Belief Revision Operators with Varying Attitudes Towards Initial Beliefs*

Adrian Haret†   Stefan Woltran

Abstract

Classical axiomatizations of belief revision include a postulate stating that if new information is consistent with initial beliefs, then revision amounts to simply adding the new information to the original knowledge base. This postulate assumes a conservative attitude towards initial beliefs, in the sense that an agent faced with the need to revise them will seek to preserve initial beliefs as much as possible. In this work we look at operators that can assume different attitudes towards original beliefs. We provide axiomatizations of these operators by varying the aforementioned postulate and obtain representation results that characterize the new types of operators using preorders on possible worlds. We also present concrete examples for each new type of operator, adapting notions from decision theory.

1 Introduction

Belief revision models rational changes of an agent's epistemic state triggered by the availability of new, trusted information. In the standard logical approach, an agent's epistemic state is represented by propositional formulas, while the standards of rationality a revision operator is expected to abide by are encoded as logical axioms [1, 12, 17, 9]. Notably, the classical set of revision postulates turn out to define a class of operators that can be looked at in two ways: on the one hand as change, guided by logical postulates, of propositional theories in response to new data; and on the other hand as choice functions over possible worlds exploiting plausibility rankings over such interpretations. This correspondence tells us that an agent faced with revision of its initial beliefs acts as if it chooses from a set of feasible possible worlds the ones it considers most plausible.

A distinguishing feature of revision operators, as typically axiomatized, is that they can be assumed to adopt a particular attitude towards initial beliefs, enforced through what are called the *Inclusion* and *Vacuity* postulates in the AGM formulation [9], or through a single postulate equivalent to their conjunction in the KM axiomatization [17]. This attitude articulates the policy by which the agent's prior information behaves with respect to new data: thus, in the KM axiomatization, the postulate in question states that if new information μ is consistent with existing beliefs κ, then the

---

result of revision is simply $\kappa \wedge \mu$. In other words, the agent retains its initial beliefs and simply supplements them with the new item of information, if it can do so in a consistent way. This is in line with a view of revision where the information κ with which the agent starts off represents the possible worlds the agent finds most plausible, information not to be given up unless challenged by conflicting new data, and spells out a conservative attitude towards initial beliefs, guided by the desire to preserve them as much as possible.

In the current work we view such a conservative attitude as one among many that an agent can have towards its initial beliefs. By varying the postulate responsible for enforcing this attitude we are able to axiomatize revision operators that embody a wider range of attitudes towards prior information, and characterize these operators in terms of the types of preorders they induce on the set of possible worlds. To illustrate these principles we provide concrete operators, constructed using two ingredients: a notion of *distance* between possible worlds and an *aggregation function* that ranks possible worlds depending on the initial beliefs. We also show, in each case, how these operators fit into the landscape of new postulates introduced. Without the theoretical apparatus of the new postulates, the concrete operators put forward would be merely classified as deviant, since they do not satisfy the traditional blend of *Inclusion* and *Vacuity*. But through the present analysis they can be viewed as encoding distinct and characterizable stances an agent can take towards its beliefs.

Alternatives to the classical revision postulates have been considered since as far back as the original publications in the field [12, 16, 14]. However, we believe that a systematic analysis of the intuition underlying the KM postulate corresponding to *Inclusion* and *Vacuity*, as we perform here, sheds new light on familiar topics.

