be seen in the same way by an agent. For example, consider a dialogue in which the option is to present the persuadee with two statements, one corresponding to a formula φ, and one to ψ, s.t. both are in CNF, $\mathsf{Sat}(\varphi) = \mathsf{Sat}(\psi)$, and φ is built out of a single clause and ψ out of 50. It is highly unlikely that a human agent will process these formulae in the same manner. Thus, while we consider *Constraint Representation Invariance* to be still desirable, *Update* (and thus *Complete*) *Representation Invariance* might not be entirely realistic.

- **Update Representation Invariance:** If Ψ_1, Ψ_2 are equivalent, i.e., $\mathsf{Sat}(\Psi_1) = \mathsf{Sat}(\Psi_2)$, then $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi_1, \lambda) = \mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi_2, \lambda)$.
- **Constraint Representation Invariance:** If $\mathcal{C}_1$ and $\mathcal{C}_2$ are equivalent, i.e., $\mathsf{Sat}(\Psi_1) = \mathsf{Sat}(\Psi_2)$ and $\mathsf{Sat}(\mathcal{C}_1) = \mathsf{Sat}(\mathcal{C}_2)$, then $\mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}_1, \Psi, \lambda) = \mathsf{DU}(P, \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle, \mathcal{C}_2, \Psi, \lambda)$.
- **Complete Representation Invariance:** *Update* and *Constrain Representation Invariance*

In addition to these postulates, we can also consider the property inheriting, which demands that properties satisfied by the update methods used in the delegated method carry over to the delegated method itself. This can be easily formulated in the following manner, where V is any of the previously considered properties:

- **Inherited V**: If for every $\mathsf{U}_i \in \langle \mathsf{U}_1, \dots, \mathsf{U}_n \rangle$, U_i satisfies the standard version of property V, then DU satisfies the delegated version of V

So, for example, if we wanted DU to satisfy *Inherited Completeness*, then we would demand that if every U_i satisfies *Completeness*, then DU satisfies *Completeness* as well.

4.4 Hypothesized Belief Update

There are various ways one could define the delegated updates. In the context of this work, of particular interest is modelling the behaviour of an agent believing an argument to a greater or lesser degree than expected due to some factors that have lead to an increased susceptibility or resistance to opinion change. We thus focus on delegated updates that alter the changes in beliefs in arguments, by which we understand the differences in the belief in an argument before and after an update is performed. We note that doing so we focus on reasoning in terms of probabilities of atomic arguments, which to humans can be more natural than shifting masses assigned to sets of arguments. In this work, we will introduce updates that speak in terms of adjusted and hypothesized belief, and start by explaining some of the core notions.

In order to perform an update, delegated or otherwise, we need at least an **update function**, a **(current) probability distribution**, a **set of constraints** we want to obey, and an **updating formula** with respect to which we want to perform the update.

Carrying out a (standard) update with the above information would, hopefully, yield a probability distribution. Such a distribution would not be taking into account any special circumstances that may have affected the agent's reasoning and as such would not necessarily be realistic. It is an example of a **hypothesized distribution**, by which we understand an updated belief distribution that an agent would have if, for example, no external factors were present, if the agent's cognitive capacity was not limited, if they were not resisting changes, and more. It can be seen as a form of somewhat idealized, "what if" reasoning. In the context of this work we will focus on only one approach, however, other possible ways of obtaining such distributions will be discussed in Section 6.

Since, the belief we would have in an argument in a hypothesized distribution might not necessarily be realistic, it needs to be modified in order to account for the nature of the agent being modelled and the wider context of the updates. The information of how the adjustment should be performed is stored in an **effect function**, which tell us whether the difference in belief in an argument between the current and hypothesized distribution needs to be decreased or increased. Such a function can be a constant factor, or depend on the distance between arguments, or more, and we will discuss various options for such functions in Section 4.5. By applying the effect function to the difference in beliefs between the current and hypothesized distributions we finally obtain the **adjusted belief**, which is an updated belief that considers the desired additional factors.

Example 12. Consider a simple graph with arguments $\mathtt{A}$ and $\mathtt{B}$ and a constraint $p(\mathtt{B}) > 0.5 \rightarrow p(\mathtt{A}) \leq 0.5$. Assume that an agent has a probability distribution P_1 s.t. $P_1(\mathtt{A}) = 0.9$ and $P_1(\mathtt{B}) = 0.1$ and that we want to perform an update with the formula $p(\mathtt{B}) = 0.7$. Normally, this would yield a probability distribution P_2 s.t. $P_2(\mathtt{A}) = 0.5$ and $P_2(\mathtt{B}) = 0.7$. However, assume that the agent is in fact resisting changes and their beliefs are updated only half as much as expected.

We therefore take P_2 as the hypothesized distribution. We also create an effect function that assigns the value of 0.5 to every argument in order to model the assumption that the difference in beliefs between hypothesized and current distribution is cut in half. Thus, we are looking for a function P_3 s.t. :

- $P_3(\mathtt{A}) - P_1(\mathtt{A}) = 0.5 \cdot (P_2(\mathtt{A}) - P_1(\mathtt{A}))$
- $P_3(\mathtt{B}) - P_1(\mathtt{B}) = 0.5 \cdot (P_2(\mathtt{B}) - P_1(\mathtt{B}))$

This means that we want $P_3(\mathtt{A}) = 0.7$ and $P_3(\mathtt{B}) = 0.4$. Given that this only tells us the beliefs in arguments, not in the sets of arguments themselves, an appropriate standard update function can be used to find the complete description of P_3.

Let us now define things more formally. The hypothesized belief update function is a delegated update function of the first degree (i.e. it delegates the execution of the updates only to a single update method). Its first input elements are the update function to delegate to, the distribution to update, the set of constraints that we need to obey, and the updating formulae. However, in order to work, it also needs the hypothesized distribution, the effect function, and input for that function. With respect to the definition of the delegated update function, they will form the elements of the update parameters set Λ.

We start by considering effect functions, which tells us how to adjust the changes in beliefs based on the current situation. Following the intuitions from Example 11, this can encompass dialogue length, date, weather, and possibly more. In other words, the effect function will be a part of the hypothesized belief that will process most of the additional information that we have discussed at the start of Section 4. We thus assume a set Θ of suitable update parameters.

Definition 4.4. Let Θ be a set of update parameters. An **effect function** is a function $\delta : \Theta \to (\mathsf{Nodes}(\mathcal{G}) \to \mathbb{Q})$ that produces an assignment of rational numbers to arguments.

We observe that the assignments of effects to arguments are not limited to positive values or values from $[0, 1]$. An agent that is susceptible to change will change his beliefs more than expected and thus be described with factors greater than 1. An agent with firmly established beliefs may exhibit backfire effect, i.e. when challenged by contradictory evidence, their beliefs can get stronger. Hence, an expected increase (decrease) in beliefs may turn out to be the opposite, and a factor smaller than 0 can be used to describe this. In practice, argumentation has the potential of damaging our position or escalating disagreement [25, 53, 54], which needs to be accounted for. A more detailed analysis of effect functions will be provided in Section 4.5.

Let us now define the hypothesized belief update. For every argument in the graph, this method produces an atomic equality constraint, where the new belief that an argument is meant to be assigned is obtained by modifying the change in beliefs between the current and hypothesized belief function. Finding an appropriate updated distribution for these constraints is then delegated to the assumed update function.

Definition 4.5. Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints, U an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. The **hypothesized belief update** function is defined as

$$\mathsf{HBU}(P, \mathsf{U}, \mathcal{C}, \Psi, [H, \delta, \theta]) = \mathsf{U}(P, \varnothing, \mathcal{C}')$$

where $\mathcal{C}' = \{\psi_{\mathtt{A}} \mid \mathtt{A} \in \mathsf{Nodes}(\mathcal{G})\}$ and

$$\psi_{\mathtt{A}} : \begin{cases} p(\mathtt{A}) = 1 & P(\mathtt{A}) + \delta(\theta)(\mathtt{A}) \cdot (H(\mathtt{A}) - P(\mathtt{A})) > 1 \\ p(\mathtt{A}) = 0 & P(\mathtt{A}) + \delta(\theta)(\mathtt{A}) \cdot (H(\mathtt{A}) - P(\mathtt{A})) < 0 \\ p(\mathtt{A}) = x \text{ where } x = P(\mathtt{A}) + \delta(\theta)(\mathtt{A}) \cdot (H(\mathtt{A}) - P(\mathtt{A})) & \text{otherwise} \end{cases}$$

We observe that with respect to the general delegated update definition, $[H, \delta, \theta]$ are the elements of Λ. We distinguish them with square brackets rather than round ones for readability purposes.

We can also observe that technically speaking, Ψ is not directly used by HBU. However, it will be used in the production of the hypothesized distribution, and while indirectly, still has a significant impact on the produced distribution. In particular, throughout this paper we will assume that $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ (i.e. the hypothesized distribution is a distribution we would obtain by applying the update in a standard manner) and that an appropriate function can always be selected. The creation of more advanced approaches as well as methods for selecting good hypothesized distributions, especially when dealing with updates that produce no distributions at all, is a research problem in its own right. We discuss it in Section 6, however, it is primarily left for future work.

Let us now consider an example of the hypothesized belief update.

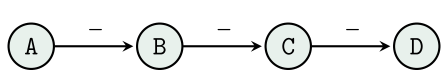


Figure 7: An argument graph

Example 13. Consider the graph from Figure 7. We assume the following set of linear constraints:

$$\mathcal{C} = \{\varphi_1 : p(\mathtt{D}) + p(\mathtt{C}) = 1, \varphi_2 : p(\mathtt{C}) + p(\mathtt{B}) = 1, \varphi_3 : p(\mathtt{B}) + p(\mathtt{A}) = 1\}$$

We now start with a uniform distribution P_0 which assigns the same probability to all sets of arguments. As a result, every argument is believed with the degree of 0.5. Consider updating the distribution using one of the standard update methods U_{d_2} and the update formulae $\Psi_1 = \{p(\mathtt{A}) = 0.9\}$. This yields the distribution P_1, in which arguments are believed as stated in Table 4. Updating P_1 with another singleton set of formulae $\Psi_2 = \{p(\mathtt{B}) = 0.3\}$ gives us the distribution P_2.

Let us now assume that a certain resistance is taking place and let us take a function $\delta_c^{0.8}$ s.t. given a set of formulae Ψ, $\delta_c(\Psi) = g$ where $g(\mathtt{A}) = 0.8$ for $\mathtt{A} \notin \mathsf{FArgs}(\Psi)$ and $g(\mathtt{A}) = 1$ otherwise. In other words, the person being modelled may agree with the information being presented, but (perhaps subconsciously) somehow resists with fully propagating the changes caused by this information. By considering the updating formulae set Ψ_1 and taking P_1 as the hypothesized distribution, the adjusted belief constraints are as follows:

$$\mathcal{C}_1' = \{p(\mathtt{A}) = 0.9, p(\mathtt{B}) = 0.18, p(\mathtt{C}) = 0.82, p(\mathtt{D}) = 0.18\}$$

We can now use HBU in order to find a distribution P_1' meeting the above properties and being a result of an update from P_0.

Let us now consider updating P_1' with Ψ_2 in the standard manner. This produces distribution H_2, and the degrees to which the arguments believed in it are listed in Table 4. We observe that while they are identical to P_2, the underlying distribution may not be[2]. Updating P_1' with Ψ_2 using HBU, where H_2 is taken as the hypothesized distribution, would make us search for a distribution P_2' satisfying the following constraints:

$$\mathcal{C}_2' = \{p(\mathtt{A}) = 0.74, p(\mathtt{B}) = 0.3, p(\mathtt{C}) = 0.724, p(\mathtt{D}) = 0.276\}$$

Finally, let us consider how a different effect function would affect the results. In particular, let us consider the function $dis_j^x(\Phi) = g$ s.t. for $\mathtt{F} \in \mathsf{Nodes}(\mathcal{G})$, $g(\mathtt{F}) = 0$ if $\mathsf{id}(\mathtt{A}, \Phi) = \infty$ and $g(\mathtt{F}) = x^v$ where $v = \max(\{\mathsf{id}(\mathtt{A}, \Phi) - j, 0\})$ otherwise. By $\mathsf{id}(\mathtt{A}, \Phi)$ we will understand the length of the shortest possible path between $\mathtt{A}$ and any of the arguments in appearing in the formulae in Φ (we will discuss this more in Section 4.5.2). For the purpose of this example, we set $x = 0.8$ and $j = 1$. It captures the idea of decreasing the belief change in arguments with respect to the distance from the arguments participating in an update. Let us start again with P_0. Updating it with Ψ_1 with P_1 as the hypothesized distribution means we need to satisfy the following formulae, and once more an appropriate distribution P_1'' can be found with HBU:

$$\mathcal{C}_3' = \{p(\mathtt{A}) = 0.9, p(\mathtt{B}) = 0.1, p(\mathtt{C}) = 0.82, p(\mathtt{D}) = 0.244\}$$

By updating P_1'' with Ψ_2 in a standard manner we obtain the distribution H_2', which has resulting beliefs in arguments similar to P_2, but a different underlying structure. Let us now consider updating P_1'' with Ψ_2 where H_2 is taken as the hypothesized distribution. This means that the resulting distribution P_2'' would need to satisfy the following:

$$\mathcal{C}_4' = \{p(\mathtt{A}) = 0.7, p(\mathtt{B}) = 0.3, p(\mathtt{C}) = 0.7, p(\mathtt{D}) = 0.289\}$$

The resulting distributions are visible in Table 4, which allows us to compare the behaviours of different effect functions.

