Probabilistic Belief Revision via Similarity of Worlds

Gavin Rens[1], Thomas Meyer[1], Gabriele Kern-Isberner[2], and Abhaya Nayak[3]

[1] University of Cape Town, South Africa
{grens,tmeyer}@cs.uct.ac.za
[2] Technical University of Dortmund, Germany
gabriele.kern-isberner@cs.tu-dortmund.de
[3] Macquarie University, Sydney, Australia
abhaya.nayak@mq.edu.au

Abstract. Similarity among worlds plays a pivotal role in providing the semantics for different kinds of belief change. Although similarity is, intuitively, a context-sensitive concept, the accounts of similarity presently proposed are, by and large, context blind. We propose an account of similarity that is context sensitive, and when belief change is concerned, we take it that the evidence (epistemic input) provides the required context. We accordingly develop and examine two accounts of probabilistic belief change that are based on such evidence-sensitive similarity. The first of these switches between two extreme behaviors depending on whether or not the evidence in question is consistent with the current knowledge. The second one gracefully changes its behavior depending on the degree to which the evidence is consistent with current knowledge. Finally, we analyze these two belief change operators with respect to a select set of plausible postulates.

Keywords: Belief revision · Probability · Similarity · Bayesian conditioning · Lewis imaging.

1 Introduction

David Lewis (1976) first proposed *imaging* to analyze conditional reasoning in probabilistic settings, and it has recently been the focus of several works on probabilistic belief change (Ramachandran et al., 2010; Chhogyal et al., 2014; Mishra and Nayak, 2016; Rens et al., 2016). Imaging is the approach of moving the belief in worlds at one moment to similar worlds compatible with evidence received at a next moment.

One of the main benefits of imaging is that it overcomes the problem with Bayesian conditioning, namely, being undefined when evidence is inconsistent with current beliefs. Gärdenfors (1988), Mishra and Nayak (2016) and Rens et al. (2016) proposed generalizations of Lewis's original definition. In this paper we propose a new generalization of imaging – ipso facto a family of imaging-based belief revision operators – and analyze other probabilistic belief revision methods with respect to it. In particular, we propose a version of imaging based on the movement of probability mass weighted by the similarity between possible worlds.

Similarity among worlds plays a pivotal role in accounts of belief change – both probabilistic and non-probabilistic. Intuitively, similarity is a context sensitive notion: Richard is similar to a lion with respect to being brave, not with respect to their food habits, for instance. We take that notion seriously, and propose that the account of similarity among worlds should be sensitive to the evidence.

We define a similarity measure parameterized by evidence. We then define the similarity modulo evidence (SME) operator employing a family of similarity functions. We prove that there is an instantiation of a similarity function for which SME is equivalent to Bayesian conditioning, and we prove that there are versions of SME equivalent to known versions of imaging. SME revision should be viewed as a generalization of probabilistic belief revision.

There are many ways to define the similarity between two stimuli. Shepard (1987) proposed a "universal generalization law" for converting measures of difference/distance to measures of similarity in an appropriately scaled psychological space. Shepard's approach has been widely adopted in cognitive psychology, and biology (concerning perception) (Jäkel et al., 2008; Yearsley et al., 2017). Suppose that the "appropriate scale" is that of probabilities, that is, $[0, 1]$, and that the "psychological space" is the epistemic notion of possible worlds. Shepard's definition of similarity is then easily applied to the possible worlds approach of formal epistemology and seems to fit well into our SME method, which employs the notion of possible worlds. We propose a version of SME based on Shepard's generalization law.

We formalize the fact that Bayesian conditioning (BC) retains probability mass of worlds modeling the evidence, whereas Shepard-based SME revision (SSR) allows the evidence to induce the mass to spread away from old beliefs towards the new evidence (which is arguably undesirable).

Due to both conditioning and SSR having desirable and undesirable properties, we propose two versions of SME revision which combine the two methods in order to maximize their desirable properties. One of the combination SME revision operators *switches* between BC and SSR depending on whether the new evidence is consistent with the current belief state. The other combination operator varies smoothly between BC and SSR depending on the *degree* to which the new evidence is consistent with the current belief state. Both combination operators satisfy three core rationality postulates, but only the switching operator satisfies all six postulates presented.

Due to space limitations, we only provide proof sketches for some of the less intuitive results.

2 Background and Related Work

It could be that the world model (belief state) is wrong, or the evidence observed is wrong. Conventionally, in traditional belief change theory, one takes the observation as primary and certain. Hence, when conflicts occur between the agent's beliefs and what it perceives, the observations dominate the knowledge base. This is the stance we take in this work.

We shall work with a finitely generated classical propositional logic. Let $\mathcal{P} = \{q, r, s, \ldots\}$ be a finite set of *atoms*. Formally, a *world* w is a unique assignment of truth values to all the atoms in $\mathcal{P}$. An agent may consider some non-empty subset $W = \{w_1, w_2, \ldots, w_n\}$ of the possible worlds. Let L be all propositional formulae which can be formed from $\mathcal{P}$ and the logical connectives $\land$ and $\neg$, with $\top$ abbreviating tautology and $\bot$ abbreviating contradiction. Let α be a sentence in L. The classical notion of satisfaction is used. World w satisfies (is a model of) α is written $w \Vdash \alpha$. $Mod(\alpha)$ denotes the set of models of α, that is, $w \in Mod(\alpha)$ iff $w \Vdash \alpha$. We call w an α-world if $w \in Mod(\alpha)$; α entails β (denoted $\alpha \models \beta$) iff $Mod(\alpha) \subseteq Mod(\beta)$; α is equivalent to β (denoted $\alpha \equiv \beta$) iff $Mod(\alpha) = Mod(\beta)$. In this paper, α and β denote evidence, by default.