2 Preliminaries

We assume a finite set $\mathcal{P}$ of propositional atoms, from which the set Prop of propositional formulas is generated using the usual propositional connectives. A propositional knowledge base κ is a finite set of propositional formulas, which we typically identify with the conjunction of its formulas $\bigwedge_{\mu \in \kappa} \mu$. The set of all propositional knowledge bases is 2^{Prop}. The universe $\mathcal{U}$ is the set of all possible interpretations (also called *possible worlds*) for formulas in Prop. The models of a propositional formula μ are the interpretations which satisfy it, and we write $[\mu]$ (respectively, $[\kappa]$) for the set of models of μ (respectively, for $\bigcap_{\mu \in \kappa}[\mu]$). If there is no danger of ambiguity, we write models as words where the letters are the atoms assigned to true, e.g., $\{\{a, b\}, \{b, c\}\}$ is written as $\{ab, bc\}$. For $\mu_1, \mu_2 \in \mathsf{Prop}$, we say that μ_1 entails μ_2, written $\mu_1 \models \mu_2$, if $[\mu_1] \subseteq [\mu_2]$, and that they are equivalent, written $\mu_1 \equiv \mu_2$, if $[\mu_1] = [\mu_2]$. A formula μ (a knowledge base κ) is consistent if $[\mu] \neq \emptyset$ ($[\kappa] \neq \emptyset$), and *complete* if it has exactly one model. The set of consistent knowledge bases is $2^{\mathsf{Prop}}_{\mathsf{cons}}$. If κ is a propositional knowledge base, then *the dual $\overline{\kappa}$ of κ* is obtained by replacing every literal appearing in κ with its negation. If w is an interpretation, the dual interpretation $\overline{w}$ is $\mathcal{P} \setminus w$. If W is a set of interpretations, its dual $\overline{W}$ is defined as $\{\overline{w} \mid w \in W\}$.

Example 1. If $\mathcal{P} = \{a, b, c\}$ and $\kappa = \{a, a \to b\}$, then $\overline{\kappa} = \{\neg a, \neg a \to \neg b\}$. We have that $[\kappa] = \{ab, abc\}$, the dual of the interpretation ab is $\overline{ab} = c$ and $\overline{[\kappa]} = \{c, \emptyset\}$.

In Example 1 we obtain that $\overline{[\kappa]} = [\overline{\kappa}]$. Though we do not provide a formal proof, we mention here that this holds more generally, i.e., for any $\kappa \in 2^{\mathsf{Prop}}$, it holds that $\overline{[\kappa]} = [\overline{\kappa}]$.

If $\mathcal{M}$ is a set, then $Bin(\mathcal{M})$ is the set of binary relations on $\mathcal{M}$. We write $<$ for the strict part of $\leq$, i.e., $x < x'$ if $x \leq x'$ and $x' \not\leq x$; moreover, $x \approx x'$ if $x \leq x'$ and $x' \leq x$. The $\leq$-*minimal elements of $\mathcal{M}$ with respect to* $\leq$ are defined as $\min_{\leq} \mathcal{M} = \{x \in \mathcal{M} \mid \nexists x' \in \mathcal{M} \text{ such that } x' < x\}$. An *assignment from $\mathcal{M}_1$ to $\mathcal{M}_2$* is a function $\alpha\colon \mathcal{M}_1 \to Bin(\mathcal{M}_2)$. We write $\leq_\kappa$ instead of $\alpha(\kappa)$ if there is no danger of ambiguity. If W is a set of interpretations, we denote by φ_W a propositional formula such that $[\varphi_W] = W$.

3 Revision: Axioms and Characterizations

A propositional revision operator is a function $\circ\colon 2^{\mathsf{Prop}}_{\mathsf{cons}} \times \mathsf{Prop} \to \mathsf{Prop}$. The intention is that $\kappa \circ \mu$ encodes changes brought to existing held beliefs κ such that new, trusted information μ is accepted. A sensible revision operator is expected to resolve inconsistencies between κ and μ and to satisfy certain rationality criteria, presented below.

3.1 Basic Postulates

If $\kappa, \kappa_1, \kappa_2 \in 2^{\mathsf{Prop}}_{\mathsf{cons}}$ and $\mu, \mu_1, \mu_2 \in \mathsf{Prop}$, we first single out a core set of axioms. Note that R_2 as we write it does not coincide with the second KM postulate. The second KM postulate shows up in the latter part of this section.

(R$_1$) $\kappa \circ \mu \models \mu$.

(R$_2$) If μ is consistent, then $\kappa \circ \mu$ is consistent.

(R$_3$) If $\kappa_1 \equiv \kappa_2$ and $\mu_1 \equiv \mu_2$, then $\kappa_1 \circ \mu_1 \equiv \kappa_2 \circ \mu_2$.

(R$_4$) $(\kappa \circ \mu_1) \wedge \mu_2 \models \kappa \circ (\mu_1 \wedge \mu_2)$.