Example 14. Let us extend Examples 10 and 11. Again, we assume that the underlying labelled graph is as presented in Figure 6 and that the set of constraints $\mathcal{C}$ is as follows:

- $\psi_1 : p(\mathtt{A}) - p(\mathtt{B}) \geq 0$
- $\psi_2 : p(\mathtt{A}) - p(\mathtt{C}) \geq 0$
- $\psi_3 : p(\mathtt{A}) - p(\mathtt{D}) \geq 0$
- $\psi_4 : p(\mathtt{A}) > 0.5 \to (p(\mathtt{B}) \geq 0.51 \vee p(\mathtt{C}) \geq 0.51 \vee p(\mathtt{D}) \geq 0.51)$

We recall we were performing updates with the sets $\Phi_1 = \{p(\mathtt{B}) = 1\}$, $\Phi_2 = \{p(\mathtt{C}) \geq 0.51 \vee p(\mathtt{D}) \geq 0.51\}$, $\Phi_3 = \{p(\mathtt{C}) = 1\}$ and $\Phi_4 = \{p(\mathtt{D}) = 1\}$. We focus on Dialogue 3 and associated updates with Φ_2 and Φ_4. Let us now consider an effect function domain $\Theta = \{(\Psi, length(L), fatigue(F), date(D), weather(W)) \mid \Psi \subseteq \mathsf{LFormulae}(\mathcal{G}), L \geq 0, F \in [0,1], D \text{ is a date}, W \in \{\text{sunny, cloudy, rainy, windy, snowy}\}\}$. L denotes the length of the dialogue, and F denotes the degree to which the agent appears fatigued (the higher the

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_0	0.5	0.5	0.5	0.5
$\{P_1\} = \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Psi_1)$	0.9	0.1	0.9	0.1
$\{P_2\} = \mathsf{U}^1_{d_2}(P_1, \mathcal{C}, \Psi_2)$	0.7	0.3	0.7	0.3
$\{P'_1\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_1, [P_1, \delta^{0.8}_c, \Psi_1)])$	0.9	0.18	0.82	0.18
$\{H_2\} = \mathsf{U}^1_{d_2}(P'_1, \mathcal{C}, \Psi_2)$	0.7	0.3	0.7	0.3
$\{P'_2\} = \mathsf{HBU}(P'_1, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_2, [H_2, \delta^{0.8}_c, \Psi_2])$	0.74	0.3	0.724	0.276
$\{P''_1\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_1, [P_1, dis^{0.8}_1, \Psi_1])$	0.9	0.1	0.82	0.244
$\{H'_2\} = \mathsf{U}^1_{d_2}(P''_1, \mathcal{C}, \Psi_2)$	0.7	0.3	0.7	0.3
$\{P''_2\} = \mathsf{HBU}(P''_1, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_2, [H'_2, dis^{0.8}_1, \Psi_2])$	0.7	0.3	0.7	0.289

Table 4: Distributions with and without diminishing effect functions applied from Example 13. We use the set of formulae $\Psi_1 = \{p(\mathtt{A}) = 0.9\}$ and $\Psi_2 = \{p(\mathtt{B}) = 0.3\}$, and effect functions $\delta^{0.8}_c$ and $dis^{0.8}_1$ as defined in the example.

					δ			
Ψ	length	fatigue	date	weather	A	B	C	D
Φ_1	1	0.7	16:00-04AUG2018	sunny	-0.6	-0.6	-0.6	-0.6
Φ_1	2	0.7	16:00-04AUG2018	sunny	-0.8	-0.8	-0.8	-0.8
Φ_1	3	0.7	16:00-04AUG2018	sunny	-1	-1	-1	-1
Φ_1	4	0.7	16:00-04AUG2018	sunny	-1	-1	-1	-1
Φ_2	1	0.7	16:00-04AUG2018	sunny	1.5	1.5	1.5	1.5
Φ_2	2	0.7	16:00-04AUG2018	sunny	1.2	1.2	1.2	1.2
Φ_2	3	0.7	16:00-04AUG2018	sunny	1	1	1	1
Φ_2	4	0.7	16:00-04AUG2018	sunny	0	0	0	0
Φ_3	1	0.7	16:00-04AUG2018	sunny	0.5	0.5	0.5	0.5
Φ_3	2	0.7	16:00-04AUG2018	sunny	0	0	0	0
Φ_3	3	0.7	16:00-04AUG2018	sunny	-0.2	-0.2	-0.2	-0.2
Φ_3	4	0.7	16:00-04AUG2018	sunny	-0.4	-0.4	-0.4	-0.4
Φ_4	1	0.7	16:00-04AUG2018	sunny	1	1	1	1
Φ_4	2	0.7	16:00-04AUG2018	sunny	0.9	0.9	0.9	0.9
Φ_4	3	0.7	16:00-04AUG2018	sunny	0.8	0.8	0.8	0.8
Φ_4	4	0.7	16:00-04AUG2018	sunny	0.7	0.7	0.7	0.7

Table 5: Excerpt of the assignments of the effect function δ from Example 10.

degree, the more exhausted the agent is). Let δ be an effect function that, for the chosen input, produces values as in Table 5.

The effect function describes the information that the system in Dialogue 3 has access to and we can assume that the effects assigned to arguments could have been learned from previous interaction of the system with the user and other similar participants. The system, based on the state of the user, can therefore see that stating Φ_1 will actually backfire and not only lead to a lack of positive change, it will make the situation worse. The system can therefore choose to go with updates with formulae towards which the agent

[2] Recall that two different distributions may still have the same resulting beliefs in arguments. For instance, a distribution assigning probability 0.5 to sets $\{\mathtt{A}\}$ and $\{\mathtt{B}\}$ will have the same beliefs in A and B as the one assigning 0.5 to $\{\mathtt{A}, \mathtt{B}\}$ and 0.5 to $\emptyset$.

has a more positive attitude, which in this case is Φ_2 followed by Φ_4. This would give us the distributions listed in Table 6.

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_0	0.5	0.5	0.5	0.5
$H_2^* \in \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Phi_2)$	0.51	0.5	0.51	0.5
$P_2^* \in \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Phi_2, [H_2^*, \delta, \theta_1])$	0.515	0.5	0.515	0.5
$H_4^* \in \mathsf{U}^1_{d_2}(P_2^*, \mathcal{C}, \Phi_4)$	1	0.5	0.515	1
$P_4^* \in \mathsf{HBU}(P_2^*, \mathsf{U}^1_{d_2}, \mathcal{C}, \Phi_4, [H_4^*, \delta, \theta_2])$	0.952	0.5	0.515	0.950

Table 6: Consecutive applications of updates from Example 14. The table presents the resulting beliefs in arguments in selected distributions and not their full descriptions. We use the effect function δ given in Table 5 and the inputs $\theta_1 = (\Phi_2, length(1), fatigue(0.7), date(16:00-04AUG2018), weather(sunny))$ and $\theta_2 = (\Phi_4, length(2), fatigue(0.7), date(16:00-04AUG2018), weather(sunny))$.

4.5 Effect Functions

There is quite a lot of freedom as to how δ effect functions used by HBU's can be defined. Their specification is crucial in obtaining new distributions through the hypothesized belief update and various options can lead to various results. In the context of this work, we will be particularly focused on the following three, non-exhaustive categories of functions: coefficient-based, distance-based, and progress-based types.

4.5.1 Coefficient-Based Effect Functions

The coefficient-based functions are the simplest possible effect functions and are meant to assign the same effect values to arguments, with some possible exceptions. For instance, they can be used for expressing that the general reluctance or susceptibility of an agent depends on external factors such as general mood, weather, and more, that are unrelated to the actual content of the arguments. The regular method produces a function that assigns the same effect to every argument. The other methods allow us to leave some arguments out. The selective approach prevents adjusting effects from applying to the arguments appearing in the input formulae, and the redundancy-free selective approach refines it by removing the protection from arguments that are logically redundant in the formulae. By redundant we understand them as not appearing in any of the minimal CNF representations of the formulae. For the functions we introduce, we set the effect function domain Θ to be a collection of sets of epistemic formulae, however, we note that this does not have to be the case for any other notions that can be developed in the future.

Definition 4.6. Let $\Theta = 2^{\mathsf{LFormulae}}(\mathcal{G})$ be a collection of sets of epistemic formulae and $\Phi \in \Theta$ be a finite set of epistemic formulae. Let $x \in \mathbb{Q}$ be a rational number. We introduce the following **coefficient-based** effect functions:

- regular: $\delta_p^x(\Phi) = g$ where $g(\mathtt{A}) = x$ for $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$
- selective: $\delta_c^x(\Phi) = g$ where $g(\mathtt{A}) = x$ for $\mathtt{A} \notin \mathsf{FArgs}(\Phi)$ and $g(\mathtt{A}) = 1$ otherwise
- redundancy-free selective : $\delta_s^x(\Phi) = g$ where $g(\mathtt{A}) = x$ for $\mathtt{A} \notin \mathsf{MinFArgs}^{\mathsf{CNF}}(\Phi)$ and $g(\mathtt{A}) = 1$ otherwise

We note that the input for δ_p does not play any role and is in fact redundant; nevertheless, for the sake of uniformity, we keep the definition in its current form.

Example 15. Consider a sunny Saturday in the countryside where a group of friends (including Peter, Jane, Susan, and John) have just arrived for a picnic in a hay field. Peter does not know John well but Jane and Susan do. Suppose Peter overhears a confidential conversation between Jane and Susan regarding John, as depicted in Figure 8. Jane says A, then Susan replies with B, and finally Jane concludes with

C. Initially, Peter does not think too much about what has been said, but being optimistic by nature, he may subconsciously form a probability distribution that assigns the following beliefs to the arguments: $P_0(\mathtt{A}) = 0.9$, $P_0(\mathtt{B}) = 0.1$, and $P_0(\mathtt{C}) = 0.9$ (see Table 7). Furthermore, he is a rather logical reasoner, and we can suppose that the constraints $\varphi_1 : p(\mathtt{A}) + p(\mathtt{B}) = 1$ and $\varphi_2 : p(\mathtt{B}) + p(\mathtt{C}) = 1$ represent how he sees the relations between A, B and C.

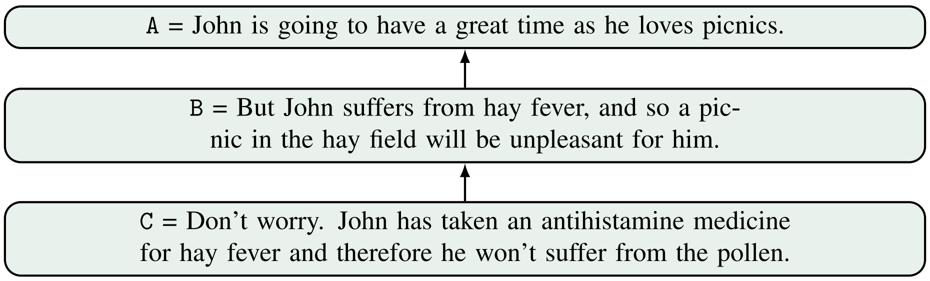


Figure 8: A dialogue between Jane (arguments A and C) and Susan (argument B) concerning their friend John.

Now suppose that Peter looks across to John and sees that John is sneezing a lot, and his eyes are red and swollen. This contradicts Peter's current views and forces him to rethink his position. We can model this with an update with the set $\Psi = \{p(\mathtt{B}) = 0.9\}$.

If we were only concerned with Peter's reasoning as expressed by the constraints, we would perform an update in a standard way with U_{d_2} and obtain the distribution $\{H\} = \mathsf{U}_{d_2}(P_0, \mathcal{C}, \Psi)$ where $\mathcal{C} = \{\varphi_1, \varphi_2\}$ as presented in Table 7. We observe that $H(\mathtt{A}) = 0.1$, $H(\mathtt{B}) = 0.9$, and $H(\mathtt{C}) = 0.1$. This is because $H(\mathtt{B}) = 0.9$ would reflect the update, and $H(\mathtt{A}) = 0.1$ and $H(\mathtt{C}) = 0.1$ would result from the constraints.

However, suppose Peter has a lot of faith in what his friends say and hence some reluctance to jump to conclusions from his observations. This could be reflected by using a coefficient-based effect function such as the regular function with $x = 0.25$. This would decrease the influence of the observation on the belief in B, and as a consequence on A and C. Performing an update with the above effect function, H as the hypothesized distribution and U_{d_2} as the underlying update method would yield the distribution P'_p visible in Table 7. We observe that only a quarter of the expected change in beliefs occurs.

Let us now consider using selective effect functions rather then the regular one. We can observe that the formula $p(\mathtt{B}) = 0.9$ is as minimal as it can get, thus causing the both of the versions of this function to behave in the same way (see P'_c and P'_s visible in Table 7). Since the selective effect functions only limit the belief changes to arguments not explicitly stated in the update, we observe that in both cases B is believed with the degree of 0.9, in contrast to 0.3 in the update with regular effect function.

In order to highlight the difference between the two selective functions, let us consider a slightly modified set of update formulae $\Psi' = \{p(\mathtt{B}) = 0.9 \wedge p(\mathtt{A}) \geq 0\}$. We observe that $p(\mathtt{A}) \geq 0$ is in fact a tautology in epistemic language, and is completely redundant in this formula. Hence, $\mathsf{Sat}(\Psi) = \mathsf{Sat}(\Psi')$. Nevertheless, due to syntactical differences, performing an update with a selective effect function with Ψ yields a different result than with Ψ' (see P'_c versus P''_c in in Table 7). Using the redundancy-free version of the selective effect function does not exhibit this behaviour, which depending on the argumentation scenario we are dealing with, may or may not be desirable.

4.5.2 Distance-Based Effect Functions

The distance-based effect functions vary the adjusting effect for an argument depending on a given notion of distance between the argument and the updating formulae. Of particular interest are approaches in which the effect of an argument depends on its graph-based distance from the arguments in the updating formulae. For example, a function in which the effect decreases with respect to the distance can be used to reflect the cognitive capacity of a given agent. Thus, in the context of this work, we will consider measuring the distance between an argument and epistemic formulae with respect to a given epistemic graph. We will understood it as the length of the shortest undirected path between the given argument and the arguments contained either in the updating formulae or in a refinement of these formulae with respect to minimality.

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$
P_0	0.9	0.1	0.9
$\{H\} = \mathsf{U}^1_{d_2}(P_0, \mathcal{C}, \Psi)$	0.1	0.9	0.1
$\{P'_p\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [H, \delta^x_p, \Psi)])$	0.7	0.3	0.7
$\{P'_c\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [H, \delta^x_c, \Psi)])$	0.7	0.9	0.7
$\{P'_s\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [H, \delta^x_s, \Psi)])$	0.7	0.9	0.7
$\{P''_p\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi', [H, \delta^x_p, \Psi)])$	0.7	0.3	0.7
$\{P''_c\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi', [H, \delta^x_c, \Psi)])$	0.1	0.9	0.7
$\{P''_s\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi', [H, \delta^x_s, \Psi)])$	0.7	0.9	0.7

Table 7: Effects of using different coefficient-based effect functions in Example 15. The table presents the resulting beliefs in arguments in selected distributions and not their full descriptions. We assume that $\mathcal{C} = \{\varphi_1, \varphi_2\}$, $\Psi = \{p(\mathtt{B}) = 0.9\}$, $\Psi' = \{p(\mathtt{B}) = 0.9 \wedge p(\mathtt{A}) \geq 0\}$, and $x = 0.25$.

