Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Probabilistic Belief Contraction Using Argumentation

Kinzang Chhogyal[1,2], **Abhaya Nayak**[2], **Zhiqiang Zhuang**[1], **Abdul Sattar**[1]
[1] Griffith University, Brisbane, Australia | [2] Macquarie University, Sydney, Australia
{kin.chhogyal, abhaya.nayak}@mq.edu.au,
{z.zhuang, a.sattar}@griffith.edu.au

Abstract

When a belief state is represented as a probability function P, the resulting belief state of the contraction of a sentence (belief) from the original belief state P can be given by the probabilistic version of the Harper Identity. Specifically, the result of contracting P by a sentence h is taken to be the mixture of two states: the original state P, and the resultant state $P^*_{\neg h}$ of revising P by the negation of h. What proportion of P and $P^*_{\neg h}$ should be used in this mixture remains an open issue and is largely ignored in literature. In this paper, we first classify different belief states by their stability, and then exploit the quantitative nature of probabilities and combine it with the basic ideas of argumentation theory to determine the mixture proportions. We, therefore, propose a novel approach to probabilistic belief contraction using argumentation.

1 Introduction

The epistemic or belief state of an intelligent agent is generally in a state of flux with new beliefs being added and old beliefs being removed. The field of belief change [Alchourrón *et al.*, 1985; Gärdenfors, 1988; Peppas, 2008] deals with modelling the different types of changes an agent's belief state may undergo. The impetus for change is usually a received piece of information (sentence). Conventionally belief change is studied under a propositional logic framework. In its most rudimentary form, the belief state of an agent is represented by a *belief set*, that is a set of sentences that is possibly closed under logical consequence.[1]

If an agent is less certain about one sentence than another, additional mechanisms are needed to capture information about the level of uncertainties. A number of approaches exist for this purpose, including *epistemic entrenchment* of beliefs [Gärdenfors, 1988; Nayak, 1994], *ranking functions* [Spohn, 1988], *probability measures* [Gärdenfors, 1988], *possibility theory* [Dubois *et al.*, 1994], and *Dempster-Shafer belief functions* [Shafer and others, 1976]. In this paper we deal with belief states represented as probability functions which effectively assign probability 1 to beliefs, 0 to disbeliefs, and different intermediate values to non-beliefs. When a belief state is represented as a probability function, a (probabilistic) belief change operation (under a given sentential input) becomes a mapping from probability functions to probability functions. Belief contraction is the process by which a sentence h that is a belief becomes a non-belief. It is well known that in a deterministic framework belief contraction can be defined through belief revision employing the *Harper Identity*.[2] The counterpart of the Harper Identity in the probabilistic setup is presented in [Gärdenfors, 1988] where probabilistic belief contraction is taken to be a *mixture* of two probability functions P and $P^*_{\neg h}$ where P represents the original belief state, and $P^*_{\neg h}$ the result of revising P by the negation of the sentence being contracted (h). This mixture takes as ingredients a proportion of P and a proportion of $P^*_{\neg h}$. However, it is not clear exactly what these proportions should be and to the best of our knowledge, there is no work in literature addressing this. The primary aim of this paper is to propose a plausible solution to this problem. The main idea is to view belief contraction as an argumentation process involving two agents $\mathcal{X}$ and $\mathcal{Y}$ with belief states P and $P^*_{\neg h}$. $\mathcal{X}$ argues for h whereas $\mathcal{Y}$ argues for $\neg h$ and the goal is to reach a compromise which is represented by the contracted belief state P^-_h. Thus, the input for a belief change process is not only a naked piece of information – it should be accompanied by arguments to support it. Different probability functions will generate arguments of different strengths for $\mathcal{X}$ and $\mathcal{Y}$ and we exploit this feature to make an informed choice as to what the proportions of the mixture should be. Section 2 introduces probabilistic belief states and Section 3 briefly reviews probabilistic belief contraction and its problems. In Section 4, we give an overview of argumentation theory and our argumentation framework for determining the mixture proportion is presented in Section 5 followed by the discussion and conclusion in Section 6.

2 Background

We assume a propositional language $\mathcal{L}$ consisting of a *finite*, *nonempty* set of propositional variables (atoms) PS along with the standard logical connectives $\wedge$, $\vee$, $\neg$ and $\rightarrow$. We

*This research is partially funded by ARC grant DP150104133.

[1] When the set of beliefs is not closed, it is customary to call it a belief base.

[2] $K^-_h = K \cap K^*_{\neg h}$, for any belief set K and sentence h.

take the background logic represented by the provability relation $\vdash$ to be the classical logic. Lower case Roman letters like a and b represent sentences of $\mathcal{L}$ and a set of sentences is represented by upper case Roman letters like A and B. Given the finitary nature of $\mathcal{L}$, the information content of an agent's belief state can be represented as a single sentence. We reserve the sentence k for this purpose. Alternatively, K is reserved to represent the information content of an agent's belief state when it is represented as a set of sentences. An *interpretation* or a *possible world* of $\mathcal{L}$ is a function from the set of propositional atoms to $\{0, 1\}$. A model of a sentence x is an interpretation that satisfies x, by $[x]$ we represent the set of models of x and by $|[x]|$ the number of models of x. We also say two sentences x and y are *equivalent*, denoted as $x \equiv y$, if they have exactly the same set of models, i.e $[x] = [y]$. By extension, $A \equiv x$ iff $\bigwedge A \equiv x$. Lower case Greek letters like α and β denote worlds, upper case Greek letters denote sets of worlds and Ω is the set of all worlds. $\omega_\perp \in \Omega$, called the *absurd world*, is a special world where every sentence holds. $\omega_\perp$ allows an agent to have inconsistent beliefs, i.e. it believes in contradictions. The letter P is reserved for representing probability (distribution) functions that assign a non-negative probability mass to each member of Ω such that the total mass of Ω is 1, with the proviso that $P(\omega_\perp) = 1$ or $P(\omega_\perp) = 0$. Note that if $P(\omega_\perp) = 1$, then all other worlds will have zero probability mass.[3] The probability of any sentence $x \in \mathcal{L}$ is the sum of the probability mass assigned by P to its models.