(R$_5$) If $(\kappa \circ \mu_1) \wedge \mu_2$ is cons., then $\kappa \circ (\mu_1 \wedge \mu_2) \models (\kappa \circ \mu_1) \wedge \mu_2$.

Postulates R_{1-5} are part of the standard set of KM postulates [17], saying that a revision operator incorporates new information μ (R_1), returns a consistent output if μ is consistent (R_2), performs its task irrespectively of how beliefs are written down (R_3) and satisfies some coherence constraints when the revision formula is varied (R_{4-5}). A revision operator is *basic* if it satisfies R_{1-5}. Throughout the paper, we work only with basic operators.

3.2 Basic Assignments

Reflection on postulates R_{1-5} reveals that an operator $\circ$ satisfying them chooses among models of μ and, in doing so, behaves as if it has preferences over units of information. Formally, this is cashed out by assigning to each consistent knowledge base κ in $2_{\text{cons}}^{\text{Prop}}$ a binary relation $\leq_\kappa$ on interpretations in $\mathcal{U}$: to revise κ by μ, then, becomes equivalent to choosing the best models of μ in $\leq_\kappa$. And, in the same way that revision operators are expected to satisfy a set of basic properties (postulates R_{1-5}), rankings on $\mathcal{U}$ must satisfy a set of properties, to be introduced in the following, that are conducive to rational choice.

For an assignment $\alpha\colon 2_{\text{cons}}^{\text{Prop}} \to Bin(\mathcal{U})$, a revision operator $\circ$ and $\kappa, \kappa_1, \kappa_2 \in 2_{\text{cons}}^{\text{Prop}}$, $w_1, w_2 \in \mathcal{U}$, we first consider the following properties:

(o_1) $\leq_\kappa$ is reflexive.

(o_2) $\leq_\kappa$ is transitive.

(o_3) If $\kappa_1 \equiv \kappa_2$, then $\leq_{\kappa_1} = \leq_{\kappa_2}$.

(o_4) $\leq_\kappa$ is total.

(o_5) $[\kappa \circ \mu] = \min_{\leq_\kappa}[\mu]$, for any propositional formula μ.

An assignment is *basic* if it satisfies properties o_{1-4}. Notice that properties o_{1-2} imply that $\leq_\kappa$ is a preorder on $\mathcal{U}$. Adding property o_4 makes $\leq_\kappa$ total, and o_3 adds an independence of syntax aspect to the assignment. If, on top of this, $\leq_\kappa$ satisfies o_5, we say that the assignment α *represents the revision operator* $\circ$ (and that $\circ$ *is represented by* α). The overloading of the term 'basic' is intentional: as Theorem 1 shows, properties o_{1-5} define a class of rankings on interpretations that fully characterize revision operators satisfying axioms R_{1-5}.

Theorem 1. A revision operator satisfies postulates R_{1-5} iff there exists an assignment $\alpha\colon 2_{\text{cons}}^{\text{Prop}} \to \mathcal{U}$ representing it such that, for any $\kappa \in 2_{\text{cons}}^{\text{Prop}}$, $\leq_\kappa$ satisfies properties o_{1-5}.

Proof. ("$\Rightarrow$") We first assume we are given an assignment $\alpha\colon 2_{\text{cons}}^{\text{Prop}} \to \mathcal{U}$ such that, for any $\kappa \in 2_{\text{cons}}^{\text{Prop}}$, $\leq_\kappa$ satisfies properties o_{1-5}. We want to show that the revision operator $\circ$ represented by it satisfies axioms R_{1-5}. Concretely, we consider the revision operator $\circ$ such that, for any knowledge base κ and formula μ, we have that:

$$[\kappa \circ \mu] = \min_{\leq_\kappa}[\mu].$$

The proof comes to checking, one by one, that axioms R_{1-5} are satisfied by operator $\circ$. The argument is not new, and can be extracted from classical works on the subject [17], but we reproduce it here for completeness.