Definition 4.7. Let $X = (\mathcal{G}, \mathcal{L}, \mathcal{C})$ be an epistemic graph, $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$ an argument, and $\Phi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of epistemic formulae. For arguments $\mathtt{B}$ and $\mathtt{C}$, let $\mathsf{sp}(\mathtt{B}, \mathtt{C})$ denote the length of the shortest path between $\mathtt{B}$ and $\mathtt{C}$ in $\mathcal{G}$. W define the following:

- **impact distance**: $\mathsf{id}(\mathtt{A}, \Phi) = \min_{\mathtt{E} \in \mathsf{FArgs}(\Phi)} \mathsf{sp}(\mathtt{E}, \mathtt{A})$
- **normalized impact distance**: $\mathsf{nid}(\mathtt{A}, \Phi) = \min_{\mathtt{E} \in \mathsf{MinFArgs}^{\mathsf{CNF}}(\Phi)} \mathsf{sp}(\mathtt{E}, \mathtt{A})$

Example 16. Consider an epistemic graph based on a graph $\mathcal{G}$ s.t. $\mathsf{Nodes}(\mathcal{G}) = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$ and $\mathsf{Arcs}(\mathcal{G}) = \{(\mathtt{A}, \mathtt{B}), (\mathtt{B}, \mathtt{C}), (\mathtt{C}, \mathtt{D})\}$. Consider a set of formulae $\Phi = \{\varphi : (p(\mathtt{A}) = 1 \wedge p(\mathtt{C}) = 1) \vee p(\mathtt{A}) = 1\}$. We observe that the minimal CNF of φ is $\varphi' : p(\mathtt{A}) = 1$. Hence, $\mathsf{FArgs}(\Phi) = \{\mathtt{A}, \mathtt{C}\}$ and $\mathsf{MinFArgs}^{\mathsf{CNF}}(\Phi) = \{\mathtt{A}\}$. The minimal distances between arguments in the graph and in the formulae according to given restrictions are tabulated below.

	A	B	C	D
$\mathsf{id}(\mathtt{A}, \Phi)$	0	1	0	1
$\mathsf{nid}(\mathtt{A}, \Phi)$	0	1	2	3

There are various ways distance-based effect functions can be defined, however, in the context of this work we consider the following two options, based on the above impact distances.

Definition 4.8. Let $\Theta = 2^{\mathsf{LFormulae}}(\mathcal{G})$ be a collection of sets of epistemic formulae and $\Phi \in \Theta$ be a finite set of epistemic formulae. Let $\langle \Phi_1, \ldots, \Phi_n \rangle \in \Theta$ be such a sequence, let $j \geq 0$ be a natural number, and let $x \in [0, 1]$ be a rational number. We introduce the following **distance–based** effect functions:

- decreasing: $dis^x_j(\Phi) = g$ s.t. for $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$,

$$g(\mathtt{A}) = \begin{cases} 0 & \mathsf{id}(\mathtt{A}, \Phi) = \infty \\ x^v \text{ where } v = \max(\{\mathsf{id}(\mathtt{A}, \Phi) - j, 0\}) & \text{otherwise} \end{cases}$$

- redundancy-free decreasing: $s\text{-}dis^x_j(\Phi) = g$ s.t. for $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$,

$$g(\mathtt{A}) = \begin{cases} 0 & \mathsf{nid}(\mathtt{A}, \Phi) = \infty \\ x^v \text{ where } v = \max(\{\mathsf{nid}(\mathtt{A}, \Phi) - j, 0\}) & \text{otherwise} \end{cases}$$

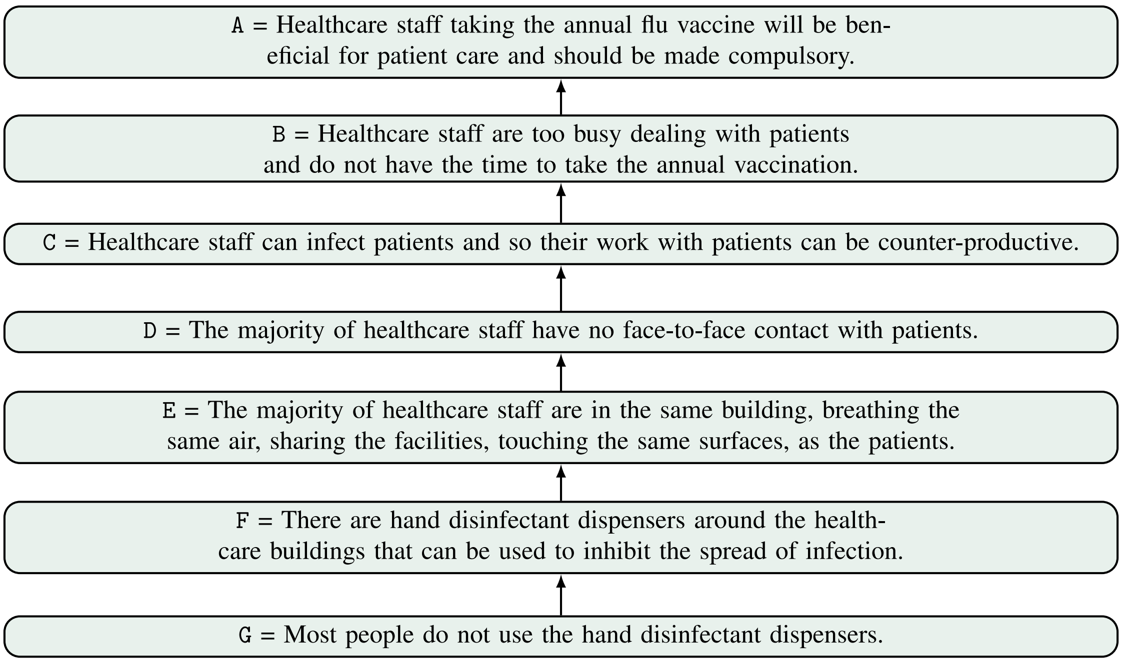


Figure 9: Example of an argument graph concerning influenza vaccination for healthcare staff

In the above definition, x plays the role of a base factor for the changes, and j adjusts the point from which the effects should be applicable. In other words, for $j = 0$, only the arguments present in the formulae (or, where appropriate, in their refinements with respect to minimality) will be unaffected by adjusting effects. For $j = 1$, it is both these arguments and their direct neighbours, and so on. We also assume that if the arguments in the formulae are disconnected from a given argument in the graph, then they have no effect on it.

Example 17. Consider the argument graph in Figure 9. Suppose that a politician in a debate is trying to persuade her colleague to accept argument A. Also suppose she assumes that her colleague has low belief in A, high belief in B, and undecided about the other arguments (perhaps he is unlikely to have given them much thought, or she simply does not know what he may think). This could be represented by with a distribution P_S as listed in Table 8. She knows he can be somewhat radical in his reasoning and decides to represent his general reasoning pattern with constraints of the form $p(\mathtt{X}) + p(\mathtt{Y}) = 1$ where X and Y are the attacked and the attacking arguments respectively. We can gather these constraints in a set $\mathcal{C}$.

Now suppose the politician wants to choose an argument to present to the said colleague. So she could choose to present one of C, E, or G, with the expectation that she gets him to fully believe the presented argument, and by a chain reaction, believe argument A. We can therefore consider the update formulae $\Psi_1 = \{p(\mathtt{C}) = 1\}$, $\{p(\mathtt{E}) = 1\}$ and $\{p(\mathtt{G}) = 1\}$, and if we were to use a standard update method such as U_{d_2}, we would obtain the distributions $P_{\mathtt{C}}$, $P_{\mathtt{E}}$ and $P_{\mathtt{G}}$ in which the arguments are believed as listed in Table 8.

We can observe that the resulting beliefs in arguments are the same and can be seen as describing the status associated with arguments under one of the classical Dung's semantics [30, 17]. This may not be entirely intuitive. An update with $p(\mathtt{C}) = 1$ will influence A via one intermediate argument (namely B) and we may expect that the influence will be quite strong. However, the update with $p(\mathtt{G}) = 1$ will influence A via 5 intermediate arguments (namely B, C, D, and E) and we may expect that the influence on A will be substantially weaker than in the previous case, perhaps even more so because we believe that the colleague has not given these arguments much thought anyway.

We can capture this using a distance-based effect function, such as the decreasing one $dis_0^{0.8}$. For instance, for $dis_0^{0.8}(\Psi_1)$ we obtain the following assignment:

$$g(\mathtt{A}) = 0.64, g(\mathtt{B}) = 0.8, g(\mathtt{C}) = 1, g(\mathtt{D}) = 0.8, g(\mathtt{E}) = 0.64, g(\mathtt{F}) = 0.512, g(\mathtt{G}) = 0.4096$$

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$	$P(\mathtt{E})$	$P(\mathtt{F})$	$P(\mathtt{G})$
P_S	0.1	0.9	0.5	0.5	0.5	0.5	0.5
$\{P_{\mathtt{C}}\} = \mathsf{U}^1_{d_2}(P_S, \mathcal{C}, \Psi_1)$	1	0	1	0	1	0	1
$\{P_{\mathtt{E}}\} = \mathsf{U}^1_{d_2}(P_S, \mathcal{C}, \Psi_2)$	1	0	1	0	1	0	1
$\{P_{\mathtt{G}}\} = \mathsf{U}^1_{d_2}(P_S, \mathcal{C}, \Psi_3)$	1	0	1	0	1	0	1
$\{P^*_{\mathtt{C}}\} = \mathsf{HBU}(P_S, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_1, [P_{\mathtt{C}}, dis_0^{0.8}, \Psi_1])$	0.676	0.18	1	0.1	0.82	0.244	0.705
$\{P^*_{\mathtt{E}}\} = \mathsf{HBU}(P_S, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_2, [P_{\mathtt{E}}, dis_0^{0.8}, \Psi_2])$	0.469	0.439	0.82	0.1	1	0.1	0.82
$\{P^*_{\mathtt{G}}\} = \mathsf{HBU}(P_S, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_3, [P_{\mathtt{G}}, dis_0^{0.8}, \Psi_3])$	0.336	0.605	0.705	0.244	0.82	0.1	1

Table 8: Differences between performing standard and hypothesized belief updates with distance-based effect functions.

Whereas for $dis_0^{0.5}(\Psi_3)$ we obtain the following effect function.

$$g(\mathtt{A}) = 0.262144, g(\mathtt{B}) = 0.32768, g(\mathtt{C}) = 0.4096, g(\mathtt{D}) = 0.512, g(\mathtt{E}) = 0.64, g(\mathtt{F}) = 0.8, g(\mathtt{G}) = 1$$

By taking $P_{\mathtt{C}}$, $P_{\mathtt{E}}$ and $P_{\mathtt{G}}$ as the hypothesized distributions, we would obtain the starred distributions from Table 8. From this, the assumption of degradation in influence over a sequence of arguments strongly suggests that the politician should use argument `C` rather than e.g. argument `G`.

4.5.3 Progress-Based Effect Functions

Progress-based effect functions are meant to vary the effects based on some notion of "progress", such as the number of the performed updates, which can be seen as corresponding to discussion length in a dialogical setting. As an example we consider the cvcl-success-rate, taken from the CMV success rate with respect to the number of exchanged messages between discussion participants [71]. The results show that the susceptibility of an agent to opinion change behaves non–monotonically with respect to the length of the dialogue. In particular, it initially increases as the dialogue progresses, only to virtually drop to 0 after five exchanges. Although the success rate of a dialogue and the susceptibility of an agent are related, but not the same thing, we believe this non-monotonism is worth highlighting and studying further.

Definition 4.9. Let $M > 0$ be a natural number denoting the maximum number of sets of formulae to be considered. Let $\Theta = \bigcup_{i=1}^{M} \mathsf{Sequence}(2^{\mathsf{LFormulae}}(\mathcal{G}), i)$ be a collection of sequences of sets of epistemic formulae of length at most M. We introduce the following **progress–based** function:

- cvcl-success-rate: $sc(\langle \Phi_1, \ldots, \Phi_n \rangle) = g$ s.t. for every argument $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$,

$$g(\mathtt{A}) = \begin{cases} 0.033 & n = 1 \\ 0.051 & n = 2 \\ 0.058 & n = 3 \\ 0.053 & n = 4 \\ 0.0003 & n = 5 \\ 0 & \text{otherwise} \end{cases}$$

In other words, depending on how many updates we have performed so far, the amount by which we can change the agent's beliefs differs. In particular, after a certain number of steps, no changes can occur anymore and the adjusting effects drop down to 0. Please note that the numbers we present are approximations extracted from the results presented in [71]. In the future, we can also consider functions with more monotonic behaviour, as exemplified in the following scenario.

Example 18. Consider a school student with a fondness for junk food getting repeatedly told to stop eating it, for instance:

`A` = *Stop eating junk food, it is bad for your health.*

This repeated message may come from different sources such as his parents, his sports coach, his doctor, etc. Each time the health message is presented, we may represent it with a formula set $\Psi = \{p(\mathtt{A}) = 1\}$ and update the model of the student. In other words, when the student gets this message, he would consider its argument true (i.e. the belief is A is 1). However, being directly told to do or not to do something can be met with certain resistance, perhaps even more so in the case of teenagers. While we can expect that initially the student will revise his opinions to some extent, it is possible that there comes a point that the more this message is repeated, the angrier the student becomes. As a result, he may completely oppose any belief changes stemming from Ψ, or even start behaving in a contrary manner.

We can represent this behaviour by a progress-based function δ s.t. resistance to changes increases with the length of the sequence of messages.