Let P be a probability function over Ω repesenting the belief state. The belief set K (*mutatis mutandis*, k) is the *top* of P, that is, the set of all sentences that have a probability of 1. Given a probability function P and its associated belief set K, a sentence h is said to be consistent with K (and, by extension, consistent with P) if $P(h) > 0$. If $P(h)$ is 1, then the sentence h is a *belief*, if $P(h)$ is 0, it is a *disbelief*, and otherwise it is a *non-belief*. The converse issue of generating a belief state from a belief set is problematic since there can be several probability functions that can be associated with K. This is an important foundational question addressed by [Lindström and Rabinowicz, 1989] [Boutilier, 1995] and [Voorbraak, 1999] among others in the context of probabilistic belief revision. We assume that a unique probability function P has been chosen through some process as the one that best represents K, subject to the condition that it assigns to each model of K except $\omega_\perp$ some *non-zero* probability mass whenever P is *stable* or *maximally stable*, which are defined below:

Definition 1. *Given a belief state P, we say:*

- *P is* stable *iff $P(\omega_\perp) = 0$,*
- *P is* unstable *iff $P(\omega_\perp) = 1$ and,*
- *P is* maximally stable *iff P is stable and for all ω, $P(\omega) > 0$, where $\omega \neq \omega_\perp$.*

Note that all maximally stable belief states are also stable but not all stable belief states are maximally stable. When we refer to a belief state as stable, we mean stable but not maximally stable unless otherwise noted. If P is unstable, $[K] = \{\omega_\perp\}$. Thus contradictions form part of the belief set K and we will say K or P is *inconsistent*. If P is stable, then $[K] \subset \Omega$ and contingent sentences form part of K. If P is maximally stable, then $[K] = \Omega$ and K only consists of tautologies. In this case, P assigns non-zero probability mass to all worlds in Ω except $\omega_\perp$. We will assume that the epistemic agent is not *opinionated*, that is, there is at least one sentence in $\mathcal{L}$ that she neither believes nor disbelieves.[4] This is captured by the assumption that the agent's beliefs allow at least three models, i.e. $|[K]| \geq 3$.[5]

3 Probability Contraction

A *probabilistic contraction function* maps a probability function, under a sentential input, to another probability function. Contraction is the process by which a sentence h that is initially a belief becomes a non-belief. Semantically, if the agent's belief state prior to the contraction is P, then P assigns non-zero probability mass to the h-worlds but not to the $\neg h$-worlds whereas after contraction both h-worlds and $\neg h$-worlds will have non-zero probability mass. We let P^-_h represent the belief state obtained as a result of contracting sentence h from P. [Gärdenfors, 1988] proposed five *basic* postulates $(P^-1)-(P^-5)$ that a rational agent should adhere to when contracting a belief. Motivation behind and justification for these postulates are provided in [Gärdenfors, 1988]. While it is possible to independently define contraction functions, we take revision as a more natural belief change process and define contraction in terms of it. For this work, it suffices to know that the revision of a belief state P by a sentence $\neg h$ results in a new belief state $P^*_{\neg h}$ where $\neg h$ is a belief and h is a disbelief, i.e. $P^*_{\neg h}(h) = 0$ and $P^*_{\neg h}(\neg h) = 1$. Given a probability distribution P and a sentence h, the probability contraction function is defined in terms of revision as follows [Gärdenfors, 1988]:

Definition 2. *Given belief state P, belief h and the revised belief state $P^*_{\neg h}$, the contracted belief state P^-_h is obtained as follows:*

$$P^-_h(x) = \epsilon \cdot P(x) + (1 - \epsilon) \cdot P^*_{\neg h}(x),$$

for all sentence x and where $0 \leq \epsilon \leq 1$.

Often it is simply written: $P^-_h = P\epsilon P^*_{\neg h}$, with $P\epsilon P^*_{\neg h}$ called the ϵ-mixture of P and $P^*_{\neg h}$. By $K^*_{\neg h}$ and K^-_h, we denote the belief sets associated with $P^*_{\neg h}$ and P^-_h respectively.

It is clear from Definition 2 that one needs to only specify ϵ and a probability function $P^*_{\neg h}$ in order to construct a probability contraction function. From a functional point of view, the role of ϵ is clear. It determines how much of the original belief state P to retain. However, in using this definition, three issues immediately surface: **1**) What exactly is ϵ? **2**) What should the value of ϵ be and how does one justify that value? and, **3**) How should $P^*_{\neg h}$ be constructed? As to the

[3] To respect $\sum_{\omega \in \Omega} P(\omega) = 1$ (Kolmogorov's Second Axiom).

[4] An agent with a unstable belief state is opinionated since it believes in every sentence.