Axiom R_1 follows from the fact that $\kappa \circ \mu$ is a formula whose set of models is, by definition, a subset of $[\mu]$. Since $[\mu]$ is a finite set and, by properties o_{1-2}, $\leq_\kappa$ is a preorder, we then have that $\min_{\leq_\kappa}[\mu] \neq \emptyset$, if $[\mu] \neq \emptyset$. This implies that axiom R_2 is satisfied. For axiom R_3, we have that if $\kappa_1 \equiv \kappa_2$, then, by property o_3, $\leq_{\kappa_1} = \leq_{\kappa_2}$. Clearly, then, if we also have that $\mu_1 \equiv \mu_2$, then it holds that $\min_{\leq_{\kappa_1}}[\mu_1] = \min_{\leq_{\kappa_2}}[\mu_2]$.

4

For axiom R_4, take $w_1 \in [(\kappa \circ \mu_1) \wedge \mu_2]$: this means that $w_1 \in \min_{\leq_\kappa}[\mu_1] \cap [\mu_2]$, and we want to show that $w_1 \in \min_{\leq_\kappa}[\mu_1 \wedge \mu_2]$. Suppose, on the contrary, that $w_1 \notin \min_{\leq_\kappa}[\mu_1 \wedge \mu_2]$. Since we can derive, from our starting assumption, that $w_1 \in [\mu_1 \wedge \mu_2]$, it follows that $[\mu_1 \wedge \mu_2] \neq \emptyset$, and hence that $\min_{\leq_\kappa}[\mu_1 \wedge \mu_2] \neq \emptyset$. Thus there exists $w_2 \in \min_{\leq_\kappa}[\mu_1 \wedge \mu_2]$; since $w_1 \notin \min_{\leq_\kappa}[\mu_1 \wedge \mu_2]$ we then conclude that $w_2 <_\kappa w_1$. But w_1 and w_2 are both in $[\mu_1]$ and $w_1 \in \min_{\leq_\kappa}[\mu_1]$, which implies that $w_1 \leq_\kappa w_2$. We have arrived at a contradiction, and thus $w_1 \in \min_{\leq_\kappa}[\mu_1 \wedge \mu_2]$. For axiom R_5, take $w_1 \in \min_{\leq_\kappa}[\kappa \circ (\mu_1 \wedge \mu_2)]$. We want to show that $w_1 \in [(\kappa \circ \mu_1) \wedge \mu_2]$. From the fact that $w_1 \in \min_{\leq_\kappa}[\kappa \circ (\mu_1 \wedge \mu_2)]$ we infer that $w_1 \in [\mu_2]$, so all we have to show is that $w_1 \in [\kappa \circ \mu_1]$. Suppose, on the contrary, that $w_1 \notin [\kappa \circ \mu_1]$: we now use the assumption that $(\kappa \circ \mu_1) \wedge \mu_2$ is consistent to conclude that there exists $w_2 \in [(\kappa \circ \mu_1) \wedge \mu_2]$, which implies that $w_2 <_\kappa w_1$. But, since $w_1 \in \min_{\leq_\kappa}[\kappa \circ (\mu_1 \wedge \mu_2)]$ and $w_2 \in [\mu_1 \wedge \mu_2]$, it also follows that $w_1 \leq_\kappa w_2$. This is a contradiction, and we conclude that $w_1 \in [\kappa \circ \mu_1]$.

("$\Leftarrow$") For this direction, we must first fix a point on notation: given a set W of interpretations, we will need to make use of a propositional formula that captures exactly the set W. More precisely, if $w_1, \ldots, w_n$ are interpretations, we write $\varphi_{1,\ldots,n}$ for a propositional formula such that $[\varphi_{1,\ldots,n}] = \{w_1, \ldots, w_n\}$.[1]

Assume, now, that we are given a revision operator $\circ$ that satisfies axioms R_{1-5}. We define as assignment $\alpha \colon 2^{\text{Prop}}_{\text{cons}} \to \mathcal{U}$ by saying, for any base κ and interpretations w_1 and w_2 that:

$$w_1 \leq_\kappa w_2 \text{ if } w_1 \in [\kappa \circ \varphi_{1,2}].$$

We show next that $\leq_\kappa$ satisfies properties o_{1-5}.