Often, in the exposition of this paper, a world will be referred to by its truth vector. For instance, if a two-atom vocabulary is placed in order $\langle q, r\rangle$ and $w \Vdash \neg q \land r$, then w may be referred to as 01. We denote the truth assignment of atom q by world w as $w(q)$. For instance, $w(q) = 0$ and $w(r) = 1$.

In this work, the basic semantic element of an agent's beliefs is a probability distribution or a *belief state* $B = \{(w_1, p_1), (w_2, p_2), \ldots, (w_n, p_n)\}$, where p_i is the agents degree of belief (the probability that she assigns to the assertion) that w_i is the actual world, and $\sum_{(w,p)\in B} p = 1$. For parsimony, let $B = \langle p_1, \ldots, p_n\rangle$ be the probabilities that belief state B assigns to $w_1, \ldots, w_n$ where, for instance, $\langle w_1, w_2, w_3, w_4\rangle = \langle 11, 10, 01, 00\rangle$, and $\langle w_1, w_2, \ldots, w_8\rangle = \langle 111, 110, \ldots, 000\rangle$. $B(\alpha)$ abbreviates $\sum_{w\in Mod(\alpha)} B(w)$.

It is not yet universally agreed what belief change means in a probabilistic setting. One school of thought says that probabilistic expansion (restricted revision) is equivalent to Bayesian conditioning.[4] This is evidenced by Bayesian conditioning (BC) being defined only when $B(\alpha) \neq 0$, thus making BC expansion equivalent to BC revision. In other words, one could define expansion to be

$$B^{\mathsf{BC}}_{\alpha} = \{(w, p) \mid w \in W, p = B(w \mid \alpha), B(\alpha) \neq 0\},$$

where $B(w \mid \alpha)$ is defined as $B(\phi_w \land \alpha)/B(\alpha)$ and ϕ_w is a sentence identifying w (i.e., a complete theory for w).[5] Note that $B^{\mathsf{BC}}_{\alpha} = \emptyset$ iff $B(\alpha) = 0$. This implies that BC is ill-defined when $B(\alpha) = 0$.

The technique of *Lewis imaging* for the revision of belief states (Lewis, 1976) requires that for each world $w \in W$ there be a unique 'closest' world $w^{\alpha} \in$

[4] Gärdenfors (1988, Chap. 5) and Voorbraak (1999) mention this, but do not necessarily agree with it.

[5] In general, we write B^{*}_{α} to mean the (the result of) revision of B with α by application of operator $*$.

$Mod(\alpha)$ for given evidence α. If we indicate Lewis's original imaging operation with LI, then his definition can be stated as

$$B_\alpha^{\mathsf{LI}} := \{(w,p) \mid w \in W, p = 0 \text{ if } w \nVdash \alpha, \text{ else } p = \sum_{\{v \in W \mid v^\alpha = w\}} B(v)\},$$

where v^α is the unique closest α-world to v. He calls B_α^{LI} the image of B on α. In words, $B_\alpha^{\mathsf{LI}}(w)$ is zero if w does not model α, but if it does, then w retains all the probability it had and accrues the probability mass from all the non-α-worlds closest to it. This form of imaging only shifts probabilities around; the probabilities in B_α^{LI} sum to 1 without the need for any normalization.

Every world having a unique closest α-world is quite a strong requirement. We now mention an approach which relaxes the uniqueness requirement. Gärdenfors (1988) describes his generalization of Lewis imaging (which he calls *general imaging*) as "... instead of moving all the probability assigned to a world W^i by a probability function P to a unique ("closest") A-world W^j, when imaging on A, one can introduce the weaker requirement that the probability of W^i be distributed among *several* A-worlds (that are "equally close")." Gärdenfors does not provide a constructive method for his approach, but insists that $B_\alpha^{\#}(\alpha) = 1$, where $B_\alpha^{\#}$ is the image of B on α. Rens et al. (2016) introduced *generalized* imaging via a constructive method. It is a particular instance of Gärdenfors' general imaging. Rens et al. (2016) use a pseudo-distance measure between worlds, as defined by Lehmann et al. (2001) and adopted by Chhogyal et al. (2014).