[5] It is 3 (not 2) to account for $\omega_\perp$ which is the model of any sentence.

first issue, it has been argued [Gärdenfors, 1988] that ϵ represents *a measure of the degree of closeness to the beliefs in P*. This view is in line with the extreme cases. When $\epsilon = 1$, contraction fails, indicating the agent totally closes up her mind and keeps its belief state unchanged. On the other hand, when $\epsilon = 0$, contraction is reduced to revision by $\neg h$ and the agent is totally open to accepting $P^*_{\neg h}$ as its new belief state. In the non-extreme cases $0 < \epsilon < 1$, the agent opens up its mind partially and is willing to part only with a certain proportion of her belief state. What makes an agent close or open up its mind is the second issue and its answer lies in how ϵ is measured. This will depend on one's interpretation of the notion of closeness. One natural candidate for this potential measure of closeness is epistemic entrenchment and its variants (such as possibility measure or ranking functions). In this paper we consider an alternative approach towards the interpretation and implementation of this measure ϵ. We imagine a dialogic process going on in between the agent and an adversary, the agent defending its belief h and the adversary pushing for $\neg h$. A typical case would be when there is some support for h, but the agent is prepared to open up its mind and consider the adversary's proposal. In this case the agent suspends the belief h without disbelieving it altogether, retains a measure of probability depending on how firmly it was defended in the first place. Hence what we need in addition to the probability measure is an argumentation framework. Regarding $P^*_{\neg h}$, as it is altogether a separate issue, in the rest of the paper we will assume that it is obtained via some suitable mechanism.

4 Argumentation Framework

While Pollock is widely considered to be the founder of formal argumentation systems [Pollock, 1987], most current work on argumentation is based on the abstract argumentation framework of [Dung, 1995] which inspired the framework in [Amgoud and Cayrol, 1998; Amgoud and Prade, 2004] that employs possibility theory. Our work below is based on the latter and on [Hunter, 2013] where probabilistic arguments are considered.

Definition 3. *An* argument *is a pair (H, h) where h is a sentence of $\mathcal{L}$, $H \subseteq \mathcal{L}$ and, a) H is consistent, b) $H \vdash h$, c) there is no proper subset H' of H such that $H' \vdash h$ and, d) $H \neq \{h'\}$, where $h' \equiv h$.*

H is called the *support* of h. First, we need to make the notion of *consistency* clear as in our framework we have the absurd world $\omega_\perp$. We will say H is consistent iff $[H] \neq \{\omega_\perp\}$. Thus, an argument that has a contradiction as support is considered invalid. Definition 3 is essentially the same as those in [Amgoud and Prade, 2004; Hunter, 2013] except that it imposes an extra condition (d). Thus, in our framework, we refrain from arguments whose conclusion and support are logically equivalent. Traditionally arguments of this sort are considered suffering from the fallacy of *petitio principii* [Iacona and Marconi, 2005] – we accuse one of this when we claim that they are begging the question, so to speak.

Let $\mathcal{A}$ denote the set of all arguments that can be generated from $\mathcal{L}$. Defeasibility of arguments is captured by the following definition:

Definition 4. *Let (H_1, h_1) and (H_2, h_2) be two arguments in $\mathcal{A}$. Argument (H_1, h_1) rebuts (H_2, h_2) iff $h_1 \equiv \neg h_2$, and (H_1, h_1) undercuts (H_2, h_2) iff $h \equiv \neg h_1$ for some $h \in H_2$.*

Now we consider the strength of an argument. We assume that the strength of an argument will be sensitive to the joint probability of its premises. This is in line with the approachs taken in [Hunter, 2013; Skyrms, 1986]. In [Pollock, 1987; Prakken and Horty, 2012], the *weakest link* principle, that an argument's strength is the minimum of the strengths of the argument's premises, is endorsed as a way to determine an argument's strength. Unfortunately, in using probabilities to quantify uncertainty, one cannot assume that the probability of the argument is equal to the minimum of the probability of its premises. However, it is known that for two events A and B, $P(A\&B) \leq P(A)$ and $P(A\&B) \leq P(B)$, and in this sense it conforms to the weakest link principle.

Definition 5. *Given a probability distribution P and an argument (H, h) where $H = \{h_1, \ldots, h_n\}$, the strength of the argument is denoted as $level(H, h)$ and is given by the probability of the conjunction of its support $P(h_1 \wedge \ldots \wedge h_n)$.*

Given a belief state P, we denote the set of arguments that can be generated for a conclusion (belief) h as $\mathcal{A}^P_h$. The strength of each argument in $\mathcal{A}^P_h$ can be determined from P as in Definition 5. Next we define how one argument may defend itself from another:

Definition 6. *Let (H_1, h_1) and (H_2, h_2) be two arguments in $\mathcal{A}$. If (H_2, h_2) rebuts or undercuts (H_1, h_1), then (H_1, h_1) defends itself against (H_2, h_2) iff $level(H_1, h_1) \geq level(H_2, h_2)$.*