- Suppose initially the student has the belief distribution P_0 such that $P_0(\mathtt{A}) = 0.3$.
- Now suppose that after the first message, we perform an update with Ψ and that $\delta(\langle\Psi\rangle)(\mathtt{A}) = 0.5$. We can let the hypothesized distribution be such that $H_1(\mathtt{A}) = 1$. Hence, the hypothesized belief update would yield a distribution P_1 where $P_1(\mathtt{A}) = 0.65$.
- After the second message, we again perform an update with Ψ and take H_1 as our hypothesized distribution. However, this time, $\delta(\langle\Psi,\Psi\rangle)(\mathtt{A}) = 0.1$. Hence, $P_2(\mathtt{A}) = 0.685$.
- After the third message, the student gets defensive and displeased. We thus take $\delta(\langle\Psi,\Psi,\Psi\rangle)(\mathtt{A}) = -0.3$ and performing an update with Ψ and H_1 leads to a distribution P_3 s.t. $P_3(\mathtt{A}) \approx 0.59$.

Please note that the above resisting behaviour is not the only one possible. There are also dual examples where the susceptibility can increase rather than decrease based on the progression of the dialogue. For instance, it is common that the more news sites report on a certain event or the more people share with us the same piece of gossip, the more likely we are to believe it. Another example is advertising. We can expect that people exposed to advertisements promoting a given smartphone, who see more people using this smartphone, and who have friends recommending it, will be more likely to purchase one. The effect can also occur in psychotherapy, when an initially resisting patient eventually starts reflecting on their behaviour and becomes more compliant and open. We will investigate susceptibility increasing progress-based effect functions in the future.

4.5.4 Properties of Effect Functions

Effect functions can be classified based on the assignment functions they produce. For example, we can distinguish the following effect functions.

Definition 4.10. Let $\delta : \Theta \to (\mathsf{Nodes}(\mathcal{G}) \to \mathbb{Q})$ be an effect function. Then δ is:

- a **diminishing effect function** if for every $\theta \in \Theta$ and $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $0 \leq \delta(\theta)(\mathtt{A}) \leq 1$
- a **boosting effect function** if for every $\theta \in \Theta$ and $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) \geq 1$
- a **contrary effect function** if for every $\theta \in \Theta$ and $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) \leq 0$
- an **altering effect function** if it is neither diminishing nor boosting nor contrary
- a **neutral effect function** if for every $\theta \in \Theta$ and $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) = 1$
- an **impervious effect function** if for every $\theta \in \Theta$ and $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) = 0$
- a **uniform effect function** if for every $\theta \in \Theta$, there exists a value $x \in \mathbb{Q}$ s.t. for every $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) = x$
- a **constant effect function** if there exists a value $x \in \mathbb{Q}$ s.t. for every $\theta \in \Theta$ and every $\mathtt{A} \in \mathsf{Nodes}(\mathcal{G})$, $\delta(\theta)(\mathtt{A}) = x$

With $\mathsf{Dim}(\Theta, \mathcal{G})$ (resp. $\mathsf{Boost}(\Theta, \mathcal{G})$, $\mathsf{Cont}(\Theta, \mathcal{G})$, $\mathsf{Alt}(\Theta, \mathcal{G})$, $\mathsf{Neu}(\Theta, \mathcal{G})$, $\mathsf{Imp}(\Theta, \mathcal{G})$, $\mathsf{Uni}(\Theta, \mathcal{G})$ and $\mathsf{Con}(\Theta, \mathcal{G})$) we denote the collections of effect functions on Θ and $\mathcal{G}$ that are diminishing (resp. boosting, contrary etc.). With $\mathcal{E}(\Theta, \mathcal{G})$ we denote the collection of all effect functions on Θ and $\mathcal{G}$.

The diminishing and boosting effect functions respectively decrease and increase the change in beliefs that would have occurred without such effects. The contrary effect function can be used to represent the reasoning of a person who, when told to do something, does the opposite. An altering function is simply one that, at different points, can exhibit any of these features. A neutral effect function is one that in no way affects the changes in belief and the adjusted distribution will be the same as the hypothesized one. The impervious function represents a person who refuses to change their mind independently of the provided information. In other words, the adjusted distribution will always be the same as the one we wanted to update. Finally, uniform and constant effect functions represent the reasoning in which all arguments are assigned the same effects, independently of their nature - the uniform approach simply adjusts the effect depending on the input, while the constant function does it across all inputs.

We can observe that the following relations hold between our properties (see also Figure 10):

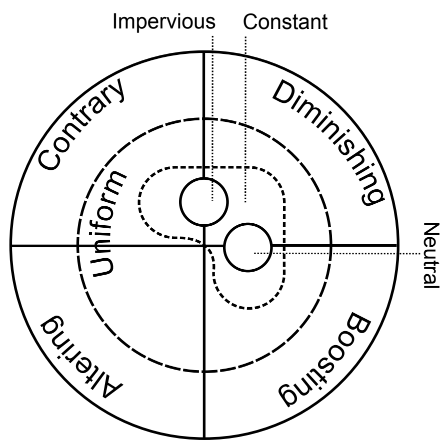


Figure 10: Relationships between the types of effect functions presented as a Venn diagram. Note that the set of impervious function is the intersection of the contrary and diminishing function sets and that the set of neutral functions is the intersection of the diminishing and boosting function sets.

Proposition 4.11. *Let* $\mathsf{Nodes}(\mathcal{G}) \neq \emptyset$ *and* $|\Theta| > 1$. *The following hold:*

- $\mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G}) = \mathcal{E}(\Theta, \mathcal{G})$
- $\mathsf{Dim}(\Theta, \mathcal{G}) \cap \mathsf{Boost}(\Theta, \mathcal{G}) = \mathsf{Neu}(\Theta, \mathcal{G})$
- $\mathsf{Dim}(\Theta, \mathcal{G}) \cap \mathsf{Cont}(\Theta, \mathcal{G}) = \mathsf{Imp}(\Theta, \mathcal{G})$
- $\mathsf{Con}(\Theta, \mathcal{G}) \subset \mathsf{Uni}(\Theta, \mathcal{G})$
- $\mathsf{Con}(\Theta, \mathcal{G}) \subset \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G})$
- $\mathsf{Uni}(\Theta, \mathcal{G}) \subseteq \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G})$
- *if* $|\mathsf{Nodes}(\mathcal{G})| \geq 2$*, then* $\mathsf{Uni}(\Theta, \mathcal{G}) \subset \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G})$

Let us now consider if and how the functions we have introduced satisfy the above properties. We assume we are working with non-empty and finite graphs. We start by considering the coefficient-based effect functions and observe that by restricting the values of parameter x, we can create functions with the desired properties.

Theorem 4.12. *The conditions in Table 9 hold.*

	δ_p^x	δ_c^x	δ_s^x
Diminishing	$0 \leq x \leq 1$	$0 \leq x \leq 1$	$0 \leq x \leq 1$
Boosting	$x \geq 1$	$x \geq 1$	$x \geq 1$
Contrary	$x \leq 0$	Never	Never
Altering	Never	$x < 0$	$x < 0$
Neutral	$x = 1$	$x = 1$	$x = 1$
Impervious	$x = 0$	Never	Never
Uniform	Always	$x = 1$	$x = 1$
Constant	Always	$x = 1$	$x = 1$

Table 9: Conditions on x under which the listed coefficient-based functions satisfy the given properties. The table should be read as "For values of x stated in the cell, the effect function in the column satisfies the property in the row". "Never" and "Always" entries mean that there is no value for x (resp. for every value of x) that would make the effect function meet the desired property.

Let us now consider the distance-based functions. Due to the fact that in order to measure the distance between formulae and arguments we are relying on the structure of the graph in consideration, the actual properties of the functions depend not only on parameters x and j, but also on how connected a given graph is, as seen in Table 10. Finally, we can consider the cvcl-success-rate function, which is easily classified as a uniform and diminishing function (and in one special case, as a constant one as well).

Theorem 4.13. *Let $\{\mathsf{sp}(\mathtt{A},\mathtt{B}) \mid \mathtt{A},\mathtt{B} \in \mathsf{Nodes}(\mathcal{G})$ and $\mathtt{A},\mathtt{B}$ are connected$\}$ be the set of lengths of shortest paths between all connected nodes in $\mathcal{G}$. Let W denote the maximal value of that set. Let $M > 0$ be a natural number denoting the maximal number of sets of formulae to be considered for cvcl-success-rate function. The conditions in Table 10 hold.*

	dis_j^x and $s\text{-}dis_j^x$	cvcl-success-rate
Diminishing	Always	Always
Boosting	$\mathcal{G}$ is connected and ($x = 1$ or $j \geq W$)	Never
Contrary	Never	Never
Altering	Never	Never
Neutral	$\mathcal{G}$ is connected and ($x = 1$ or $j \geq W$)	Never
Impervious	Never	Never
Uniform	$\mathcal{G}$ is connected and ($x = 1$ or $j \geq W$)	Always
Constant	$\mathcal{G}$ is connected and ($x = 1$ or $j \geq W$)	$M = 1$

Table 10: Conditions on $\mathcal{G}$, x and j under which the listed distance-based functions satisfy given properties and conditions on M under which the cvcl-success-rate progress function satisfies given properties. The table should be read as "Under the conditions stated in the cell, the effect function in the column satisfies the property in the row". "Never" and "Always" entries mean that there are no restrictions on $\mathcal{G}$, j and x (or M) (resp. for all possible scenarios) that would make the effect function meet the desired property.

There are further properties that one can investigate, particularly assuming there exist means of comparing the elements of Θ. This can include topics such as monotonicity, and we will investigate this in the future. We close this section with an example comparing the proposed effect functions.

Example 19. Consider the argument graph from Figure 11 and assume the following set of constraints $\mathcal{C}$:

- $\varphi_1 : p(\mathtt{A}) + p(\mathtt{C}) = 1$
- $\varphi_2 : p(\mathtt{B}) + p(\mathtt{E}) = 1$

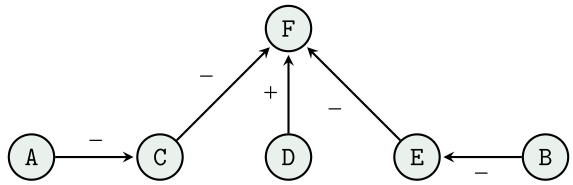


Figure 11: An argument graph

- $\varphi_3 : p(\mathtt{F}) - p(\mathtt{D}) + 0.5 \cdot p(\mathtt{C}) + 0.5 \cdot p(\mathtt{E}) \geq 0$
- $\varphi_4 : p(\mathtt{F}) + 0.5 \cdot p(\mathtt{C}) + 0.5 \cdot p(\mathtt{E}) \leq 1$

We now update a uniform distribution P_0 with a singleton formulae set $\Psi = \{p(\mathtt{A}) = 0.3 \wedge p(\mathtt{B}) = 0.1\}$ and see how different effect functions may change the result. We consider the following distributions, presented in Figure 12:

- P_0 - starting uniform distribution
- $\{P_1\} = \mathsf{U}_{d_2}(P_0, \mathcal{C}, \Psi)$
- $\{P_2\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, \delta^{0.6}_c, \Psi])$
- $\{P_3\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, \delta^{0.6}_p, \Psi])$
- $\{P_4\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, \delta^{1.5}_c, \Psi])$
- $\{P_5\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, \delta^{1.5}_p, \Psi])$
- $\{P_6\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, dis^{0.8}_0, \Psi])$
- $\{P_7\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, sc, \langle\Psi\rangle])$

4.6 Hypothesized Belief Update Properties

Let us now consider the properties of HBU. We first observe that the set of constraints $\mathcal{C}'$ produced by HBU is a set of equality atoms, one atom per argument. Such a set of atoms is always satisfiable [41]:

Proposition 4.14. *Let $B = \{\mathtt{B}_1, \dots, \mathtt{B}_k\} \subseteq \mathsf{Nodes}(\mathcal{G})$ be a non-empty set of arguments and let $\Psi = \{p(\mathtt{B}_1) = x_1, \dots, p(\mathtt{B}_k) = x_k\}$, where $x_i \in [0,1]$, be a set of non-strict epistemic atoms. Then $\mathsf{Sat}(\Psi) \neq \emptyset$.*

According to the definition of HBU (Definition 4.5), only such a set of constraints will be passed to the underlying update function. This is due to the fact that the hypothesized distribution is taken as input, rather than produced internally. This leads us to the observation that as long as the update function U satisfies *Uniqueness* or *Completeness*, so does HBU:

Proposition 4.15. *Hypothesized belief update satisfies* Inherited Uniqueness *and* Inherited Completeness*.*

However, just because U may not meet *Uniqueness* or *Completeness* (or at least, not for all possible inputs), it does not mean that HBU does not. This means as long as the underlying update method to which HBU is delegating meets these properties in the fragment of non-strict epistemic atoms (i.e. not necessarily for the whole language), HBU will always produce exactly one answer.

Proposition 4.16. *If hypothesized belief update delegates to an update function that satisfies* Uniqueness *and* Completeness *in the fragment of non-strict epistemic atoms, then it always produces exactly one distribution.*

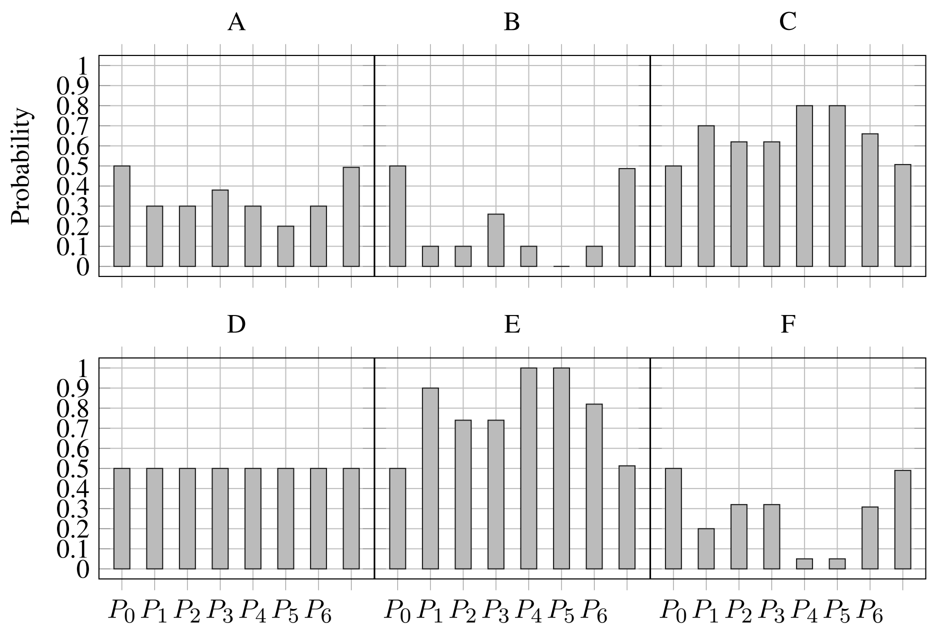


Figure 12: Beliefs in arguments in different possible updated distributions from Example 19.