Definition 1. *A pseudo-distance function $d : W \times W \to \mathbb{Z}$ satisfies the following four conditions: for all worlds $w, w', w'' \in W$,*

1. $d(w,w') \geq 0$ *(Non-negativity)*
2. $d(w,w) = 0$ *(Identity)*
3. $d(w,w') = d(w',w)$ *(Symmetry)*
4. $d(w,w'') \leq d(w,w') + d(w',w'')$ *(Triangle Inequality)*

One may also want to impose a condition on a distance function such that any two distinct worlds must have some distance between them: For all $w, w' \in W$, if $w \neq w'$, then $d(w,w') > 0$. This condition is called *Faithfulness*.[6]

Rens et al. (2016) defined $Min(\alpha, w, d)$ to be the set of α-worlds closest to w with respect to pseudo-distance d. Formally,

$$Min(\alpha, w, d) := \{w' \Vdash \alpha \mid \forall w'' \Vdash \alpha, d(w', w) \leq d(w'', w)\},$$

where $d(\cdot)$ is some pseudo-distance function between worlds (e.g., Hamming or Dalal distance). *Generalized imaging* (Rens et al., 2016) (denoted GI) is then defined as

$$\begin{aligned} B_\alpha^{\mathsf{GI}} := \big\{&(w,p) \mid w \in W, p = 0 \text{ if } w \nVdash \alpha, \text{ else} \\ &p = \sum_{\{w' \in W \mid w \in Min(\alpha, w', d)\}} B(w')/|Min(\alpha, w', d)|\big\}. \end{aligned}$$

[6] The term *faithfulness* has been defined differently by different authors. We take the term from Boutilier (1998).

B_α^{GI} is the new belief state produced by taking the generalized image of B on α. In words, the probability mass of non-α-worlds is shifted to their closest α-worlds, such that if a non-α-world $w^\times$ with probability p has n closest α-worlds (equally distant), then each of these closest α-worlds gets p/n mass from $w^\times$.

Recently, Mishra and Nayak (2016) proposed an imaging-based expansion operator $\langle prem^{cl}\rangle$ based on the notion of closeness, where closeness between two worlds is defined as "the gap between the distance between them and the maximum distance possible between any two worlds" (in a neighbourhood of relevance). Formally,

$$B\langle prem^{cl}\rangle R := \{(w,p) \mid w \in W, p = B(w) + \sigma^{cl}(w,S,R)\},$$

where R is the set of non-α-worlds (for some observation α), S is the α-worlds and $\sigma^{cl}(w,S,R)$ is the share of the overall probability salvaged from R going to $w \in S$. To re-iterate, $\langle prem^{cl}\rangle$ is an expansion operator; it does not deal with conflicting evidence.

"The most widely adopted function linking distances and similarities is Shepard's (1987) law of generalization, according to which $\mathit{Similarity} = e^{-\mathit{distance}}$." (Yearsley et al., 2017), where e is Euler's number (≈ 2.71828). (See also, e.g., Jäkel et al. (2008).) Here, *distance* is a term used to refer to the difference in perceived observations (*stimuli* in the jargon of psychology) in an appropriately scaled psychological space. Suppose $\sigma(w,w')$ represents the similarity between worlds w and w'. Then we could define $\sigma(w,w') := e^{-d(w,w')}$. This implies that $d(w,w') = -\ln\sigma(w,w')$.

$$\sigma(w,w'') \geq \sigma(w,w')\cdot\sigma(w',w''). \tag{1}$$

Yearsley et al. (2017) derive (1) from the triangle inequality and call it the multiplicative triangle inequality (MTI).

Imaging falls into the class of probabilistic belief change methods that rely on distance or similarity between worlds. There is another class of methods that rely on definitions of distance or similarity between *distributions* over worlds. The most popular of the latter methods employs the notion of (information theoretic) entropy optimization (Jaynes, 1978; Paris and Vencovská, 1997; Kern-Isberner, 2001). Recently, Beierle et al. (2017) presented a knowledge management system with the core belief change method based on entropy optimization. The present work focuses a method that relies on the notion of similarity between *worlds*.

To further contextualize the present work, we do not consider uncertain evidence (Chan and Darwiche, 2005) nor the general case when instead of a single belief state being known, only a *set* of them is known to hold (Grove and Halpern, 1998; Mork, 2013; Rens et al., 2016). Other related literature worth mentioning is that of Boutilier (1995), Makinson (2011), Chhogyal et al. (2015) and Zhuang et al. (2017). Space limitations prevent us from relating all these approaches to SME revision.

3 Similarity Modulo Evidence (SME)

Let $\sigma : W \times W \to \mathbb{R}$ be a function signature for a family of *similarity* functions. Let σ_α be a sub-family of similarity function, one sub-family for every $\alpha \in L$. Function $\sigma_\alpha(w, w')$ denotes the similarity between worlds w and w' in the context of evidence α. We consider the following set of arguably plausible properties of a similarity function modulo evidence.

For all $w, w', w'', w''' \in W$ and for all $\alpha, \beta \in L$,

1. $\sigma_\alpha(w, w') = \sigma_\alpha(w', w)$ (Symmetry)
2. $0 \leq \sigma_\alpha(w, w') \leq 1$ (Unit Boundedness)
3. $\sigma_\alpha(w, w) = 1$ (Identity)
4. $\sigma_\alpha(w, w'') \geq \sigma_\alpha(w, w') \cdot \sigma_\alpha(w', w'')$ (MTI)
5. If $w, w' \in Mod(\alpha)$ and $w'' \notin Mod(\alpha)$, then $\sigma_\alpha(w, w') > \sigma_\alpha(w, w'')$ (Model Preference)
6. If $w \neq w'$, then $\sigma_\alpha(w, w') < \sigma_\alpha(w, w)$ (Faithfulness)

A property we assume to be satisfied is, if $\alpha \equiv \beta$, then $\sigma_\alpha(w, w') = \sigma_\beta(w, w')$. We now discuss the listed properties.