5 Contraction via Argumentation

Assume there is an agent $\mathcal{X}$ whose belief state is P and amongst other sentences has h as one of its beliefs. Recall that the new belief state P^-_h of $\mathcal{X}$ after h has been removed from its set of beliefs can be computed by the mixture $P\epsilon P^*_{\neg h}$ as given in Definition 2. Our primary goal is to propose a plausible way of determining ϵ. Though $P^*_{\neg h}$ represents the revision of $\mathcal{X}$'s belief state P by $\neg h$, it is convenient for illustrative purposes to think of $P^*_{\neg h}$ as the belief state of another agent $\mathcal{Y}$. We now give the basic idea of the argumentation process and its role in determining ϵ. We imagine that $\mathcal{X}$ has been approached by $\mathcal{Y}$ to give up its belief in h. However, $\mathcal{X}$ is reluctant to do so and an argumentation process ensues. Since h is the sentence of interest, $\mathcal{X}$ equips itself with the set of arguments $\mathcal{A}^P_h$ for h and similarly $\mathcal{Y}$ with $\mathcal{A}^{P^*}_{\neg h}$ for $\neg h$.[6] $\mathcal{X}$ and $\mathcal{Y}$ begin by presenting to each other their best possible arguments. If $\mathcal{X}$ defends itself against $\mathcal{Y}$'s argument and $\mathcal{Y}$ can't do the same, then the situation favors $\mathcal{X}$ and P^-_h should be closer to $\mathcal{X}$'s belief state P. The value assigned to ϵ should then reflect this. Should the situation favor $\mathcal{Y}$, we must again take this into account in determining ϵ. If however, there is a tie, i.e. both $\mathcal{X}$ and $\mathcal{Y}$ defend against each other's arguments, then the argumentation process continues with the agents resorting to comparing the number of arguments that they pos-

[6] We have abused notation here to improve readability and represented $\mathcal{A}^{P^*_{\neg h}}_{\neg h}$ as $\mathcal{A}^{P^*}_{\neg h}$.

sess in order to break the tie. The details of the argumentation process are presented later.

Any sentence considered as a belief by an agent could be one of three types: a *tautology*, a *contradiction* or a *contingent* sentence. The combination of different types of belief states and beliefs give rise to what we call *contraction scenarios* as shown in Table 1. The first step in our work is to identify which scenarios are possible and which are not. This will help us to focus on what kind of belief state and beliefs we should be looking out for. It is worth keeping in mind that the semantic interpretation of revision of P by $\neg h$ means none of the h worlds will have non-zero probability mass assigned by $P^*_{\neg h}$ or in other words, the total probability mass is completely contained within the $\neg h$ worlds.

	P	h	$\neg h$	$P^*_{\neg h}$
(S1)	stable	contingent	contingent	stable
(S2)*	stable	contra.	tautology	stable
(S3)	stable	tautology	contra.	unstable
(M1)*	maximal	contingent	contingent	stable
(M2)*	maximal	contra.	tautology	maximal
(M3)	maximal	tautology	contra.	unstable
(U1)	unstable	contingent	contingent	stable
(U2)	unstable	contra.	tautology	maximal
(U3)*	unstable	tautology	contra.	unstable

Table 1: Table showing different contraction scenarios according to the type of belief state P, belief to be contracted h, its negation $\neg h$, and the revised belief state $P^*_{\neg h}$ that may be used for contraction. * indicates the scenario is either not possible or uninteresting.

In Table 1,[7] h represents a current belief that we desire to remove, and hence $\neg h$ is the sentence by which we must revise the belief state in the process. The cases (S2, M1, M2) and (U3) need some discussion. (S2) is not "possible" since the agent in a stable belief state does not have any contradictory beliefs to remove. Similarly, in (M1) and (M2), an agent in a maximal belief states has only tautological beliefs, and there are no contingent or contradictory beliefs to remove. As to (U3), we stipulate that removal of a tautology from an unstable belief state leaves it unchanged, that is contraction is not successful. The last column describes the intermediate state $P^*_{\neg h}$ that the process is assumed to go through. The five contraction scenarios that interest us are $(S1)$, $(S3)$, $(M3)$, $(U1)$ and $(U2)$. As we will show later, the latter four are limiting cases where ϵ takes on extremal values.

Any belief state and belief in $(S1)$, $(S3)$, $(M3)$, $(U1)$ and $(U2)$ belong to one of five types: $(T1)-(T5)$ that are shown in Table 2. This is useful as it will allows us to learn about the existence and level of arguments in the various contraction scenarios as shown in Propositon 1 below and summarized in Table 2 as well. For instance for $(S3)$, we need to consider only argument sets that can be generated from $(T2)$ and $(T4)$. Furthermore, arguments in $T2$ are of level 0, level 1 and level between 0 and 1 as seen in Table 2.

[7]We recall that by stable, we mean stable but not maximally stable belief states.

Proposition 1. *Given a belief state P and a belief h:*

- *if P is stable and h is contingent or a tautology, then there are arguments of level 0, level 1 and level between 0 and 1 in $\mathcal{A}^P_h$,*
- *if P is maximal and h is a tautology, then there are only arguments of level between 0 and 1 in $\mathcal{A}^P_h$,*
- *if P is unstable and h is a contradiction, then there are no arguments in $\mathcal{A}^P_h$,*
- *If P is unstable and h is contingent or a tautology, then there are only arguments of level 1 in $\mathcal{A}^P_h$.*

	P	h	$[0]$	$(0,1)$	$[1]$
(T1)	stable	contingent	✓	✓	✓
(T2)	stable	tautology	✓	✓	✓
(T3)	max	tautology	×	✓	×
(T4)	unstable	contra.	×	×	×
(T5)	unstable	contingent	×	×	✓

Table 2: Table showing different belief states and the possible arguments levels. The last three columns represent the possible level of arguments where the closed set $[0]$, $[1]$ mean levels 0, 1 respectively and the open set $(0, 1)$ means level between 0 and 1.