For property o_1 (reflexivity), notice that, using axioms R_1 and R_3, we can conclude that $[\kappa \circ \varphi_1] = \{w_1\}$, where φ_1 is a propositional formula such that $[\varphi_1] = \{w_1\}$. This implies that $w_1 \leq_\kappa w_1$.

For property o_2 (transitivity), assume there are interpretations w_1, w_2 and w_3 such that $w_1 \leq_\kappa w_2$ and $w_2 \leq_\kappa w_3$. We want to show that $w_1 \leq_\kappa w_3$. We will do this in two steps. The first step consists in showing that $w_1 \in [\kappa \circ \varphi_{1,2,3}]$, where $\varphi_{1,2,3}$ is a propositional formula such that $[\varphi_{1,2,3}] = \{w_1, w_2, w_3\}$. First, notice that, by axioms R_1 and R_2, we have that $\emptyset \subset [\kappa \circ \varphi_{1,2,3}] \subseteq [\varphi_{1,2,3}]$. In other words, $[\kappa \circ \varphi_{1,2,3}]$ contains at least one of the interpretations w_1, w_2 and w_3. We will do a case analysis to show that $w_1 \in [\kappa \circ \varphi_{1,2,3}]$.

Case 1. If $w_1 \in [\kappa \circ \varphi_{1,2,3}]$, the conclusion is immediate.

Case 2. If $w_2 \in [\kappa \circ \varphi_{1,2,3}]$, then $(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,2}$ is consistent. Using axioms R_{4-5}, this implies that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,2} \equiv \kappa \circ (\varphi_{1,2,3} \wedge \varphi_{1,2}).$$

We have that $\varphi_{1,2,3} \wedge \varphi_{1,2} \equiv \varphi_{1,2}$, and thus, by axiom R_3, it holds that $\kappa \circ (\varphi_{1,2,3} \wedge \varphi_{1,2}) \equiv \kappa \circ \varphi_{1,2}$. We can thus conclude that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,2} \equiv \kappa \circ \varphi_{1,2}.$$

Since $w_1 \in [\kappa \circ \varphi_{1,2}]$ (because $w_1 \leq_\kappa w_2$), we have that $w_1 \in [\kappa \circ \varphi_{1,2,3}]$.

---

[1] Such a formula is guaranteed to exist in propositional logic: we can, for instance, take the DNF representation of the set $\{w_1, \ldots, w_n\}$.

Case 3. If $w_3 \in [\kappa \circ \varphi_{1,2,3}]$, $(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{2,3}$ is consistent. Using axioms R_{4-5}, this implies that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{2,3} \equiv \kappa \circ (\varphi_{1,2,3} \wedge \varphi_{2,3}),$$

and by axioms R_{1-2} again, we have that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{2,3} \equiv \kappa \circ \varphi_{2,3}.$$

Since $w_2 \in [\kappa \circ \varphi_{2,3}]$ (because $w_2 \leq_\kappa w_3$), we get that $w_2 \in [\kappa \circ \varphi_{1,2,3}]$. We can now reproduce the reasoning from Case 2 to conclude that $w_1 \in [\kappa \circ \varphi_{1,2,3}]$.

Once we have that $w_1 \in [\kappa \circ \varphi_{1,2,3}]$, it follows that $(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,3}$ is consistent. So, by axioms R_{4-5}, we have that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,3} \equiv \kappa \circ (\varphi_{1,2,3} \wedge \varphi_{1,3}),$$

and by axioms R_{1-2}, we have that:

$$(\kappa \circ \varphi_{1,2,3}) \wedge \varphi_{1,3} \equiv \kappa \circ \varphi_{1,3}.$$

Since $w_1 \in [\kappa \circ \varphi_{1,2,3}]$ and $w_1 \in [\varphi_{1,3}]$, we conclude that $w_1 \in [\kappa \circ \varphi_{1,3}]$. This implies that $w_1 \leq_\kappa w_3$.

Property o_3 follows from the fact that the definition of $\leq_\kappa$ does not depend in any way on the syntax of κ.