1. *Symmetry*: Typically, symmetry of similarity is assumed. However, it is not always the case.
2. *Unit Boundedness*: This is a convention to simplify reasoning.
3. *Identity*: Objects are maximally similar to themselves.
4. *Multiplicative Triangle Inequality (MTI)*: Note that even if a similarity function is not symmetric, it could satisfy MTI (and non-symmetric distance functions could satisfy the (additive) triangle inequality). In general, if one suspects that a similarity function is non-symmetric, one would have to check for every combination of orderings of arguments in the inequality (eight such) to ascertain whether MTI holds.
5. *Model Preference*: Any two worlds which agree on a piece of evidence should be more similar to each other than any two worlds, one of which agrees on that evidence and one which does not.
6. *Faithfulness*: It seems intuitive that non-identical worlds should not be maximally similar. It is, however, conceivable that two non-identical worlds cannot be distinguished, given the evidence, in which case they might be deemed (completely) similar.

Definition 2. *Let B be a belief state, α a new piece of information and σ a similarity function. Then the new belief state changed with α via similarity modulo evidence (SME) is defined as*

$$B^{\mathsf{SME}}_\alpha := \big\{(w, p) \mid p = 0 \textit{ if } w \not\Vdash \alpha, \textit{ else } p = \frac{1}{\gamma} \sum_{w' \in W} B(w')\sigma_\alpha(w, w')\big\},$$

where $\gamma := \sum_{w \in W, w \Vdash \alpha} \sum_{w' \in W} B(w')\sigma_\alpha(w, w')$ is a normalizing factor.

We use some identifier ID to identify a similarity function as a particular instantiation σ^{ID}. By $\mathsf{SME}ID$ we mean SME employing σ^{ID}. For any probabilistic belief revision operator $*$, we say that $*$ is SME-compatible *iff there exists a similarity function σ^{ID} such that $B^*_\alpha = B^{\mathsf{SME}ID}_\alpha$ for all B and α.*

An example of revision with SME is provided in Section 4.3.

4 Belief Revision Operations via SME

In this section we investigate various probabilistic belief revision operations simulated or defined as SME operations. We simulate Bayesian conditioning, Lewis imaging and generalized imaging via SME. Finally, we present a new SME-based probabilistic belief revision operation with the similarity function based on Shepard's generalization law.

4.1 Bayesian Conditioning via SME

Bayesian conditioning can be simulated as an SME operator. Let σ^{BC} be defined as follows.

$$\sigma^{BC}_{\alpha}(w, w') := \begin{cases} 1 \textit{ if } w = w' \\ 0 \textit{ otherwise.} \end{cases}$$

Proposition 1. $B^{\mathsf{BC}}_{\alpha} = B^{\mathsf{SME}BC}_{\alpha}$ *iff* $B(\alpha) > 0$. *That is, BC is SME-compatible iff* $B(\alpha) > 0$.

Proof-sketch: σ^{BC}_{α} acts like an indicator function, picking out only α-worlds; non-α-worlds are also picked but are never considered, that is, are assigned zero probability according to the definition of SME. □

Proposition 2. σ^{BC} *satisfies all the similarity function properties, except Model Preference.*

4.2 Imaging via SME

In this sub-section we show that Lewis and generalized imaging are both SME-compatible, and that their corresponding similarity functions satisfy only four of the similarity function properties.

Let $Max(\alpha, w, \sigma)$ be the set of α-worlds most similar to w with respect to similarity function σ. Formally, $Max(\alpha, w, \sigma) := \{w' \in W \mid w' \Vdash \alpha, \forall w'' \Vdash \alpha, \sigma_{\alpha}(w', w) \geq \sigma_{\alpha}(w'', w)\}$.

Lewis imaging can be simulated as an SME operator: Let

$$\sigma^{LI1}_{\alpha}(w, w') := \begin{cases} 1 \text{ if } Max(\alpha, w', \sigma^{L}) = \{w\} \\ 0 \text{ otherwise,} \end{cases}$$

where σ^{L} is defined such that Faithfulness holds and $Max(\alpha, w, \sigma^{L})$ is always a singleton, that is, σ^{L} identifies the unique most similar world to w, for each $w \in W$. Note that due to σ^{L} being faithful, if $w \Vdash \alpha$, then $Max(\alpha, w, \sigma^{L}) = \{w\}$.

Assume $w \neq w'$, $w \Vdash \alpha$ and $w' \nVdash \alpha$. Then $Max(\alpha, w, \sigma^{L}) = \{w\}$, implying that $\sigma^{LI1}_{\alpha}(w, w') = 0$. But it could be that $Max(\alpha, w', \sigma^{L}) = \{w\}$. Then

$\sigma^{LI1}_\alpha(w', w) = 1$. Hence, σ^{LI1} does not satisfy Symmetry. To obtain Symmetry, we define σ^{LI2}. Let

$$\sigma^{LI2}_\alpha(w, w') := \begin{cases} 1 \text{ if } w = w' \\ 1 \text{ if } Max(\alpha, w', \sigma^L) = \{w\} \\ 1 \text{ if } Max(\alpha, w, \sigma^L) = \{w'\} \\ 0 \text{ otherwise.} \end{cases}$$

Proposition 3. *$B^{\mathsf{LI}}_\alpha = B^{\mathsf{SME}\,LI1}_\alpha = B^{\mathsf{SME}\,LI2}_\alpha$. That is, LI is SME-compatible.*

Proof-sketch:

$$B^{\mathsf{LI}}_\alpha(w) = \sum_{\substack{v \in W \\ v = w^\alpha}} B(v) = \sum_{\substack{v \in W \\ Max(\alpha, v, \sigma^L) = \{w\}}} B(v)$$
$$= \sum_{v \in W} B(v)\sigma^{LI1}_\alpha(v, w) = \frac{1}{\gamma} \sum_{v \in W} B(v)\sigma^{LI1}_\alpha(v, w),$$

where $\gamma = 1 = \sum_{w \in W} \sum_{v \in W} B(v)\sigma^{LI1}_\alpha(v, w)$ due to the definition of σ^L.