Given an agent's argument set $\mathcal{A}^P_h$, ideally we would not want an argument $(H, h) \in \mathcal{A}^P_h$ to be attacked by other arguments in $\mathcal{A}^P_h$. An argument that is attacked should at least be able to defend itself. The next propositions offer us this reassurance.

Proposition 2. *Given a belief state P and a belief h, if:*

- *P is stable and h is contingent (T1) or,*
- *P is stable and h is a tautology (T2, T3) or,*[8]
- *P is unstable and h is a contradiction (T4) or,*
- *P is unstable and h is contingent (T5),*

then there are no rebuts nor undercuts in $\mathcal{A}^P_h$.

We are interested in contraction scenarios $(S1)$, $(S3)$, $(M3)$, $(U1)$ and $(U2)$. In each scenario, there are two argument sets we are interested in: argument set $\mathcal{A}^P_h$ of $\mathcal{X}$ and $\mathcal{A}^{P^*}_{\neg h}$ of $\mathcal{Y}$. An argumentation process will generally involve both rebuttals and undercuts. However, as the following simple theorem shows $\mathcal{A}^P_h$ and $\mathcal{A}^{P^*}_{\neg h}$ do not have any undercutting arguments:

Theorem 1. *For any $(S1)$, $(S3)$, $(M3)$, $(U1)$ or $(U2)$, if (H, h) and $(H', \neg h)$ are arguments in $\mathcal{A}^P_h$ and $\mathcal{A}^{P^*}_{\neg h}$ respectively, then (H, h) and $(H', \neg h)$ only rebut but do not undercut each other.*

Theorem 1 greatly simplifies the argumentation process as $\mathcal{X}$ and $\mathcal{Y}$ no longer have to worry about their arguments being undercut by the other and we use this feature to define when one argument set is preferred over the other simply based on the strength of the arguments that they possess.

Definition 7. *Two arguments are equivalent iff $h \equiv h'$ and $H \equiv H'$.*

[8]Here, stable includes maximally stable belief states as well.

Definition 8. *Given $\mathcal{A}_h^P$, by $\mathcal{A}_h^P/\sim$, we denote the set of equivalence classes of $\mathcal{A}_h^P$ where each class consists of arguments of the same level.*

Definition 9. *Given $\mathcal{A}_h^P/\sim$ and $a \in \mathcal{A}_h^P/\sim$,*

- *$level(a)$ is the level of a and,*
- *$uniq(a)$ is the number of arguments unique upto logical equivalence in a,*
- *c^P is the equivalence class of the highest level not equal to 1,*
- *if c^P exists then $best(\mathcal{A}_h^P/\sim) = level(c^P)$ else $best(\mathcal{A}_h^P/\sim) = 0$.*

Example 1. *Let $PS = \{a, b, c\}$ and $k \equiv a \vee b \rightarrow a \wedge b \wedge \bar{c}$. Thus, $[k] = \{\bar{a}\bar{b}\bar{c}, \bar{a}\bar{b}c, ab\bar{c}\}$. Let $P(\bar{a}\bar{b}\bar{c}) = 0.35, P(\bar{a}\bar{b}c) = 0.4, P(ab\bar{c}) = 0.25$.[9] Let $h = a \leftrightarrow b$. Since $[h] = [k] \cup \{abc\}$ and $P(h) = 1$, h is a belief. Any argument (H, h) such that $[H] \subset [h]$ is a valid argument. However, only if $[H] \cap [k] \neq \emptyset$ will $level(H, h) > 0$. It is clear that for any argument (H, h) with level 1, it must be that $[k] \subset [H]$. In this example, all arguments with level 1 must be equivalent to k, since $[k]$ is the largest proper subset of $[h]$. Thus, if a is the equivalence class in $\mathcal{A}_h^P/\sim$ s.t. level(a) = 1, then $uniq(a) = 1$. The next highest level possible for arguments is 0.75. For any argument (H, h) of level 0.75, it must be that either $[H] = \{\bar{a}\bar{b}\bar{c}, \bar{a}\bar{b}c\}$ or $[H] = \{\bar{a}\bar{b}\bar{c}, \bar{a}\bar{b}c, abc\}$ as these are the only two subsets of $[h]$ whose probabilities sum to 0.75. Let $a' \in \mathcal{A}_h^P/\sim$ s.t. $level(a) = 0.75$, then $c^P = a'$, $uniq(c^P) = 2$ and $best(\mathcal{A}_h^P/\sim) = 0.75$.*

Definition 10. *Given $\mathcal{A}_h^P/\sim$ and $\mathcal{A}_{h'}^{P'}/\sim$:*

> ***Case I**: $\mathcal{A}_h^P$ and $\mathcal{A}_{h'}^{P'}$ are equally preferred ($\mathcal{A}_h^P \approx \mathcal{A}_{h'}^{P'}$) iff $best(\mathcal{A}_h^P/\sim) = best(\mathcal{A}_{h'}^{P'}/\sim)$ and $uniq(c^P) = uniq(c^{P'})$,*
>
> ***Case II**: $\mathcal{A}_{h'}^{P'}$ is strictly preferred to $\mathcal{A}_h^P$ ($\mathcal{A}_h^P \prec \mathcal{A}_{h'}^{P'}$) iff*
>
> > ***a**: $best(\mathcal{A}_h^P/\sim) < best(\mathcal{A}_{h'}^{P'}/\sim)$ OR,*
> >
> > ***b**: $best(\mathcal{A}_h^P/\sim) = best(\mathcal{A}_{h'}^{P'}/\sim)$ and $uniq(c^P) < uniq(c^{P'})$.*