We note that lack of *Uniqueness* and *Completeness* guarantees on the non–strict epistemic atom language fragment will lead to lack of such guarantees for HBU, independently of the language fragment of the constraints and update formulae.

The above property also implies the *Uniqueness* and *Completeness* of HBU when delegating to the update functions that satisfy *Uniqueness* and *Completeness* in the fragment of non-strict epistemic atoms. However, we observe that it represents a stronger influence due to the fact that *Contradiction* is violated. Additionally, given the results in Theorems 3.5 and 3.11, we can also state the following.

Corollary 4.17. *If hypothesized belief update delegates to an (atomic or standard) distance-minimizing update function, then it always produces exactly one distribution.*

Nevertheless, without any restrictions on the nature of U, neither *Uniqueness* nor *Completeness* can be guaranteed for HBU. There might also be update functions that will lead to HBU satisfying the *Contradiction* property (e.g. consider a trivial update function that for every input returns $\varnothing$), though based on Proposition 4.16, such functions would not give us any guarantees even on the simplest fragment of epistemic language. This can be summarized as follows.

Proposition 4.18. *The hypothesized belief update is not guaranteed to satisfy* Inherited Contradiction*. If the update function to which* HBU *delegates satisfies* Completeness*, then it cannot satisfy* Contradiction*.*

Since HBU depends primarily on the hypothesized distribution and the effect function, then as long as these components and the input distribution and update function remain the same, then the order or the way in which the existing constraints and updating formulae are passed makes no difference. Thus, *Indiscrimination* and various *Representation Invariance* properties are trivially satisfied by HBU:

Proposition 4.19. *The hypothesized belief update function satisfies* Indiscrimination*,* Update Representation Invariance*,* Constraint Representation Invariance*,* Complete Representation Invariance *and their inherited versions.*

However, we would like to observe that satisfaction of these properties in the hypothesized belief update methods is not necessarily informative. The actual information that we use to guide the update is stored within the $[H, \delta, \theta]$ tuple, and as seen in Section 4.5, the effect function input θ can be the set of update formulae themselves. Consequently, any input change concerning Ψ should be paired with a change in

θ in order to obtain meaningful executions. The above properties do not account for this and in order to have a full picture, one should ensure that the chosen effect functions would produce the same effects independently of the Ψ syntax and the choice of a hypothesized distribution would not rely on it either.

The properties we have considered so far concerned the number of answers that HBU could produce and the dependence of the answers on, for example, syntactical, features of the input. Hence, the results depended primarily on the update function U or were trivially satisfied. However, there are also properties such as *Success*, which force certain dependencies between input and output of the update functions, and their satisfaction in many cases goes against the intuition of hypothesized belief updates. We therefore observe that they are easily, and purposefully, violated.

Proposition 4.20. HBU *is not guaranteed to satisfy any of* Epistemic Consistency *and* Success *or their inherited versions.*

We notice that this violation can also happen even with the least intrusive update functions, and this is due to the way the adjusted belief constraints $\mathcal{C}'$ from Definition 4.5 are formed (see Example 20). While it is possible that these properties can be salvaged in simpler classes of epistemic graphs, further investigation is left for future work.

Example 20. Let us consider a graph on arguments $\{\mathtt{A}, \mathtt{B}\}$, a uniform probability distribution P_0, empty set of constraints $\mathcal{C} = \emptyset$ and set of update formulae $\Psi = \{p(\mathtt{A} \wedge \mathtt{B}) = 0.5\}$. Performing an atomic-distance minimizing update yields the distribution P_1 listed in Table 11 (we take a simple weight function w assigning 1 to every argument). Even though $P_1 \neq P_0$, we observe that $P_1(\mathtt{A}) = P_0(\mathtt{A})$ and $P_1(\mathtt{B}) = P_0(\mathtt{B})$. By taking P_1 as a hypothesized distribution for HBU and a neutral regular function δ_p^1, we obtain the distribution P_2 which is identical to P_0. This is due to the fact that the produced adjusted beliefs constraints would be $p(\mathtt{A}) = 0.5$ and $p(\mathtt{B}) = 0.5$, which P_0 already satisfies. Consequently, even though P_1 and P_2 may be equivalent in terms of beliefs they assign to arguments, their structure is different and P_2 is not a satisfying distribution of Ψ.

P	$P(\emptyset)$	$P(\{\mathtt{A}\})$	$P(\{\mathtt{B}\})$	$P(\{\mathtt{A},\mathtt{B}\})$
P_0	0.25	0.25	0.25	0.25
$\{P_1\} = \mathsf{U}^1_{d_2}(P, \mathcal{C}, \Psi)$	0.5	0	0	0.5
$\{P_2\} = \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi, [P_1, \delta_p^1, \Psi])$	0.25	0.25	0.25	0.25

Table 11: Violation of *Epistemic Consistency* and *Success* in delegated updates in Example 20.

Fortunately, the *Tautology*, *Conservatism* and *Idempotence* properties can be satisfied as long as some restrictions are put on the hypothesized distribution, or the effect function and the update function we delegate to.

Proposition 4.21. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If δ is neutral or impervious and* U *satisfies* Conservatism*, then* HBU *satisfies* Idempotence.

Proposition 4.22. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ and* U *satisfies* Conservatism*, then* HBU *satisfies* Conservatism.

Proposition 4.23. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ and* U *satisfies* Tautology *and* Conservatism*, then* HBU *satisfies* Tautology.

Our results are summarized in Table 12, where we list the properties that HBU can satisfy depending on the update functions it delegates to.

Property	U	U_d	U_d^w
Uniqueness	$\times$	✓	✓
Completeness	$\times$	✓	✓
Epistemic Consistency	$\times$	$\times$	$\times$
Success	$\times$	$\times$	$\times$
Tautology	★	☆	☆
Contradiction	$\times$	$\times$	$\times$
Update Representation Invariance	✓	✓	✓
Constraint Representation Invariance	✓	✓	✓
Complete Representation Invariance	✓	✓	✓
Idempotence	$\times$	☆☆	☆☆
Indiscrimination	✓	✓	✓
Conservatism	★	☆	☆
Inherited Uniqueness	✓	✓	✓
Inherited Completeness	✓	✓	✓
Inherited Epistemic Consistency	$\times$	$\times$	$\times$
Inherited Success	$\times$	$\times$	$\times$
Inherited Tautology	★	☆	☆
Inherited Contradiction	$\times$	$\times$	$\times$
Inherited Update Representation Invariance	✓	✓	✓
Inherited Constraint Representation Invariance	✓	✓	✓
Inherited Complete Representation Invariance	✓	✓	✓
Inherited Idempotence	★★	☆☆	☆☆
Inherited Indiscrimination	✓	✓	✓
Inherited Conservatism	☆	☆	☆

Table 12: Satisfaction of delegated update function properties by HBU based on the nature of the update function it delegates to (arbitrary, distance minimizing, and atomic distance minimizing).
☆ - under some assumptions as to the hypothesized distribution
★ - under some assumptions as to the hypothesized distribution and the update function
☆☆ - under some assumptions as to the effect function
★★ - under some assumptions as to the effect function and the update function

5 Related Work

Epistemic graphs are a generalization of epistemic probabilistic argumentation to a setting with more advanced relations between arguments. In [38] it was shown how the epistemic postulates [72, 34, 40, 58] and abstract dialectical frameworks [15, 47], which themselves generalize a wide range of existing argumentation formalisms [56], can be expressed within epistemic graphs. The ability to represent constraints not limited to arguments that are directly connected in the graph also allows epistemic graphs to handle constrained argumentation frameworks [26].

5.1 Weighted, Gradual, and Ranking-based Semantics

Given the fine–grained nature of the epistemic approach, it is natural to compare our proposal to the graded and ranking–based semantics proposed for a number of argumentation frameworks [2, 3, 4, 5, 6, 7, 13, 18, 19, 45, 65, 28, 10, 61, 63, 64]. Although in most of these approaches what the semantics produce can be seen as "assigning numbers from $[0,1]$" to arguments (either as a side or end product), probabilities in the epistemic approach are interpreted as belief, while in the remaining works they are typically left abstract. Thus, many of the postulates set out in the aforementioned methods are not really applicable in the epistemic approach, even though they can be perfectly suitable in other scenarios. We can for instance consider the principles from [5]. Many postulates state how an increase or decrease in beliefs in attackers (or supporters) should be matched with an appropriate decrease or increase in the belief of the target argument. Such properties can, but do not have to be satisfied a given epistemic graph. This can be caused either by the constraints themselves simply not adhering to a given axiom on purpose, or by the constraints not being specific enough. Already a simple formula such as $p(\mathtt{A}) > 0.5 \rightarrow p(\mathtt{B}) \leq 0.5$, which embodies one of the core concepts of the classical epistemic approach [72, 34, 40], violates what is referred to as *Weakening* and *Strengthening*. Furthermore, with the exception of [16], the patterns set out by the graded and ranking–based semantics have to be global, while in our case we can choose to define the way parents of an argument affect it differently for every argument. For instance, in the above frameworks, it would be difficult for an attack relation $(\mathtt{A},\mathtt{B})$ described with a constraint $p(\mathtt{A}) + p(\mathtt{B}) = 1$ to co-exist in the same graph with another attack relation $(\mathtt{C},\mathtt{D})$ described through $p(\mathtt{C}) > 0.5 \leftrightarrow p(\mathtt{D}) \leq 0.5$. We note that this analysis should not be taken as a criticism of the weighted or epistemic approaches, but only as a highlight of striking conceptual differences between them. Further details can be found in [38].

Our approach also shares certain conceptual similarity with the variable-depth propagation semantics from [14, 12] defined for Dung's graphs (i.e. graphs with only binary attack relation). Similarly like in the case of the distance-based effect functions, they allow long lines of argumentation to become ineffective, and do not demand unattacked arguments to be the highest ranked ones. However, unlike in our case, this approach is based on propagation (i.e. repeated re-calculations of values) and is strongly tied to the nature of the underlying graph. The length of the path is used to determine whether the effect of an argument should be positive or negative, and we can expect that any generalization to a graph with more relations would involve a series of case distinctions. Our approach bypasses that, and is based on a general concept that could be straightforwardly reproduced in other frameworks as long as a suitable standard update method was supplied. Finally, it is also designed for a dynamic setting, with explicit considerations of effects of how stating given arguments can impact the underlying belief distribution. The work in [14, 12], while definitely having a practical potential, still requires consideration of how the propagation semantics should account for the progress of, for instance, a dialogue.

5.2 Probabilistic Argumentation

The epistemic approach is not the only form of probabilistic argumentation. One can also name the constellation approach [46, 34], in which we consider a probability distribution over subgraphs of a given graph. The probability of each subgraph is interpreted as its chances of being the "real graph", which is quite distinct from the belief interpretation of the epistemic approach. Hence, despite the fact that both formalisms focus on probabilities, there are significant differences between how they model and use them. Further analysis can be found in [34, 59]. We are also not aware of any works on probabilistic argumentation that would deal with performing updates in the constellation approach, we note that the methods presented in [68] could potentially be harnessed to achieve this.

We also note that there are other works concerning updating epistemic states in argumentation. Applying the standard epistemic approach to modelling persuadee's beliefs in arguments has produced methods for updating beliefs during a dialogue [35, 39]. However, these methods are not equipped to handle epistemic graphs, and, in particular, do not consider the positive relations between arguments. They are also not equipped to handle the concept of refining beliefs through additional information. The work in [37] can be seen as a successor to the previous approaches to a more general setting inspired by the empirical studies we have carried out in [58, 36]. This paper takes [37] further and incorporates it in a more advanced system that aims to provide more realistic update behaviour.

5.3 Belief Revision

The problem of updating an epistemic state with respect to new knowledge has been studied extensively in the belief revision literature that evolved from the AGM theory developed in [1]. An up-to-date discussion of the main ideas can be found in [32]. Our postulates can be roughly related to the AGM postulates. For example, *Uniqueness* and *Completeness* roughly correspond to the *Closure* postulate, which demands that an update yields a well-defined belief state again. *Representation Invariance* corresponds to the *Extensionality* postulate. Our remaining properties can be roughly related to properties in AGM theory as well, but it gets more complicated because probabilistic knowledge bases behave in a different way than classical knowledge bases. Both are monotonic in the sense that adding knowledge can only decrease the number of models. However, the semantical consequences are very different. For classical knowledge bases, monotonicity means that the set of entailed formulas can only increase. For probabilistic knowledge bases, it generally only means that the derived probabilities can change. In particular, while inconsistency in classical logic means that everything is a logical consequence, in probabilistic logics, no probabilities can be derived anymore because there are no probability distributions that satisfy the knowledge base. Therefore, some AGM postulates are difficult to connect to our setting. The closest relative to our setting may be the probabilistic belief change framework from [43]. A more elaborate discussion of relationships between classical and probabilistic belief changes can be found in [43] and [44].

Due to the differences between the classical and probabilistic knowledge bases, classical and probabilistic belief change approaches mainly share high-level ideas. One central common idea is to create a new epistemic state that satisfies the new knowledge, but remains as close as possible to the original state. While this can be difficult to accomplish in classical logics due to the discrete nature of classical models, it is often straightforward for probabilistic logics. This is because probability distributions can often be identified with probability vectors in real spaces, so that common measures can be applied. Furthermore, probability theory provides a variety of measures that can be applied for this purpose. Most commonly, variants of the KL divergence are applied to measure the distance, but other measures have been considered as well. Some examples can be found in [23, 11, 62, 67, 66].