We then show that $B^{\mathsf{SME}\,LI1}_\alpha = B^{\mathsf{SME}\,LI2}_\alpha$ via the lemma: For all $w \in W$, if $w \Vdash \alpha$, then $\sigma^{LI1}_\alpha(w, w') = \sigma^{LI2}_\alpha(w, w')$. □

Proposition 4. *Of the similarity function properties, σ^{LI2} satisfies only Symmetry, Unit Boundedness, Identity and MTI.*

Generalized imaging can also be simulated as an SME operator: Let

$$\sigma^{GI1}_\alpha(w, w') := \begin{cases} 1 \textit{ if } w \in Min(\alpha, w', d) \\ 0 \textit{ otherwise,} \end{cases}$$

where d is a pseudo-distance function defined to allow multiple worlds sharing the status of being most similar to w', for each $w' \in W$, that is, such that $|Min(\alpha, w', d)|$ may be greater than 1.

For similar reasons as for σ^{LI1}, σ^{GI1} does not satisfy Symmetry. To obtain Symmetry, we define σ^{GI2}. Let

$$\sigma^{GI2}_\alpha(w, w') := \begin{cases} 1 \text{ if } w = w' \\ 1 \text{ if } w \in Min(\alpha, w', d) \\ 1 \text{ if } w' \in Min(\alpha, w, d) \\ 0 \text{ otherwise.} \end{cases}$$

Proposition 5. *$B^{\mathsf{GI}}_\alpha = B^{\mathsf{SME}\,GI1}_\alpha = B^{\mathsf{SME}\,GI2}_\alpha$. That is, GI is SME-compatible.*

Proof-sketch: The proof follows the same pattern as for Proposition 3, just more complicated due to GI being more general than LI. □

Proposition 6. *Of the similarity function properties, σ^{GI2} satisfies only Symmetry, Unit Boundedness, Identity and MTI.*

4.3 A Similarity Function for SME based on Shepard's Generalization Law

We now define a model preferred, Shepard-based similarity function:

$$\sigma^{Sh}_{\alpha}(w, w') := \begin{cases} e^{-d(w,w')} & \text{if } w = w' \text{ or if } w, w' \Vdash \alpha \\ e^{-d(w,w')-d_{max}} & \text{otherwise,} \end{cases}$$

where d is a pseudo-distance function and $d_{max} := \max_{w,w' \in W}\{d(w, w')\}$. Subtracting d_{max} in the second case of the definition of σ^{Sh} is exactly to achieve Model Preference, and the least value to guarantee Model Preference. Note that $\sigma^{Sh}_{\alpha}(w, w') \in (0, 1]$, for all $w, w' \in W$.

Example 1. Quinton knows only three kinds of birds: quails (q), ravens (r) and swallows (s). Quinton thinks Keaton has only a quail and a raven, but he is unsure whether Keaton has a swallow. Quinton's belief state is represented as $B = \{(111, 0.5), (110, 0.5), (101, 0), \ldots , (000, 0)\}$. Now Keaton's sister Cirra tells Quinton that Keaton definitely has no quails, but she has no idea whether Keaton has ravens or swallows. Cirra's information is represented as evidence $\neg q$.

We assume d is Hamming distance. Note that $B^{\mathsf{SME}Sh}_{\neg q}(w) = 0$ for $w \in Mod(q)$ and that $B^{\mathsf{SME}Sh}_{\neg q}(w') = \frac{1}{\gamma}[B(111)\sigma^{Sh}_{\neg q}(w', 111) + B(110)\sigma^{Sh}_{\neg q}(w', 110)]$ for $w' \in Mod(\neg q)$. That is, $B^{\mathsf{SME}Sh}_{\neg q}(w') = \frac{1}{\gamma}0.5[e^{-d(w',111)-d_{max}} + e^{-d(w',110)-d_{max}}] = \frac{0.5}{\gamma}[e^{-d(w',111)-3} + e^{-d(w',110)-3}]$.

For instance, $B^{\mathsf{SME}Sh}_{\neg q}(011) = \frac{0.5}{\gamma}[e^{-1-3}+e^{-2-3}]$ and γ turns out to be 0.0342. Finally, $B^{\mathsf{SME}Sh}_{\neg q}$ is calculated as $\langle 0, 0, 0, 0, 0.365, 0.365, 0.135, 0.135\rangle$. Observe that all $\neg q$-worlds are possible, and that worlds in which Keaton has a raven (but no quail) are more than double as likely than worlds in which Keaton has no raven (and no quail) – due to raven-no-quail-worlds being more similar to Keaton's initially believed worlds than no-raven-no-quail-worlds.