In Definition 9, $best(\mathcal{A}_h^P/\sim)$ returns the highest level amongst all equivalence classes in $\mathcal{A}_h^P/\sim$ that are not of level 1, and 0 if there is no such class. We have adopted the view in Definition 10 that where possible the winner of an argument should be the one who possesses stronger (higher level) arguments (II-a). However, if two agents have arguments of the same strength, then the winner is decided by counting the number of arguments (II-b). Thus, $level(\cdot)$ has precedence over $uniq(\cdot)$. In [Amgoud and Prade, 2004], arguments with a small number of formulas in the support are preferred in order to reduce the chances of being undercut. However, as we do not have undercutting arguments (Theorem 1), we do not differentiate between the arguments in c^P or $c^{P'}$. The reader will note that in the definition of $best(\cdot)$, level 1 classes are ignored. Consider scenario $(S1)$ where both argument sets for $\mathcal{X}$ and $\mathcal{Y}$ are of type $(T1)$ and thus always have level 1 arguments. This means preference over the arguments are determined using Case II-b. It is easily seen then that there is no need to use a probabilistic belief state in the first place as ties may be broken simply based on the number of models of an argument. Thus to exploit probabilities, we refrain from classes of level one. Finally, for the sake of simplicity, we focussed only on one equivalences class in Definition 10 but it is certainly possible to take other classes into consideration.

Determining ϵ

Recall that the contracted belief state P_h^- will be constructed according to the formula $P_h^- = \epsilon \cdot P + (1 - \epsilon) \cdot P_{\neg h}^*$. For each case in Definition 10, we justify and show how ϵ is determined. If the argument sets of $\mathcal{X}$ and $\mathcal{Y}$ are equally preferred, then the proportion of $\mathcal{X}$ and $\mathcal{Y}$'s belief state to be used for computing P_h^- should be equal as shown below.

Case I: If $\mathcal{A}_h^P \approx \mathcal{A}_{\neg h}^{P^*}$ then $\epsilon = 0.5$.

Now if $\mathcal{X}$'s argument set is preferred to $\mathcal{Y}$'s, because its arguments are at a higher level compared to $\mathcal{Y}$'s, then a greater proportion of $\mathcal{X}$'s belief state should be "retained" in P_h^-. This proportion should depend on how much stronger $\mathcal{X}$'s argument is and this is captured in Case II-a below. If $\mathcal{X}$ wins then $\epsilon > 0.5$ and if $\mathcal{X}$ loses, then $\epsilon < 0.5$.

Case II-a: If $\mathcal{A}_h^P \not\approx \mathcal{A}_{\neg h}^{P^*}$, then:

$$\epsilon = \frac{best(\mathcal{A}_h^P/\sim)}{best(\mathcal{A}_h^P/\sim) + best(\mathcal{A}_{\neg h}^{P^*}/\sim)}.$$

The final case is when $\mathcal{X}$'s argument set is preferred to $\mathcal{Y}$'s because it has more arguments than $\mathcal{Y}$ even though their arguments are of the same level. It is tempting to normalize the number of arguments but this is not very desirable. Let $uniq(c^P) = 1$ for $\mathcal{X}$ and $uniq(c^{P^*}) = 2$ for $\mathcal{Y}$. By normalizing $uniq(c^P)$, we get $\epsilon \approx 0.33$. ϵ becomes disproportionately skewed and a large proportion $\mathcal{Y}$'s belief state will be used for P_h^- even though $\mathcal{Y}$ has only one more argument than $\mathcal{X}$. Instead, in such a case, we imagine that a *losing* agent (one with fewer arguments) commits to retaining *at most* a certain proportion of its belief state, represented by α_{max} which must be less than 0.5. At the same time, no matter how many more arguments the winning agent has, the losing agent is determined to retain a certain proportion of its belief state and this is represented by α_{min}, where $0 < \alpha_{min} < \alpha_{max}$. The actual proportion of the losing agent's belief state to be used in the mixture, α, will be between α_{min} and α_{max}, and depend on the difference, *diff*, between the number of arguments of the two agents. The minimum value of *diff* will be 1 at which point we need $\alpha = \alpha_{max}$ and for all other values of *diff*, $\alpha < \alpha_{max}$. A desirable feature for α is for it to not jump too much between small changes of *diff*, i.e. we would like it to decrease slowly towards α_{min} as *diff* becomes bigger. A function for α that has these properties is the slowly decreasing log function shown below:

Case II-b: If $\mathcal{A}_h^P \not\approx \mathcal{A}_{\neg h}^{P^*}$, let:

$$\alpha = \frac{1}{\ln(\mathit{diff} + \delta)} + \alpha_{min}$$

[9] Recall that only k-worlds have non-zero probabilities.

where $\delta = e^{\frac{1}{\alpha_{max}-\alpha_{min}}} - 1$ and $diff = |uniq(c^P) - uniq(c^{P^*})|$.[10]

(i) If $\mathcal{A}^P_h \prec \mathcal{A}^{P^*}_{\neg h}$, then $\epsilon = \alpha$, else

(ii) If $\mathcal{A}^{P^*}_{\neg h} \prec \mathcal{A}^P_h$, then $\epsilon = 1 - \alpha$.