Property o_4 follows by axioms R_{1-2}: since, for any two interpretations w_1 and w_2, we get that $\emptyset \subset [\kappa \circ \varphi_{1,2}] \subseteq [\varphi_{1,2}]$, we have that $w_1 \leq_\kappa w_2$ or $w_2 \leq_\kappa w_1$, or both.

For property o_5, we show the double inclusion.

("$\subseteq$") Take, first, $w_1 \in [\kappa \circ \mu]$, and some arbitrary interpretation $w_2 \in [\mu]$. We get that $w_1 \in [(\kappa \circ \mu) \wedge \varphi_{1,2}]$ and, by axioms R_4 and R_3, it follows that $w_1 \in [\kappa \circ (\mu \wedge \varphi_{1,2})]$ and then that $w_1 \in [\kappa \circ \varphi_{1,2}]$. Thus, $w_1 \leq_\kappa w_2$ and hence, $w_1 \in \min_{\leq_\kappa}[\mu]$.

("$\supseteq$") Take, now, $w_1 \in \min_{\leq_\kappa}[\mu]$. We want to show that $w_1 \in [\kappa \circ \mu]$. Suppose, on the contrary, that $w_1 \notin [\kappa \circ \mu]$. Since, due to our assumption, it follows that μ is consistent, we have, by axioms R_{1-2}, that there exists $w_2 \in [\kappa \circ \mu]$. By axiom R_5, we have that:

$$(\kappa \circ \mu) \wedge \varphi_{1,2} \models \kappa \circ (\mu \wedge \varphi_{1,2}).$$

By axiom R_3, we have that $\kappa \circ (\mu \wedge \varphi_{1,2}) \equiv \kappa \circ \varphi_{1,2}$. Since $w_1 \notin [\kappa \circ \mu]$, it follows that $w_1 \notin [\kappa \circ \varphi_{1,2}]$, and hence $w_2 <_\kappa w_1$. But we also have that $w_1 \in \min_{\leq_\kappa}[\mu]$ and $w_2 \in [\mu]$, which implies that $w_1 \leq_\kappa w_2$. We have thus arrived at a contradiction. $\qquad\square$

Theorem 1 tells us that an agent revising beliefs along the lines of postulates R_{1-5} behaves as if it ranks interpretations in $\mathcal{U}$ in a total preorder $\leq_\kappa$ that depends on initial beliefs κ, and always picks the minimal models of μ according to $\leq_\kappa$. Such an agent, then, behaves like a rational agent, in the sense of rational choice theory [26, 11], choosing the best elements from a given menu of options: the menu, here, would be the models of μ, i.e., the possible worlds the agent is allowed to believe in light of new information, while the best elements are decided with reference to $\leq_\kappa$. Thus, a revision operator can be seen as a *choice function* over sets of interpretations: accordingly, Theorem 1 aligns with standard choice theoretic results [2, 26, 22]. That a similar
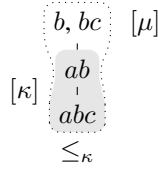
Figure 1: Revision scenario of Example 2, showing the relevant section of the preorder $\leq_\kappa$ on the basis of which the revision result is constructed: only models of $\mu = b$ are depicted; the models of κ among this set are highlighted in grey.

mathematical formalism underlies both belief revision and rational choice is, by itself, not a new insight, the topic having been studied under various guises [24, 5]. The main difference is one of interpretation, with $\leq_\kappa$ usually thought of in revision as a plausibility ranking, i.e., the agent's assessment as to which possible worlds are more or less plausible.

Example 2. A doctor knows that the patient has been diagnosed with asthma (a), finds out that the patient is suffering from shortness of breath (b) and infers that chest pain (c) is also present: the two often go together in asthma, and such stereotypical examples are more salient in the doctor's mind than less frequent, more exotic examples. In other words, the doctor has initial information $\kappa = a$ and a plausibility ranking $\leq_\kappa$ over possible worlds (with respect to the alphabet $\mathcal{P} = \{a, b, c\}$) depicted in Figure 1. The doctor then revises by $\mu = b$ and settles on a possible state of affairs that is most plausible according to their plausibility ranking $\leq_\kappa$, i.e., $[\kappa \circ \mu] = \min_{\leq_\kappa}[\mu] = \{abc\}$. Note that the doctor, in this case, takes the situation represented by $a \wedge b \wedge c$ to be more likely than $a \wedge b \wedge \neg c$.