Our hypothesized belief updates are based on the idea that beliefs may change to a lesser or greater extent. Related ideas have been considered in non-prioritized belief revision. An overview of the main ideas can be found in [33]. Again, non-prioritized belief change operations are usually designed for classical knowledge bases and so the actual mechanics are very different from what we do here. For example, the new beliefs may either be completely accepted or ignored entirely so that the original beliefs remain unchanged. In between these extremes, non-prioritized belief change operations can balance between which parts of the original beliefs remain unchanged and which parts of the new beliefs will be accepted. We accomplish similar things here. However, instead of removing or replacing formulas or models of a classical knowledge base, we adapt probabilities (degrees of beliefs) here. Non-prioritized belief change operations for structured probabilistic argumentation have been investigated in [70]. Knowledge bases are given as special probabilistic logic programs and the belief change task is, roughly speaking, to incorporate a new fact or rule. This is accomplished by adapting the knowledge base similar to classical approaches. In particular, some classical axioms for (non-prioritized) belief changes are satisfied by this approach [70]. However, there is again no simple translation to our framework because we adapt an epistemic state that is represented by a probability distribution rather than by a knowledge base.

6 Conclusions & Future Work

In this paper, we have investigated the formal modelling of an agent in terms of the beliefs they may have in arguments and means of updating such models when new information is presented to an agent. We take the recently introduced epistemic graphs as the base framework for our investigations due to their flexibility in modelling fine-grained acceptability, different relationships between arguments, context-sensitivity, and imperfect agents. We have extended our previous work on the belief updates [39, 38, 37] by introducing the notion of delegated updates, which allow us to take into account the external and internal features of the agent being modelled in order to modulate the belief changes. Following this we have proposed hypothesized belief updates as a specific proposal for delegated updates which allow us to better model the belief change in agents that, for various reasons, can be more susceptible or resistant to modifying their opinions in light of new information. We have then investigated the properties of these updates and their components, thus producing a rich but viable formalism for modelling participants in argumentation. These models can then be used in applications such as automated persuasion systems, conversational agents, decision support systems, simulations of communities of reasoning agents, and more.

There is still a wide range of questions and topics to be investigated in the context of delegated updates. For instance, in this work, we had assumed that whenever possible, the input hypothesized distribution would be a result of performing a standard update first, which is later adjusted and refined. Nevertheless, as the approach below shows, this is not the only possible approach.

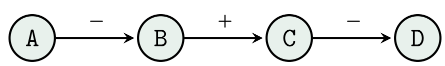


Figure 13: An argument graph

Example 21. Let us consider the graph presented in Figure 13 and the following set of constraints $\mathcal{C}$:

- $\varphi_1 : p(\texttt{B}) + p(\texttt{A}) = 1$
- $\varphi_2 : (p(\texttt{B}) > 0.8 \rightarrow p(\texttt{C}) \geq 0.6) \wedge (p(\texttt{B}) < 0.5 \rightarrow p(\texttt{C}) \leq 0.3)$
- $\varphi_3 : p(\texttt{B}) - p(\texttt{C}) \geq 0.1$
- $\varphi_4 : p(\texttt{C}) > 0.5 \rightarrow p(\texttt{D}) \leq 0.4$
- $\varphi_5 : p(\texttt{C}) < 0.5 \rightarrow p(\texttt{D}) \geq 0.8$

Let us start with a uniform distribution P_0 and consider updates with formulae $\Psi_1 = \{p(\texttt{A}) = 0.6\}$, $\Psi_2 = \{p(\texttt{B}) = 0.7\}$ and $\Psi_3 = \{p(\texttt{C}) = 0.5\}$, one update at a time. The updated distributions after each set is applied without any adjusting effects are visible in Table 13 (distributions P_1, P_2 and P_3).

By considering the decreasing distance-based function $dis_0^{0.9}$ and taking as hypothesized distributions those we would obtain from the previous step through a standard update, we would obtain distributions P_1' to P_3'. This exemplifies the approach we have been using so far in this paper.

By taking P_1, P_2 and P_3 as the hypothesized distributions for respective steps, we would obtain the updated distributions P_1'', P_2'' and P_3''. This exemplifies an approach in which the hypothesized distribution is one we would obtain as if no adjusting effects were present at all from the very first update. We can observe that while in the beginning, both delegated updates have produced similar approaches, they slowly start diverging as the dialogue progresses. It is possible that in more complex scenarios, these differences can start adding up and lead to more significant changes. Thus, in the future, we would like to investigate various options for producing and selecting an appropriate hypothesized distributions for the updates.

Another interesting line of inquiry concerns the effects functions. The examples we have considered are relatively straightforward and our analysis is by no means exhaustive. There are various other methods that are highly interesting to investigate. For instance, we can expect that how well (or how poorly) a given argument aligns with the beliefs of an agent will have a bearing on the degree to which certain arguments are updated. We can consider using measures of similarity and inconsistency between arguments and other related formalisms [8, 31, 29] for creating new effect functions. Furthermore, it is possible that more complex behaviours call for more advanced functions, and there can be a need to investigate approaches exhibiting both distance and progress-based characteristics, approaches that take into account not only the distance between arguments but also whether they are (indirectly) supporting or attacking, and more. Finally, the types of effect functions and inputs we can consider, such as user's mood and the way it changes during a dialogue, can depend on the sensory modules that the artificial agent is equipped with and the way it communicates with the agents it is trying to model. Hence, this topic is another, more advanced line of inquiry.

The task of obtaining epistemic graphs is also another topic we would like to investigate. In this paper, we have not explicitly considered how epistemic graphs or modifiers for updates can be sourced.

P	$P(\mathtt{A})$	$P(\mathtt{B})$	$P(\mathtt{C})$	$P(\mathtt{D})$
P_0	0.5	0.5	0.5	0.5
$P_1 \in \mathsf{U}_{d_2}(P_0, \mathcal{C}, \Psi_1)$	0.6	0.4	0.3	0.8
$P_2 \in \mathsf{U}_{d_2}(P_1, \mathcal{C}, \Psi_2)$	0.3	0.7	0.3	0.8
$P_3 \in \mathsf{U}_{d_2}(P_2, \mathcal{C}, \Psi_3)$	0.3	0.7	0.5	0.8
$P'_1 \in \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_1, [P_1, dis_0^{0.9}, \Psi_1])$	0.6	0.41	0.338	0.719
$P'_2 \in \mathsf{HBU}(P'_1, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_2, [\mathsf{U}_{d_2}(P'_1, \mathcal{C}, \Psi_2), dis_0^{0.9}, \Psi_2])$	0.33	0.7	0.338	0.785
$P'_3 \in \mathsf{HBU}(P'_2, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_3, [\mathsf{U}_{d_2}(P'_2, \mathcal{C}, \Psi_3), dis_0^{0.9}, \Psi_3])$	0.318	0.687	0.5	0.785
$P''_1 \in \mathsf{HBU}(P_0, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_1, [P_1, dis_0^{0.9}, \Psi_1])$	0.6	0.41	0.338	0.719
$P''_2 \in \mathsf{HBU}(P''_1, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_2, [P_2, dis_0^{0.9}, \Psi_2])$	0.33	0.7	0.304	0.785
$P''_3 \in \mathsf{HBU}(P''_2, \mathsf{U}^1_{d_2}, \mathcal{C}, \Psi_3, [P_3, dis_0^{0.9}, \Psi_3])$	0.306	0.7	0.5	0.798

Table 13: Distributions considering various hypothesized distributions from Example 21. We use the sets of formulae $\Psi_1 = \{p(\mathtt{A}) = 0.6\}$, $\Psi_2 = \{p(\mathtt{B}) = 0.7\}$ and $\Psi_3 = \{p(\mathtt{C}) = 0.5\}$, and the decreasing distance-based function $dis_0^{0.9}$.

However, recent developments [21, 36, 58, 48, 22] in empirical studies show that there is a potential of using crowdsourced or other kinds of data (such as obtained from online discussions [71]) for learning how people perceive the relations between arguments and how they react to certain kinds of arguments depending on other factors in the discussion, such as its length, number of participants, and more. We will investigate this topic in more detail in the future.

Last, but not least, hypothesized update is only one of the possible forms of delegated update. Another important approach worth considering relates to performing updates when the constraints and new information are inconsistent together. This calls for additional steps to be employed in order to retrieve some notion of coherence. This could potentially be done by removing (parts of) constraints or updating formulae, by developing methods that would relax them to the point an answer can be obtained, or by considering inconsistency measures for picking the "least wrong" distributions. Furthermore, there is also the option of considering higher level update functions which can delegate to other delegated update methods, not only the standard ones, or employing other chaining techniques. This could, for instance, allow us to produce input distributions for hypothesized belief update even when dealing with inconsistent scenarios, as well as offer an alternative for handling more advance effect functions through combining updates equipped with simpler ones. These are all interesting questions on compounding updates for modelling complex behaviours that are worth exploring.

Acknowledgments

Sylwia Polberg and Anthony Hunter were supported by EPSRC Project EP/N008294/1 "Framework for Computational Persuasion".

7 References

[1] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, 1985.

[2] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In W. Liu, V. Subrahmanian, and J. Wijsen, editors, *Proceedings of SUM'13*, volume 8078 of *LNCS*, pages 134–147. Springer, 2013.

[3] L. Amgoud and J. Ben-Naim. Axiomatic foundations of acceptability semantics. In C. Baral, J. Delgrande, and F. Wolter, editors, *Proceedings of KR'16*, pages 2–11. AAAI Press, 2016.

[4] L. Amgoud and J. Ben-Naim. Evaluation of arguments from support relations: Axioms and semantics. In S. Kambhampati, editor, *Proceedings of IJCAI'16*, pages 900–906. AAAI Press, 2016.

[5] L. Amgoud and J. Ben-Naim. Evaluation of arguments in weighted bipolar graphs. In A. Antonucci, L. Cholvy, and O. Papini, editors, *Proceedings of ECSQARU'17*, volume 10369 of *LNCS*, pages 25–35. Springer, 2017.

[6] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Ranking arguments with compensation-based semantics. In C. Baral, J. Delgrande, and F. Wolter, editors, *Proceedings of KR'16*, pages 12–21. AAAI Press, 2016.

[7] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In C. Sierra, editor, *Proceedings of IJCAI'17*, pages 56–62. International Joint Conferences on Artificial Intelligence, 2017.

[8] L. Amgoud, E. Bonzon, J. Delobelle, D. Doder, S. Konieczny, and N. Maudet. Gradual semantics accounting for similarity between arguments. In M. Thielscher, F. Toni, and F. Wolter, editors, *Proceedings of KR'18*, pages 88–97. AAAI Press, 2018.

[9] C. A. Anderson, M. R. Lepper, and L. Ross. Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39(6):1037–1049, 1980.

[10] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, and G. Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015.

[11] C. Beierle and G. Kern-Isberner. A conceptual agent model based on a uniform approach to various belief operations. In B. Mertsching, M. Hund, and Z. Aziz, editors, *Proceedings of KI'09*, volume 5803 of *LNCS*, pages 273–280. Springer, 2009.

[12] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. Argumentation ranking semantics based on propagation. In P. Baroni, T. F. Gordon, T. Scheffler, and M. Stede, editors, *Proceedings of COMMA'16*, volume 287 of *FAIA*, pages 139–150. IOS Press, 2016.

[13] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In D. Schuurmans and M. Wellman, editors, *Proceedings of AAAI'16*, pages 914–920. AAAI Press, 2016.

[14] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A parametrized ranking-based semantics for persuasion. In S. Moral, O. Pivert, D. Sánchez, and N. Marín, editors, *Proceedings of SUM'17*, volume 10564 of *LNCS*, pages 237–251. Springer, 2017.

[15] G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks revisited. In F. Rossi, editor, *Proceedings of IJCAI'13*, pages 803–809. AAAI Press, 2013.

[16] G. Brewka, H. Strass, J. P. Wallner, and S. Woltran. Weighted abstract dialectical frameworks. In S. A. McIlraith and K. Q. Weinberger, editors, *Proceedings of AAAI'18*, pages 1779–1786. AAAI Press, 2018.

[17] M. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93:109–145, 2009.

[18] C. Cayrol and M. Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In L. Godo, editor, *Proceedings of ECSQARU'05*, volume 3571 of *LNCS*, pages 366–377. Springer, 2005.

[19] C. Cayrol and M. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[20] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning*, 54(7):876–899, 2013.

[21] F. Cerutti, N. Tintarev, and N. Oren. Formal arguments, preferences, and natural language interfaces to humans: an empirical evaluation. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *Proceedings of ECAI'14*, volume 263 of *FAIA*, pages 1033–1034. IOS Press, August 2014.

[22] L. A. Chalaguine, F. L. Hamilton, A. Hunter, and H. W. W. Potts. Argument harvesting using chatbots. In S. Modgil, K. Budzynska, and J. Lawrence, editors, *Proceedings of COMMA'18*, volume 305 of *FAIA*, pages 149–160. IOS Press, 2018.

[23] H. Chan and A. Darwiche. A distance measure for bounding probabilistic belief change. *International Journal of Approximate Reasoning*, 38(2):149–174, 2005.

[24] R. Cialdini. *Influence: The Psychology of Persuasion*. Collins Business Essentials. HarperCollins, 2009.

[25] D. H. Cohen. Arguments that backfire. In D. Hitchcock, editor, *Proceedings of OSSA'05: The Uses of Argument*, pages 58–65. OSSA, 2005.

[26] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In P. Doherty, J. Mylopoulos, and C. Welty, editors, *Proceedings of KR'06*, pages 112–122. AAAI Press, 2006.

[27] I. Csiszar. I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability*, 3(1):146–158, 1975.

[28] C. da Costa Pereira, A. G. B. Tettamanzi, and S. Villata. Changing one's mind: Erase or rewind? Possibilistic belief revision with fuzzy argumentation based on trust. In T. Walsh, editor, *Proceedings of IJCAI'11*, pages 164–171. AAAI Press, 2011.

[29] G. De Bona, J. Grant, A. Hunter, and S. Konieczny. Towards a unified framework for syntactic inconsistency measures. In S. A. McIlraith and K. Q. Weinberger, editors, *Proceedings of AAAI'18*, pages 1803–1810. AAAI Press, 2018.