Finally, remember that ϵ is the proportion of $\mathcal{X}$'s belief state to keep. Therefore, if $\mathcal{X}$ is the losing agent (IIb-i) then $\epsilon = \alpha$, whereas if $\mathcal{X}$ wins (IIb-ii) , then $\epsilon = 1 - \alpha$.

Argumentation Scenarios

	$\mathcal{A}^P_h$	$\mathcal{A}^{P*}_{\neg h}$	$best(\mathcal{A}^P_h/\sim)$	$best(\mathcal{A}^{P*}_{\neg h}/\sim)$	ϵ
(S3)	(T2)	(T4)	(0,1)	[0]	1
(M3)	(T3	(T4)	(0,1)	[0]	1
(U1)	(T5)	(T1)	[0]	(0,1)	0
(U2)	(T4)	(T3)	[0]	(0,1)	0

Table 3: Contractions scenarios except $(S1)$ and the value of ϵ.

Table 3 shows the ϵ values of scenarios $(S3)$, $(M3)$, $(U1)$ and $(U2)$. We discuss here only (S3), but the reasoning is similar for the others. (S1) which is not included in the table is discussed separately. In $(S3)$, $\mathcal{X}$'s argument set $\mathcal{A}^P_h$ is of type $(T2)$ consisting of a stable belief state P and a tautological-belief h. $\mathcal{X}$ is challenged by $\mathcal{Y}$, whose belief state $P^*_{\neg h}$ is unstable and pushes for $\mathcal{X}$ to instead accept a contradiction $\neg h$ as a belief. From Table 2, we see that $best(\mathcal{A}^P_h/\sim)$ is some non-zero number less than 1, where as $best(\mathcal{A}^{P*}_{\neg h}/\sim) = 0$ (see Definition 10). Thus, $\mathcal{A}^{P*}_{\neg h} \prec \mathcal{A}^P_h$ (Case II-a) and we get $\epsilon = 1$. If $\mathcal{X}$'s belief state is stable or maximally stable, and is challenged by $\mathcal{Y}$ whose belief state is unstable (as in cases S3 and M3) then $\mathcal{X}$ wins outright and refuses to give up any of its beliefs and P^-_h is the same as P. On the other hand, if $\mathcal{X}$'s belief state is unstable and $\mathcal{Y}$ with a stable or maximally stable belief state challenges it (as in cases U1 and U2), then $\mathcal{Y}$ wins outright and $\mathcal{X}$ is forced to give up all its beliefs. In the latter case, P^-_h is simply $P^*_{\neg h}$. The value of ϵ for $(S1)$ is best illustrated by the Example 2 below. In $(S1)$, P and $P^*_{\neg h}$ are both stable, and beliefs h and $\neg h$ are both contingent.

Example 2 (continued from Example 1). *Recall that* $h = a \leftrightarrow b$, $uniq(c^P) = 2$ *and* $best(\mathcal{A}^P_h/\sim) = 0.75$. *Let* $P^*_{\neg h}$ *be such that* $P^*_{\neg h}(\bar{a}b\bar{c}) = 0.5, P^*_{\neg h}(\bar{a}bc) = 0.3, P^*_{\neg h}(a\bar{b}\bar{c}) = 0.2$. *Note that* $P^*_{\neg h}(h) = 0$. *By similar reasoning to Example 1, we get* $uniq(c^{P^*_{\neg h}}) = 2$ *and* $best(\mathcal{A}^{P^*}_{\neg h}/\sim) = 0.8$. *Since* $best(\mathcal{A}^P_h/\sim) < best(\mathcal{A}^{P^*}_{\neg h}/\sim)$, $\mathcal{A}^P_h \prec \mathcal{A}^{P^*}_{\neg h}$ *(Case II-a), and* $\epsilon \approx \frac{0.75}{0.75+0.8} \approx 0.48$. *Thus,* $P^-_h = 0.48P + 0.52P^*_{\neg h}$. *Consider now* $P^*_{\neg h}(\bar{a}b\bar{c}) = 0.5, P^*_{\neg h}(\bar{a}bc) = 0.25, P^*_{\neg h}(a\bar{b}\bar{c}) = 0.25$, *then* $best(\mathcal{A}^P_h/\sim) = best(\mathcal{A}^{P^*}_{\neg h}/\sim) = 0.75$. *However,* $uniq(c^{P^*_{\neg h}}) = 4 > uniq(c^P)$. *Thus* $\mathcal{A}^P_h \prec \mathcal{A}^{P^*}_{\neg h}$ *(Case II-b). Let* $\alpha_{max} = 0.49$ *and* $\alpha_{min} = 0.1$. *We get* $diff = 2$, $\delta \approx 11.99$ *and* $\alpha \approx 0.48$. *Since* $\mathcal{A}^P_h \prec \mathcal{A}^{P^*}_{\neg h}$, *we set* $\epsilon = \alpha \approx 0.48$ *and* $P^-_h = 0.48P + 0.52P^*_{\neg h}$.

As we can see, $(S1)$ is the only scenario where $0 < \epsilon < 1$ and neither $\mathcal{X}$ nor $\mathcal{Y}$ are willing to give up all of their beliefs. In $\mathcal{X}$'s original belief state P, the maximum level of support for $\neg h$ is 0, where as in P^-_h, it has a maximum level of $1 - \epsilon$.

[10] e is Euler's number. α_{min} is the horizontal asymptote.