Proposition 7. *Similarity function properties 1 - 4 are satisfied for σ^{Sh}. Model Preference and Faithfulness are satisfied for σ^{Sh} iff d is Faithful.*

Proof-sketch: The most challenging was to prove that σ^{Sh} satisfies MTI. It was tackled with a lemma stating that $e^{-d(w,w'')-x} \geq e^{-d(w,w')-x} \cdot e^{-d(w',w'')-x} \iff d(w, w'') \leq d(w, w') + d(w', w'')$ for $x \geq 0$, and by considering cases where (i) $w = w''$ (ii) $w \neq w''$, with sub-cases (ii.i) $w = w'$ (or $w' = w''$), and (ii.ii) $w \neq w' \neq w''$, with sub-sub-cases (ii.ii.i) exactly one of w, w' or w'' is in $Mod(\alpha)$, (ii.ii.ii) w, w' and w'' are all in $Mod(\alpha)$, and (ii.ii.iii) exactly one of the three worlds is not in $Mod(\alpha)$. □

4.4 Combined Shepard-based and Bayesian SME Operators

Suppose that $B(\alpha) > 0$ and $\beta \models \alpha$. Then we would expect the current belief in β (i.e., $B(\beta)$) not to change due to finding out that α. After all, α tells us nothing new about β; β entails α. We want belief in β to be *retained* when revising by α while $B(\alpha) > 0$ and $\beta \models \alpha$.

Definition 3. *Let $B(\alpha) > 0$ and $\beta \models \alpha$, and let $*$ be a probabilistic belief revision operator. We say that $*$ is* retentive *if $B^*_\alpha(\beta) = B(\beta)$, else we say that $*$ is* inductive.

There might be conditions under which it makes sense for $B^*_\alpha(\beta)$ not to equal $B(\beta)$, for instance, with a belief update operation. In such cases, we presume that an *inductive* process is occurring.

Proposition 8. *$\mathsf{SME}BC$ is retentive and $\mathsf{SME}Sh$ is inductive.*

When $\mathsf{SME}BC$ is defined ($B(\alpha) > 0$), it has the retention property. However, when $B(\alpha) = 0$, an operation other than $\mathsf{SME}BC$ is required. It might, therefore, be desirable to switch between retention and induction. We define an SME revision function which deals with the cases of $B(\alpha) > 0$ and $B(\alpha) = 0$ using $\mathsf{SME}BC$, respectively, $\mathsf{SME}Sh$:

$$B^{\mathsf{SME}Cmb}_\alpha := \begin{cases} B^{\mathsf{SME}BC}_\alpha & \text{if } B(\alpha) > 0 \\ B^{\mathsf{SME}Sh}_\alpha & \text{otherwise.} \end{cases}$$

Switching is arguably a harsh approach due to its discontinuous behavior. Can we gradually trade off between retention and induction? Let $\tau \in [0, 1]$ be the 'degree of retention' desired. Then $\mathsf{SME}BC$ and $\mathsf{SME}Sh$ can be linearly combined as $\mathsf{SME}BCSh$ by defining

$$\sigma^{BCSh}_{\alpha,\tau}(w, w') := \tau \cdot \sigma^{BC}_\alpha(w, w') + (1 - \tau)\sigma^{Sh}_\alpha(w, w').$$

We shall write $\mathsf{SME}BCSh(\tau)$ to mean: $\mathsf{SME}BCSh$ using $\sigma^{BCSh}_{\alpha,\tau}$.

What should τ be? If we use σ^{BC}_α when α is (completely) consistent with B, then we reason that we should use σ^{BC}_α *to the degree* that α is consistent with B. In other words, we set $\tau = B(\alpha)$. We thus instantiate $\sigma^{BCSh}_{\alpha,\tau}$ as

$$\sigma^{\Theta}_\alpha(w, w') := B(\alpha) \cdot \sigma^{BC}_\alpha(w, w') + (1 - B(\alpha)) \cdot \sigma^{Sh}_\alpha(w, w').$$

We analyze $\mathsf{SME}Cmb$ and $\mathsf{SME}\Theta$ with respect to a set of rationality postulates in the next section.

Conjecture 1. Let $x, y \in [0, 1]$ such that $x+y = 1$ and let σ^f and σ^g be similarity functions. If σ^f and σ^g satisfy MTI, then $\sigma^{fg}_{\alpha,\tau}(w, w') := \tau \cdot \sigma^f_\alpha(w, w') + (1 - \tau) \cdot \sigma^g_\alpha(w, w')$ satisfies MTI.

In other words, it is unknown at this stage whether σ^Θ satisfies MTI.