One way of thinking of postulates R_{1-5} is that they axiomatize total preorders on interpretations. These preorders nominally depend on κ, but nothing in postulates R_{1-5} touches on how models of κ should influence these preorders. In other words, there is as yet no information about the attitude of an agent towards its initial epistemic state, and postulates R_{1-5} are consistent with arbitrary attitudes towards κ. How should the models of κ stand in relation to all other interpretations? Example 2 offers a glimpse into one possible answer: the agent starts off with some information κ and differentiates among possible worlds consistent with κ: some of these worlds are more plausible than others, perhaps as a result of being more salient, or because of a systematic bias [15]. Still, as a whole, models of κ are more plausible than any *other* interpretations consistent with the new information μ. In other words, the agent is biased towards the possible worlds consistent with κ. Are there, now, other ways of arranging the models of κ in $\leq_\kappa$, ways that span the space of possible such attitudes? We study this question through the lens of additional axioms.

3.3 Attitudes Towards Initial Beliefs

If $\kappa, \kappa' \in 2^{\mathsf{Prop}}_{\mathsf{cons}}$, $\mu \in \mathsf{Prop}$, consider the following postulates:

(R$_6$) If $\kappa \wedge \mu$ is consistent, then $\kappa \wedge \mu \models \kappa \circ \mu$.

(R$_7$) If $\kappa \wedge \mu$ is consistent, then $\kappa \circ \mu \models \kappa \wedge \mu$.

(R$_8$) If $\kappa \circ \mu \models \overline{\kappa}$, then $\kappa \circ \mu \equiv \mu$.

(R$_9$) If $\mu \not\models \overline{\kappa}$, then $(\kappa \circ \mu) \wedge \overline{\kappa}$ is inconsistent.

(R$_{10}$) If κ' is complete and $\kappa' \models \kappa \circ \mu$, then $\kappa' \models (\kappa \vee \kappa') \circ \mu$.

Each of these postulates encodes a particular type of attitude towards initial beliefs, and they are intended to be thought of in conjunction with the basic set of postulates R$_{1-5}$. Some clarification is in order. Postulate R$_6$ models an agent that incorporates all information in $\kappa \wedge \mu$, and possibly extends this to cover more ground. Postulate R$_7$ models an agent that reserves the right to drop information from κ if it so sees fit, even if that information is consistent with μ: we may imagine this is done on the basis of certain preferences over the information encoded by κ, i.e., the agent is partial towards certain parts of κ to the detriment of others. Taken together, postulates R$_{6-7}$ imply that $\kappa \circ \mu$ is equivalent to $\kappa \wedge \mu$, when $\kappa \wedge \mu$ is consistent. This property models an agent who wants to preserve as much of κ as it can, and does not have any bias towards either of the models of κ. Postulate R$_6$ can be equated with the *Inclusion* postulate in the AGM formulation and R$_7$ corresponds to *Vacuity* [9], while in the KM axiomatization R$_6$ and R$_7$ are packaged together in one postulate (i.e., KM postulate R$_2$) and presented alongside R$_{1-5}$ as the default set of rational properties for revision [17].

Postulates R$_{8-9}$ focus on the dual knowledge base $\overline{\kappa}$ obtained by replacing every literal in κ with its negated version. If κ is a conjunction of literals, or if it is a complete (i.e., with exactly one model) formula, then $\overline{\kappa}$ will be a formula whose models are complements of the models of κ.

Example 3. If $\mathcal{P} = \{a, b, c\}$ and $\kappa = \{a \wedge b\}$ is a knowledge base over the alphabet $\mathcal{P}$, then $\overline{\kappa} = \{\neg a \wedge \neg b\}$, and we have that $[\kappa] = \{ab, abc\}$, while $[\overline{\kappa}] = \{\emptyset, c\}$

Thus, if κ is very specific (e.g., is a conjunction of literals), then $\overline{\kappa}$ can be thought of as the point of view opposite to that of κ, and situations can be imagined in which it is desirable to put bounds on the revision function in terms of how it treats information encoded by $\overline{\kappa}$. This is the case if the agent has, or is required to have, a definite opinion on every item from an agenda, as is typically the case in Judgment Aggregation [8]; if κ is a 'vivid' knowledge base [21]; or, if it encodes something like an agent's preferred bundle from a set of available items. In all these cases κ can be required to be a conjunction of literals or a complete formula.