6 Discussion and Conclusion

The work in [Hunter, 2013] is closely related to ours and it considers how two agents that have arguments over two different probability distributions, P_i and P_j, may combine their probability distributions to reflect their possibly conflicting views. The combined probability distribution is defined as: $P_{i\oplus j}(A) = max(P_i(A), P_j(A))$, where A is any argument in $\mathcal{A}$ and $P_i(A)$ is the level of the argument (Definition 5). However, it is also shown that $P_{i\oplus j}(A)$ will only be consistent iff P_i and P_j do not *diverge*, which only occurs when $P_i(\omega) = P_j(\omega)$, for all $\omega \in \Omega$. In all the contractions scenarios that we considered in our work, P and $P^*_{\neg h}$ are fully divergent since they do not agree on any models and using $P_{i\oplus j}(A)$ will result in an inconsistent probability distribution which violates the first postulate (P^-1) for probabilistic belief contraction. In [Amgoud and Prade, 2004], the main focus is on two agents that use argumentation to negotiate over goals under a possibilistic logic setting. In relation to belief revision, they define the revision of a belief set by an argument (H, h) as a new belief where the argument is forced to hold, i.e. with a possibility degree of 1. Since our work deals with contraction, we do not fully accept arguments from $\mathcal{A}^{P^*}_{\neg h}$ except in scenarios $(U1)$ and $(U2)$ where $\mathcal{X}$'s initial belief state is unstable. In [Ramachandran *et al.*, 2012; Olsson, 1995] probabilistic belief contraction is also addressed but the rationale behind their approach is based on information theory. They assume that the entropy of the contracted belief state should be higher than original belief state since information is lost and attempt to construct the contracted belief state directly.

When the contraction of a probabilistic belief state is viewed as the mixture of the original belief set and the revised belief state, the question of what mixture proportion should be has been largely ignored. By exploiting the quantitative nature of probabilities and using ideas from argumentation theory, we presented a framework that can determine a credible value for the proportions of the mixture. It is worth noting that a belief state is not simply a probability distribution P, but rather a pair $\langle P, \sigma \rangle$ where σ represents an appropriate similarity measure such as a ranking or a distance function. Although we don't have any explicit similarity measure, we suspect that the argumentation support structure for the input sentence h somehow captures the relevant aspects of a similarity measure and further exploration of this idea will be taken up in future. We plan to look at refining the procedure to determine preferred argument sets by examining more than one equivalence class as done here and to consider argument sets with beliefs other than the belief to be removed as conclusion. Argumentation theory offers an alternative way of accounting for and thinking about probabilistic belief change in general, and this work is a first step in that direction.

References

[Alchourrón *et al.*, 1985] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530, 1985.

[Amgoud and Cayrol, 1998] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In *UAI*, pages 1–7, 1998.

[Amgoud and Prade, 2004] Leila Amgoud and Henri Prade. Reaching agreement through argumentation: A possibilistic approach. In *KR*, pages 175–182, 2004.

[Boutilier, 1995] Craig Boutilier. On the revision of probabilistic belief states. *Notre Dame Journal of Formal Logic*, 36(1):158–183, 1995.

[Dubois *et al.*, 1994] D. Dubois, J. Lang, and H. Prade. Possibilistic logic. In Dov M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming (Vol. 3)*, pages 439–513. Clarendon Press, 1994.

[Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[Gärdenfors, 1988] Peter Gärdenfors. *Knowledge in Flux. Modelling the Dymanics of Epistemic States*. MIT Press, 1988.

[Hunter, 2013] Anthony Hunter. A probabilistic approach to modelling uncertain logical arguments. *Int. J. Approx. Reasoning*, 54(1):47–81, 2013.

[Iacona and Marconi, 2005] Andrea Iacona and Diego Marconi. Petitio principii: What's wrong? *Facta Philosophica*, 7:19–34, 2005.

[Lindström and Rabinowicz, 1989] Sten Lindström and Wlodzimierz Rabinowicz. On probabilistic representation of non-probabilistic belief revision. *Journal of Philosophical Logic*, 18(1):69–101, 1989.

[Nayak, 1994] Abhaya C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

[Olsson, 1995] Erik J Olsson. Can information theory help us remove outdated beliefs? *Lund University Cognitive Studies*, (36), 1995.

[Peppas, 2008] Pavlos Peppas. Belief revision. *Handbook of knowledge representation*, 3:317–359, 2008.

[Pollock, 1987] John L Pollock. Defeasible reasoning. *Cognitive science*, 11(4):481–518, 1987.

[Prakken and Horty, 2012] Henry Prakken and John Horty. An appreciation of john pollock's work on the computational study of argument. *Argument & Computation*, 3(1):1–19, 2012.

[Ramachandran *et al.*, 2012] Raghav Ramachandran, Arthur Ramer, and Abhaya C. Nayak. Probabilistic belief contraction. *Minds and Machines*, 22(4):325–351, 2012.

[Shafer and others, 1976] Glenn Shafer et al. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, 1976.

[Skyrms, 1986] Brian Skyrms. *Choice and chance: An introduction to inductive logic*. Wadsworth Publishing Company, 3rd edition, 1986.

[Spohn, 1988] Wolfgang Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W.L. Harper and B. Skryms, editors, *Causation in Decision, Belief Change, and Statistics, II*, pages 105–134. Kluwer, 1988.

[Voorbraak, 1999] Frans Voorbraak. Probabilistic belief change: Expansion, conditioning and constraining. In *UAI*, pages 655–662, 1999.