Proposition 9. *Similarity function properties 1 - 3 are satisfied for σ^Θ. (i) Faithfulnes is satisfied for σ^Θ iff d is Faithful and (ii) Model Preference is satisfied for σ^Θ iff d is Faithful and $B(\alpha) < 1$.*

Proof-sketch: We sketch only the proof of case (ii). If $B(\alpha) = 1$, then $\sigma^\Theta = \sigma^{BC}$, implying that Model Preference fails. Recall that if d is Faithful, then σ^{Sh} satisfies Model Preference. If $B(\alpha) < 1$, then $1 - B(\alpha) > 0$, giving σ^{Sh} enough weight in σ^Θ to satisfy Model Preference. □

5 Probabilistic Revision Postulates

First, we discuss the operation called *expansion*, because it is mentioned in the postulates below. Let K be a set of sentences closed under logical consequence. Conventionally, (classical) expansion (denoted $+$) is the logical consequences of $K \cup \{\alpha\}$, where α is new information and K is the current belief set. Or if the current beliefs can be captured as a single sentence β, expansion is defined simply as $\beta + \alpha \equiv \beta \wedge \alpha$. We denote the expansion of belief state B with α as B^{+}_{α}. Furthermore, we shall equate $+$ with Bayesian conditioning (BC). Let $*$ be a probabilistic belief revision operator. Unless stated otherwise, it is assumed that α is logically satisfiable. The probabilistic belief revision postulates are

(P^*1) B^*_α is a belief state
(P^*2) $B^*_\alpha(\alpha) = 1$
(P^*3) If $\alpha \equiv \beta$, then $B^*_\alpha = B^*_\beta$
(P^*4) If $B(\alpha) > 0$, then $B^*_\alpha = B^+_\alpha$
(P^*5) If $B^*_\alpha(\beta) > 0$, then $B^*_{\alpha\wedge\beta} = (B^*_\alpha)^+_\beta$
(P^*6) If $B(\alpha) > 0$ and $\beta \models \alpha$, then $B^*_\alpha(\beta) = B(\beta)$

(P^*1) - (P^*5) are adapted from Gärdenfors (1988) and written in our notation. (P^*6) is a new postulate. We take (P^*1) - (P^*3) to be self explanatory, and to be the three core postulates. (P^*4) is an interpretation of the AGM postulate Alchourrón et al. (1985) which says that if the evidence is consistent with the currently held beliefs, then revision amounts to expansion. (P^*5) says that if β is deemed possible in the belief state revised with α, then expanding the revised belief state with β should be equal to revising the original belief state with the conjunction of α and β. (P^*6) states the requirement for retention (cf. Def. 3) as a rationality postulate.

Proposition 10. $\mathsf{SME}\mathit{Cmb}$ *satisfies (*P^*1*) - (*P^*6*).*

Proof-sketch: The most challenging was the proof that $\mathsf{SME}\mathit{Cmb}$ satisfies (P^*5). The proof depends on the observation that it is known that if $B(\alpha \wedge \beta) > 0$, then $(B^{\mathsf{BC}}_\alpha)^{\mathsf{BC}}_\beta = B^{\mathsf{BC}}_{\alpha\wedge\beta}$ and a lemma stating that if $B^{\mathsf{SME}\mathit{Sh}}_\alpha(\beta) > 0$, then $(B^{\mathsf{SME}\mathit{Sh}}_\alpha)^{\mathsf{SME}BC}_\beta = B^{\mathsf{SME}\mathit{Sh}}_{\alpha\wedge\beta}$. □

Proposition 11. $\mathsf{SME}\Theta$ *satisfies (*P^*1*) - (*P^*3*) but not (*P^*4*) - (*P^*6*).*

Propositions 10 and 11 make the significant difference between $\mathsf{SME}\mathit{Cmb}$ and $\mathsf{SME}\Theta$ obvious.

6 Conclusion

The main contributions of this paper are (i) the definition of SME, a probabilistic belief revision operator derived from Lewis imaging and (ii) $\mathsf{SME}\mathit{Cmb}$ which combines and switches between $\mathsf{SME}\mathit{Sh}$, an instance of SME employing a version of Shepard's generalization law for similarity, and an instance of SME which simulates Bayesian conditioning.

The key mechanism in SME revision is the weighting of world probabilities by the worlds' similarity to the world whose probability is being revised. SME revision was not developed as a competitor to Bayesian Conditioning; nonetheless, SME is more general and with the availability of a similarity function as a weighting mechanism, it allows for tuning of the 'behavior' of revision. SMESh has several advantages over previous operators: It can deal with evidence inconsistent with current beliefs (other imaging methods also have this property), and it is more general than Lewis's original imaging and generalized imaging. Furthermore, σ^{Sh} satisfies most properties one might expect from a similarity measure, notably the multiplicative triangle inequality and model preference. Finally, SMECmb satisfies all the rationality postulates for probabilistic revision investigated in this study.

We have defined notions of retention and induction for probabilistic belief change operators, but we did not say which is preferred. A combined belief revision approach was proposed, which allows the user or agent to choose the degree of retention/induction. We proposed that the trade-off factor be $B(\alpha)$, the degree to which evidence α is consistent with current beliefs B. We saw, however, that the three non-core rationality postulates are not satisfied. Nonetheless, the idea of trading off between SMEBC and SMESh via $B(\alpha)$ seems intuitively appealing. But what is the effect of retention versus induction and when is one more appropriate than the other?

Suppose evidence β contains no new information with respect to evidence α, (i.e. $\alpha \models \beta$). Should two worlds then be at least as similar to each other in the context of β as they are in the context of α? If the question is answered in the positive, we call the applicable similarity function *weakly monotonic*. The strict version of the monotonicity property is: If $\alpha \models \beta$, $\alpha \not\equiv \beta$ and $w \neq w'$, then $\sigma_\alpha(w, w') < \sigma_\beta(w, w')$. It is unclear whether either of the monotonicity properties ought to be satisfied, but it seems like a question meriting further investigation. At present we can show that weak monotonicity holds for σ^{BC}, σ^{Sh} and σ^θ, and that strict monotonicity fails for all our definitions of σ.