[30] P. M. 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.

[31] J. Grant and A. Hunter. Analysing inconsistent information using distance-based measures. *International Journal of Approximate Reasoning*, 89:3 – 26, 2017. Special Issue on Theories of Inconsistency Measures and Their Applications.

[32] S. Hansson. Logic of belief revision. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

[33] S. O. Hansson. A survey of non-prioritized belief revision. *Erkenntnis*, 50(2-3):413–427, 1999.

[34] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54(1):47–81, 2013.

[35] A. Hunter. Modelling the persuadee in asymmetric argumentation dialogues for persuasion. In Q. Yang and M. Wooldridge, editors, *Proceedings of IJCAI'15*, pages 3055–3061. AAAI Press, 2015.

[36] A. Hunter and S. Polberg. Empirical methods for modelling persuadees in dialogical argumentation. In J. Guerrero, editor, *Proceedings of ICTAI'17*, pages 382–389. IEEE, 2017.

[37] A. Hunter, S. Polberg, and N. Potyka. Updating Belief in Arguments in Epistemic Graphs. In M. Thielscher, F. Toni, and F. Wolter, editors, *Proceedings of KR'18*, pages 138–147. AAAI Press, 2018.

[38] A. Hunter, S. Polberg, and M. Thimm. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *ArXiv*, 2018.

[39] A. Hunter and N. Potyka. Updating probabilistic epistemic states in persuasion dialogues. In A. Antonucci, L. Cholvy, and O. Papini, editors, *Proceedings of ECSQARU'17*, volume 10369 of *LNCS*, pages 46–56. Springer, 2017.

[40] A. Hunter and M. Thimm. Probabilistic argumentation with incomplete information. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *Proceedings of ECAI'14*, volume 263 of *FAIA*, pages 1033–1034. IOS Press, 2014.

[41] A. Hunter and M. Thimm. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, 59:565–611, August 2017.

[42] K. Kathy and C. Tim. Classifying compliance gaining messages: Taxonomic disorder and strategic confusion. *Communication Theory*, 4(1):3–60, 1994.

[43] G. Kern-Isberner. *Conditionals in nonmonotonic reasoning and belief revision: considering conditionals as agents*. Springer, 2001.

[44] G. Kern-Isberner. Linking iterated belief change operations to nonmonotonic reasoning. In G. Brewka and J. Lang, editors, *Proceedings of KR'08*, pages 166–176. AAAI Press, 2008.

[45] J. Leite and J. Martins. Social abstract argumentation. In T. Walsh, editor, *Proceedings of IJCAI'11*, pages 2287–2292. AAAI Press, 2011.

[46] H. Li, N. Oren, and T. J. Norman. Probabilistic argumentation frameworks. In S. Modgil, N. Oren, and F. Toni, editors, *Proceedings of TAFA'11*, volume 7132 of *LNCS*, pages 1–16. Springer, 2011.

[47] T. Linsbichler, M. Maratea, A. Niskanen, J. Wallner, and S. Woltran. Novel algorithms for abstract dialectical frameworks based on complexity analysis of subclasses and sat solving. In J. Lang, editor, *Proceedings of IJCAI'18*, pages 1905–1911. International Joint Conferences on Artificial Intelligence, 2018.

[48] S. Lukin, P. Anand, M. Walker, and S. Whittaker. Argument strength is in the eye of the beholder: Audience effects in persuasion. In M. Lapata, P. Blunsom, and A. Koller, editors, *Proceedings of EACL'17*, pages 742–753. Association for Computational Linguistics, 2017.

[49] H. Nguyen and J. Masthoff. Designing persuasive dialogue systems: Using argumentation with care. In H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerståhl, and P. Øhrstrøm, editors, *Proceedings of PERSUASIVE'08*, volume 5033 of *LNCS*, pages 201–212. Springer, 2008.

[50] R. S. Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2):175–220, 1998.

[51] B. Nyhan and J. Reifler. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, 33(3):459 – 464, 2015.

[52] J. Ogden. *Health Psychology: A Textbook*. Open University Press, 5 edition, 2012.

[53] F. Paglieri. Ruinous arguments: Escalation of disagreement and the dangers of arguing. In J. Ritola, editor, *Proceedings of OSSA'09: Argument Cultures*, pages 1–15. OSSA, 2009.

[54] F. Paglieri and C. Castelfranchi. Why argue? Towards a cost–benefit analysis of argumentation. *Argument & Computation*, 1(1):71–91, 2010.

[55] S. Pluviano, C. Watt, and S. Della Sala. Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLOS ONE*, 12(7):1–15, 2017.

[56] S. Polberg. Understanding the Abstract Dialectical Framework. In L. Michael and A. Kakas, editors, *Proceedings of JELIA'16*, volume 10021 of *LNCS*, pages 430–446. Springer, 2016.

[57] S. Polberg. Intertranslatability of abstract argumentation frameworks. Technical Report DBAI-TR-2017-104, Institute for Information Systems, Technical University of Vienna, 2017.

[58] S. Polberg and A. Hunter. Empirical evaluation of abstract argumentation: Supporting the need for bipolar and probabilistic approaches. *International Journal of Approximate Reasoning*, 93:487 – 543, 2018.

[59] S. Polberg, A. Hunter, and M. Thimm. Belief in attacks in epistemic probabilistic argumentation. In S. Moral, O. Pivert, D. Sánchez, and N. Marín, editors, *Proceedings of SUM'17*, volume 10564 of *LNCS*, pages 223–236. Springer, 2017.

[60] E. M. Pomerantz, S. Chaiken, and R. S. Tordesillas. Attitude strength and resistance processes. *Journal of Personality and Social Psychology*, 69(3):408–419, 1995.

[61] N. Potyka. Continuous dynamical systems for weighted bipolar argumentation. In M. Thielscher, F. Toni, and F. Wolter, editors, *Proceedings of KR'18*, pages 148–157. AAAI Press, 2018.

[62] N. Potyka, C. Beierle, and G. Kern-Isberner. Changes of relational probabilistic belief states and their computation under optimum entropy semantics. In I. J. Timm and M. Thimm, editors, *Proceedings of KI'13*, volume 8077 of *LNCS*, pages 176–187. Springer, 2013.

[63] F. Pu, J. Luo, Y. Zhang, and G. Luo. Argument ranking with categoriser function. In R. Buchmann, C. V. Kifor, and J. Yu, editors, *Proceedings of KSEM'14*, volume 8793 of *LNCS*, pages 290–301. Springer, 2014.

[64] F. Pu, J. Luo, Y. Zhang, and G. Luo. Attacker and defender counting approach for abstract argumentation. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, and P. P. Maglio, editors, *Proceedings of CogSci'15*, page 1. cognitivesciencesociety.org, 2015.

[65] A. Rago, F. Toni, M. Aurisicchio, and P. Baroni. Discontinuity-free decision support with quantitative argumentation debates. In C. Baral, J. Delgrande, and F. Wolter, editors, *Proceedings of KR'16*, pages 63–73. AAAI Press, 2016.

[66] G. Rens, T. Meyer, K.-I. Gabriele, and A. Nayak. Probabilistic belief revision via similarity of worlds modulo evidence. In F. Trollmann and A.-Y. Turhan, editors, *Proceedings of KI'18*, volume 11117 of *LNCS*, page 343. Springer, 2018.

[67] G. B. Rens and T. Meyer. A new approach to probabilistic belief change. In I. Russell and W. Eberle, editors, *Proceedings of FLAIRS'15*, pages 581–587. AAAI Press, 2015.

[68] R. Riveret and G. Governatori. On learning attacks in probabilistic abstract argumentation. In C. M. Jonker, S. Marsella, J. Thangarajah, and K. Tuyls, editors, *Proceedings of AAMAS'16*, pages 653–661. IFAAMAS, 2016.

[69] W. Rudin. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1976.

[70] P. Shakarian, G. I. Simari, G. Moores, D. Paulo, S. Parsons, M. A. Falappa, and A. Aleali. Belief revision in structured probabilistic argumentation. *Annals of Mathematics and Artificial Intelligence*, 78(3-4):259–301, 2016.

[71] C. Tan, V. Niculae, C. Danescu-Niculescu-Mizil, and L. Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In J. Bourdeau, J. Hendler, R. Nkambou, I. Horrocks, and B. Y. Zhao, editors, *Proceedings of WWW'16*, pages 613–624. ACM, 2016.

[72] M. Thimm. A probabilistic semantics for abstract argumentation. In L. D. Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, and P. Lucas, editors, *Proceedings of ECAI'12*, volume 242 of *FAIA*, pages 750–755. IOS Press, 2012.

8 Proof Appendix

Proposition 3.4. *[Extended version of [37]] Every distance-minimizing update functions satisfies* Epistemic Consistency, Success, Tautology, Contradiction, Update Representation Invariance, Constraint Representation Invariance, Complete Representation Invariance, Idempotence, Indiscrimination, *and* Conservatism.

Proof. Satisfaction of the *Epistemic Consistency*, *Success*, *Tautology*, *Contradiction*, *Update Representation Invariance* and *Idempotence* properties has been shown in [37].

We observe that $\mathsf{Sat}(\mathcal{C}_1 \cup \Psi_1) = \mathsf{Sat}(\mathcal{C}_2 \cup \Psi_1) = \mathsf{Sat}(\mathcal{C}_1 \cup \Psi_2) = \mathsf{Sat}(\mathcal{C}_2 \cup \Psi_2)$ when $\mathsf{Sat}(\mathcal{C}_1) = \mathsf{Sat}(\mathcal{C}_2)$ and $\mathsf{Sat}(\Psi_1) = \mathsf{Sat}(\Psi_2)$. Consequently, the corresponding optimization problems are equivalent and U meets the *Constraint Representation Invariance* and *Complete Representation Invariance*. In a similar fashion we can show that *Indiscrimination* holds.

We can show that due to the positive definiteness of d, if $P \in \mathsf{Sat}(\mathcal{C} \cup \Psi)$ then $\{P\} = \arg\min_{P' \in \mathsf{Sat}(\mathcal{C} \cup \Psi)} d(P, P')$. Hence, *Conservatism* holds.

□

Theorem 3.11. *[Extended version of [37]] Every atomic distance-minimizing update functions satisfies* Epistemic Consistency, Success, Tautology, Contradiction, Update Representation Invariance, Constraint Representation Invariance, Complete Representation Invariance, Idempotence, Indiscrimination, *and* Conservatism. *In the fragment of non-strict epistemic formulae,* Completeness *is satisfied as well and* $\mathsf{U}_d^w(P, \Psi)$ *is guaranteed to be finite. In the fragment of non-strict epistemic atoms,* Uniqueness *is also satisfied.*

Proof. General satisfaction of the *Epistemic Consistency*, *Success*, *Tautology*, *Contradiction*, *Update Representation Invariance* and *Idempotence* properties and of *Completeness* and *Uniqueness* in specific language fragments has been noted in [37]. Satisfaction of other *Representation Invariance* properties, *Indiscrimination* and *Conservatism*, can be shown similarly as in Proposition 3.4. □

Proposition 4.11. *Let* $\mathsf{Nodes}(\mathcal{G}) \neq \emptyset$ *and* $|\Theta| > 1$. *The following hold:*

- $\mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G}) = \mathcal{E}(\Theta, \mathcal{G})$
- $\mathsf{Dim}(\Theta, \mathcal{G}) \cap \mathsf{Boost}(\Theta, \mathcal{G}) = \mathsf{Neu}(\Theta, \mathcal{G})$
- $\mathsf{Dim}(\Theta, \mathcal{G}) \cap \mathsf{Cont}(\Theta, \mathcal{G}) = \mathsf{Imp}(\Theta, \mathcal{G})$
- $\mathsf{Con}(\Theta, \mathcal{G}) \subset \mathsf{Uni}(\Theta, \mathcal{G})$
- $\mathsf{Con}(\Theta, \mathcal{G}) \subset \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G})$
- $\mathsf{Uni}(\Theta, \mathcal{G}) \subseteq \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G})$
- *if* $|\mathsf{Nodes}(\mathcal{G})| \geq 2$, *then* $\mathsf{Uni}(\Theta, \mathcal{G}) \subset \mathsf{Dim}(\Theta, \mathcal{G}) \cup \mathsf{Boost}(\Theta, \mathcal{G}) \cup \mathsf{Cont}(\Theta, \mathcal{G}) \cup \mathsf{Alt}(\Theta, \mathcal{G})$

Proof. These properties follow easily from the definitions of these functions. □

Theorem 4.12. *The conditions in Table 9 hold.*

Proof. The properties of δ_p^x follow easily from its definition.

The selective function δ_c^x can never be contrary nor impervious for a non-empty graph due to the fact that $\delta_c^x(\{\top\})$ produces a function assigning effect of 1 to every argument. The same holds for δ_s^x. This also explains why for $x < 0$, these functions are altering.

The remaining properties easily follow from the definitions of these functions. □

Theorem 4.13. *Let* $\{\mathsf{sp}(\mathtt{A}, \mathtt{B}) \mid \mathtt{A}, \mathtt{B} \in \mathsf{Nodes}(\mathcal{G})$ *and* $\mathtt{A}, \mathtt{B}$ *are connected*$\}$ *be the set of lengths of shortest paths between all connected nodes in* $\mathcal{G}$. *Let* W *denote the maximal value of that set. Let* $M > 0$ *be a natural number denoting the maximal number of sets of formulae to be considered for cvcl-success-rate function. The conditions in Table 10 hold.*

Proof. Consider the definitions of the distance functions from Definition 4.8. We observe that based on the properties of j and the considered distances, $v \geq 0$ is a natural number.

We observe that since $0 \geq x \geq 1$, $0 \geq g(\mathtt{A}) \geq 1$. Consequently, our distance-based functions are diminishing independently of other factors.

For $j \geq W$, $v = 0$ whenever applicable. This means that for $j \geq W$, the effects assigned to arguments are 0 (if a given argument is not connected to the arguments in the formulae) or 1 (otherwise). Thus, for a connected graph, the $j \geq W$ condition leads to boosting, uniform, neutral, and constant functions.