Our view is that when it comes to probabilistic revision, (P^*4) - (P^*6) might be too strong. Perhaps they should be weakened just enough to accommodate SMEΘ. A theorem states that a particular set of rationality postulates identify, characterize or represent a (class of) belief change operator(s), and that the (class of) operator(s) satisfies all the postulates. In general, it would be nice if we could make general statements about the relations between the revision postulates and the similarity properties. This is left for future work. We acknowledge that representation theorems are desirable, but consider them as a second step after clarifying what properties are adequate for a novel belief revision operator in general. We consider our paper as a first step of presenting and elaborating on a completely novel type of revision operator. The shown relationships to well-known revision operators prove its basic foundation in established traditions of belief change theory.

Bibliography

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

Beierle, C., Finthammer, M., Potyka, N., Varghese, J., and Kern-Isberner, G. (2017). A framework for versatile knowledge and belief management. *IF-CoLog Journal of Logics and Their Applications*, 4(7):2063–2095.

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

Boutilier, C. (1998). A unified model of qualitative belief change: A dynamical systems perspective. *Artif. Intell.*, 98(1–2):281–316.

Chan, H. and Darwiche, A. (2005). On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163:67–90.

Chhogyal, K., Nayak, A., and Sattar, A. (2015). Probabilistic belief contraction: Considerations on epistemic entrenchment, probability mixtures and kl divergence. In Pfahringer, B. and Renz, J., editors, *Proceedings of the Australasian Joint Conference on Artificial Intelligence*, volume 9457 of *LNAI*, pages 109–122. Springer Intl. Publishing, Switzerland.

Chhogyal, K., Nayak, A., Schwitter, R., and Sattar, A. (2014). Probabilistic belief revision via imaging. In Pham, D.-N. and Park, S.-B., editors, *PRICAI 2014: Trends in Artif. Intell.: Thirteenth Pacific Rim Intl. Conf. on Artif. Intell.*, volume 8862 of *LNAI*, pages 694–707. Springer Intl. Publishing, Switzerland.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Massachusetts/England.

Grove, A. and Halpern, J. (1998). Updating sets of probabilities. In *Proceedings of the Fourteenth Conf. on Uncertainty in Artif. Intell.*, UAI'98, pages 173–182, San Francisco, CA, USA. Morgan Kaufmann.

Jäkel, F., Schölkopf, B., and Wichmann, F. (2008). Similarity, kernels, and the triangle inequality. *Journal of Mathematical Psychology*, 52(5):297–303.

Jaynes, E. (1978). Where do we stand on maximum entropy? In *The Maximum Entropy Formalism*, pages 15–118. MIT Press.

Kern-Isberner, G. (2001). Revising and updating probabilistic beliefs. In Williams, M.-A. and Rott, H., editors, *Frontiers in Belief Revision*, volume 22 of *Applied Logic Series*, pages 393–408. Kluwer Academic Publishers, Springer Netherlands.

Lehmann, D., Magidor, M., and Schlechta, K. (2001). Distance semantics for belief revision. *Journal of Symboloc Logic*, 66(1):295–317.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85(3):297–315.

Makinson, D. (2011). Conditional probability in the light of qualitative belief change. *Philosophical Logic*, 40(2):121–153.

Mishra, S. and Nayak, A. (2016). Causal basis for probabilistic belief change: Distance vs. closeness. In *Proceedings of the Tenth Multi-disciplinary International Workshop on Artificial Intelligence (MIWAI)*, LNAI, pages 112–125. Springer International Publishing AG.

Mork, J. C. (2013). Uncertainty, credal sets and second order probability. *Synthese*, 190(3):353–378.

Paris, J. and Vencovská, A. (1997). In defense of the maximum entropy inference process. *Intl. Journal of Approximate Reasoning*, 17(1):77–103.

Ramachandran, R., Nayak, A., and Orgun, M. (2010). Belief erasure using partial imaging. In Li, J., editor, *AI 2010: Advances in Artif. Intell.: Proceedings of the Twenty-third Australasian Joint Conf.*, volume 6464 of *LNAI*, pages 52–61, Berlin/Heidelberg. Springer Verlag.

Rens, G., Meyer, T., and Casini, G. (2016). On revision of partially specified convex probabilistic belief bases. In M. Fox, G. K., editor, *Proceedings of the Twenty-second European Conference on Artificial Intelligence (ECAI-2016)*, pages 921–929. IOS Press.

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323.

Voorbraak, F. (1999). Probabilistic belief change: Expansion, conditioning and constraining. In *Proceedings of the Fifteenth Conf. on Uncertainty in Artif. Intell.*, UAI'99, pages 655–662, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Yearsley, J. M., Barque-Duran, A., Scerrati, E., Hampton, J. A., and Pothos, E. M. (2017). The triangle inequality constraint in similarity judgments. *Progress in Biophysics and Molecular Biology*, 130(Part A):26 – 32. Quantum information models in biology: from molecular biology to cognition.

Zhuang, Z., Delgrande, J., Nayak, A., and Sattar, A. (2017). A unifying framework for probabilistic belief revision. In Bacchus, F., editor, *Proceedings of the Twenty-fifth Intl. Joint Conf. on Artif. Intell. (IJCAI-17)*, Menlo Park, CA. AAAI Press. To appear.