We observe that it is not possible to create an impervious effect function. For it to hold, the shortest possible path between any two nodes in the graph should be infinite. However, the shortest path between a node and itself is 0. Hence, in a non-empty graph, we can always choose an argument $\mathtt{A}$ and a formula such as $\{p(\mathtt{A}) = 0.5\}$ that would force any of our distance-based function to assign 1 to $\mathtt{A}$. For the same reason, we cannot create a contrary function.

It is easy to see that a graph that is not connected cannot have a boosting or neutral effect function (i.e. we will always find some set of formulae and some argument s.t. the resulting effect would be 0). We therefore focus on connected graphs only. We observe that by setting $x = 1$, we ensure that only value 1 is assigned to arguments. Thus, such functions would be not only boosting and neutral, but also uniform and constant, independently of j.

Let us now focus on altering functions. An effect function δ is altering if and only if we can find sets of formulae Φ_1, Φ_2 and arguments $\mathtt{A}_1$, $\mathtt{A}_2$ s.t. any of the following cases hold:

- $\delta(\Phi_1)(\mathtt{A}_1) > 0$ and $\delta(\Phi_2)(\mathtt{A}_2) < 0$
- $\delta(\Phi_1)(\mathtt{A}_1) > 1$ and $\delta(\Phi_2)(\mathtt{A}_2) < 1$

Given the fact that $0 \geq g(\mathtt{A}) \geq 1$, this can never happen.

The properties of the cvcl-success-rate effect function follow straightforwardly from the definition. We only note that $M = 1$ (i.e. given that $M > 0$, all the sequences we are considering are of length 1), only the $n = 1$ case of the definition is relevant, thus producing a constant function.

□

Proposition 4.14. *Let $B = \{\mathtt{B}_1, \ldots, \mathtt{B}_k\} \subseteq \mathsf{Nodes}(\mathcal{G})$ be a non-empty set of arguments and let $\Psi = \{p(\mathtt{B}_1) = x_1, \ldots, p(\mathtt{B}_k) = x_k\}$, where $x_i \in [0,1]$, be a set of non-strict epistemic atoms. Then $\mathsf{Sat}(\Psi) \neq \varnothing$.*

Proof. Please note that this proof is an adjustment of the one from [41], which contained certain issues.

Let $B = \{\mathtt{B}_1, \ldots, \mathtt{B}_k\} \subseteq \mathsf{Nodes}(\mathcal{G})$ be a non-empty set of arguments and let $\Psi = \{p(\mathtt{B}_1) = x_1, \ldots, p(\mathtt{B}_k) = x_k\}$, where $x_i \in [0,1]$, be a set of non-strict epistemic atoms. Without the loss of generality, we assume that $x_1 \leq \ldots \leq x_k$.

Define $P : 2^{\mathsf{Nodes}(\mathcal{G})} \to [0,1]$ via (let $i = 1, \ldots, k$):

$$P(\{\mathtt{B}_i, \ldots, \mathtt{B}_k\}) = x_i - \sum_{j=1}^{j=i-1} P(\{\mathtt{B}_j, \ldots, \mathtt{B}_k\})$$

$$P(\varnothing) = 1 - \sum_{j=1}^{k} P(\{\mathtt{B}_j, \ldots, \mathtt{B}_k\})$$

$$P(Y) = 0 \text{ for all remaining sets } Y$$

We observe that this is equivalently expressed with:

$$P(\{\mathtt{A}_1, \ldots, \mathtt{A}_n\}) = x_1$$

$$P(\{\mathtt{A}_i, \ldots, \mathtt{A}_n\}) = x_i - x_{i-1} \text{ for } i = 2, \ldots, n$$

$$P(\varnothing) = 1 - x_k$$

$$P(Y) = 0 \text{ for all remaining sets } Y$$

It is therefore easy to see that for every set of arguments Z, $P(Z) \geq 0$. Furthermore, $\sum_{Z \subseteq \mathsf{Nodes}(\mathcal{G})} P(Z) = x_1 + x_2 - x_1 + x_3 - x_2 + \ldots + x_k - x_{k-1} + 1 - x_k = 1$. Hence, P is indeed a probability distribution.

We can now show that for every B_i, $x_i = P(\mathsf{B}_i)$. Recall that $P(\mathsf{B}_i) = \sum_{Z \subseteq \mathsf{Nodes}(\mathcal{G}), \mathsf{B}_i \in Z} P(Z)$. Given the construction of P, $P(\mathsf{B}_i) = \sum_{j=1}^{i} P(\{\mathsf{B}_j, \ldots, \mathsf{B}_k\})$, which means that $P(\mathsf{B}_i) = x_1 + x_2 - x_1 + \ldots + x_i - x_{i-1} = x_i$. □

Proposition 4.15. *Hypothesized belief update satisfies* Inherited Uniqueness *and* Inherited Completeness.

Proof. Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints, U an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. Let HBU be the hypothesized belief update function and let $\mathcal{C}'$ denote the constraints created by HBU as seen in Definition 4.5.

Let U satisfy *Uniqueness*. This means that $|\mathsf{U}(P, \emptyset, \mathcal{C}')| \leq 1$ and therefore that $|\mathsf{HBU}(P, \mathsf{U}, \mathcal{C}, \Psi, [H, \delta, \theta])| \leq 1$. Hence, HBU satisfies *Uniqueness*, and as a result also *Inherited Uniqueness*.

Let U satisfy *Completeness*. This means that if $\mathsf{Sat}(\mathcal{C}') \neq \emptyset$, then $|\mathsf{U}(P, \emptyset, \mathcal{C}')| \geq 1$. As shown in Proposition 4.14, $\mathsf{Sat}(\mathcal{C}') \neq \emptyset$, hence $|\mathsf{HBU}(P, \mathsf{U}, \mathcal{C}, \Psi, [H, \delta, \theta])| \geq 1$, independently of whether $\mathsf{Sat}(\mathcal{C} \cup \Psi) \neq \emptyset$. Hence, HBU satisfies *Completeness* and *Inherited Completeness* □

Proposition 4.16. *If hypothesized belief update delegates to an update function that satisfies* Uniqueness *and* Completeness *in the fragment of non-strict epistemic atoms, then it always produces exactly one distribution.*

Proof. Follows easily from Proposition 4.15 and the construction of constraints $\mathcal{C}'$ from Definition 4.5. □

Corollary 4.17. *If hypothesized belief update delegates to an (atomic or standard) distance-minimizing update function, then it always produces exactly one distribution.*

Proof. Based on Theorems 3.5 and 3.11, atomic and standard distance-minimizing updates satisfy *Uniqueness* and *Completeness* in the fragment of non-strict epistemic atoms. Thus, based on Proposition 4.16, hypothesized belief updates delegating to such update functions will always produce exactly one distribution. □

Proposition 4.18. *The hypothesized belief update is not guaranteed to satisfy* Inherited Contradiction*. If the update function to which* HBU *delegates satisfies* Completeness*, then it cannot satisfy* Contradiction.

Proof. Let U satisfy *Completeness*. Based on proofs of Propositions 4.14 and 4.15, we can show that HBU always produces at least one distribution, independently of whether the constraints and updating formulae are jointly satisfiable or not. Thus, HBU cannot satisfy *Contradiction*.

We observe that every distance-minimizing function satisfies *Contradiction* and that based on Proposition 4.16, hypothesized belief update functions equipped with such standard functions always produce exactly one answer, thus violating *Inherited Contradiction*. □

Proposition 4.19. *The hypothesized belief update function satisfies* Indiscrimination*,* Update Representation Invariance*,* Constraint Representation Invariance*,* Complete Representation Invariance *and their inherited versions.*

Proof. Based on Definition 4.5 we observe that the set of constraints $\mathcal{C}'$ is the same for $\mathsf{HBU}(P, \mathsf{U}, \mathcal{C}, \Psi, [H, \delta, \theta])$, $\mathsf{HBU}(P, \mathsf{U}, \emptyset, \mathcal{C} \cup \Psi, [H, \delta, \theta])$ and $\mathsf{HBU}(P, \mathsf{U}, \mathcal{C} \cup \Psi, \emptyset, [H, \delta, \theta])$. Hence, in all of these cases, the produced distributions are $\mathsf{U}(P, \emptyset, \mathcal{C}')$, and the *Indiscrimination* property holds. The same can be shown for the *Representation Invariance* properties, and since they are trivially satisfied, their inherited versions follow as well. □

Proposition 4.20. HBU *is not guaranteed to satisfy any of* Epistemic Consistency *and* Success *or their inherited versions.*

Proof. We can consider a simple graph with arguments A and B and a constraint set $\mathcal{C} = \{p(\texttt{A}) = 1 - p(\texttt{B})\}$. Updating a uniform probability distribution P_0 with the set of formulae $\Psi = \{p(\texttt{A}) = 0.6\}$ and an update function U_{d_2} would produce the distribution $P_1 = \mathsf{U}_{d_2}(P_0, \mathcal{C}, \Psi)$ where $P_1(\texttt{A}) = 0.6$ and $P_1(\texttt{B}) = 0.4$. Updating P_0 using the hypothesized belief update with the above input and a selective effect function $\delta_c^{0.8}$ would give us the function $\{P_1'\} = \mathsf{HBU}(P_0, \mathsf{U}_{d_2}, \mathcal{C}, \Psi, [P_1, \delta_p^{0.8}, \Psi])$ s.t. $P_1'(\texttt{A}) = 0.6$ and $P_1'(\texttt{B}) = 0.42$. Clearly, $P_1' \notin \mathsf{Sat}(\mathcal{C})$. Thus, HBU violates *Epistemic Consistency*. By using a regular update function $\delta_p^{0.8}$ instead of $\delta_c^{0.8}$ we would obtain a function P_1'' s.t. $P_1''(\texttt{A}) = 0.58$ and $P_1''(\texttt{B}) = 0.42$, which violates *Success*.

Given that U_{d_2} meets *Epistemic Consistency* and *Success* in the above examples and HBU does not, the inherited versions of these properties do not hold for HBU. □

Proposition 4.21. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If δ is neutral or impervious and* U *satisfies* Conservatism*, then* HBU *satisfies* Idempotence*.*

Proof. We recall that *Idempotence* is defined in the following way: if $\mathsf{DU}(P, \langle \mathsf{U}_1, \ldots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P^*\}$ then $\mathsf{DU}(P^*, \langle \mathsf{U}_1, \ldots, \mathsf{U}_n \rangle, \mathcal{C}, \Psi, \lambda) = \{P^*\}$.

Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints, U an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. Let HBU be the hypothesized belief update function. Let $\mathcal{C}'$ denote the constraints created by HBU as seen in Definition 4.5 for $\mathsf{HBU}(P, \mathsf{U}, \varnothing, \mathcal{C} \cup \Psi, [H, \delta, \theta])$ and $\mathcal{C}''$ denote the constraints created for $\mathsf{HBU}(P^*, \mathsf{U}, \varnothing, \mathcal{C} \cup \Psi, [H, \delta, \theta])$.

Assume that δ is a neutral effect function and that U satisfies *Conservatism*. This means that for every argument $\texttt{A} \in \mathsf{Nodes}(\mathcal{G})$, $P^*(\texttt{A}) = H(\texttt{A})$. Additionally, every constraint $\varphi \in \mathcal{C}''$ is of the form $p(\texttt{A}) = x$ where $x = P^*(\texttt{A})$. Thus, $\{P^*\} = \mathsf{U}(P^*, \varnothing, \mathcal{C}'')$ due to *Conservatism*, and HBU satisfies *Idempotence*.

Assume that δ is an impervious effect function and that U satisfies *Conservatism*. This means that for every argument $\texttt{A} \in \mathsf{Nodes}(\mathcal{G})$, $P^*(\texttt{A}) = P(\texttt{A})$. Additionally, every constraint $\varphi \in \mathcal{C}''$ is of the form $p(\texttt{A}) = x$ where $x = P^*(\texttt{A})$. Thus, $\{P^*\} = \mathsf{U}(P^*, \varnothing, \mathcal{C}'')$ due to *Conservatism*, and HBU satisfies *Idempotence*. □

Proposition 4.22. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ and* U *satisfies* Conservatism*, then* HBU *satisfies* Conservatism*.*

Proof. Let $\mathcal{C}'$ denote the constraints created by HBU as seen in Definition 4.5 for $\mathsf{HBU}(P, \mathsf{U}, \varnothing, \mathcal{C} \cup \Psi, [H, \delta, \theta])$.

If U satisfies *Conservatism*, then $H = P$. Hence, every constraint $\varphi \in \mathcal{C}'$ is of the form $p(\texttt{A}) = x$ where $x = P(\texttt{A})$. As a result, $P \in \mathsf{Sat}(\mathcal{C}')$ and $\{P\} = \mathsf{U}(P, \varnothing, \mathcal{C}')$ due to *Conservatism*. Hence, $\mathsf{HBU}(P, \mathsf{U}, \varnothing, \mathcal{C} \cup \Psi, [H, \delta, \theta]) = \{P\}$ and HBU satisfies *Conservatism* as well. □

Proposition 4.23. *Let $\mathcal{G}$ be a graph, $\mathcal{C} \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of constraints,* U *an update method, $\Psi \subseteq \mathsf{LFormulae}(\mathcal{G})$ a set of linear epistemic formulae, $P, H \in \mathsf{Dist}(\mathcal{G})$ the current and a hypothesized probability distribution. Let δ be an effect function and $\theta \in \Theta$ its input. If $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ and* U *satisfies* Tautology *and* Conservatism*, then* HBU *satisfies* Tautology*.*

Proof. Let $P \in \mathsf{Sat}(\mathcal{C})$ and let $\mathsf{Sat}(\Psi) = \mathsf{Sat}(\top)$. Since $H \in \mathsf{U}(P, \mathcal{C}, \Psi)$ and U satisfies *Tautology*, it holds that $P = H$. Consequently, all the constraints in $\mathcal{C}'$ in Definition 4.5 are of the form $p(\texttt{A}) = x$ where $x = P(\texttt{A})$ given an argument $\texttt{A} \in \mathsf{Nodes}(\mathcal{G})$. Hence, $P \in \mathsf{Sat}(\mathcal{C}')$, and due to *Conservatism*, $\mathsf{U}(P, \varnothing, \mathcal{C}') = \{P\} = \mathsf{HBU}(P, \mathsf{U}, \mathcal{C}, \Psi, [H, \delta, \theta])$. □