Postulate R$_8$ says that if κ undergoes revision by a formula μ embodying such an adverse perspective, then the agent must adopt μ: in other words, the agent has no room for maneuvering towards a more amenable middle ground. Such a revision policy makes more sense when considered alongside postulate R$_9$, which specifies that if the agent has the option of believing states of affairs *not* compatible with $\overline{\kappa}$, it should wholeheartedly adopt those as the most plausible stance. Taken together, postulates R$_{8-9}$ inform the agent to believe states of affairs compatible with $\overline{\kappa}$ only if it has no other choice in the matter: the models of $\overline{\kappa}$ should be part of a viewpoint one is willing to accept only as a last resort. Postulate R$_{10}$ is best understood through an example.

Example 4. An agent intends to go to an art museum, the beach and a concert, i.e., $\kappa = \{a \wedge b \wedge c\}$, with $\mathcal{P} = \{a, b, c\}$. The agent then learns that it only has time for one of these activities and chooses the art museum, i.e., $\kappa \circ \mu \equiv a \wedge \neg b \wedge \neg c$. If the agent's initial intentions were less specific, for instance that it would either go to all three places or only to the art museum (i.e., $\kappa = \{(a \wedge b \wedge c) \vee (a \wedge \neg b \wedge \neg c)\}$), then, faced with the same new information μ, $a \wedge \neg b \wedge \neg c$ should still feature as one of its most preferred options.

A clearer view of postulates R_{6-10} emerges when looking at how they situate the models of κ in a total preorder $\leq_\kappa$, for $\kappa, \kappa' \in 2_{\mathsf{cons}}^{\mathsf{Prop}}$, $w_1, w_2, w' \in \mathcal{U}$:

(o_6) If $w_1 \in [\kappa]$, then $w_1 \leq_\kappa w_2$.

(o_7) If $w_1 \in [\kappa]$ and $w_2 \notin [\kappa]$, then $w_1 <_\kappa w_2$.

(o_8) If $w_1 \in [\overline{\kappa}]$, then $w_2 \leq_\kappa w_1$.

(o_9) If $w_1 \in [\overline{\kappa}]$ and $w_2 \notin [\overline{\kappa}]$, then $w_2 <_\kappa w_1$.

(o_{10}) If $w' \leq_\kappa w$ and $[\kappa'] = \{w'\}$, then $w' \leq_{\kappa \vee \kappa'} w$.

Properties o_{6-10} turn out to characterize axioms R_{6-10} on the semantic level, as per the following representation result.

Theorem 2. If $\circ$ is a basic revision operator and $\alpha: 2_{\mathsf{cons}}^{\mathsf{Prop}} \to \mathcal{U}$ is a basic assignment that represents it, then, for any $\kappa \in 2_{\mathsf{cons}}^{\mathsf{Prop}}$ and $\mu \in \mathsf{Prop}$, the following equivalences hold:

(1) $\circ$ satisfies axiom R_6 iff $\leq_\kappa$ satisfies property o_6;

(2) $\circ$ satisfies axiom R_7 iff $\leq_\kappa$ satisfies property o_7;

(3) $\circ$ satisfies axiom R_8 iff $\leq_\kappa$ satisfies property o_8;

(4) $\circ$ satisfies axiom R_9 iff $\leq_\kappa$ satisfies property o_9;

(5) $\circ$ satisfies axiom R_{10} iff $\leq_\kappa$ satisfies property o_{10}.

Proof. Recall that we denote by $\varphi_{1,2}$ a propositional formula for which $[\varphi_{1,2}] = \{w_1, w_2\}$.

For equivalence 1, we show each direction in turn. ("$\Rightarrow$") Take, first, an assignment α satisfying property o_6, and the revision operator $\circ$ represented by it. Let us assume that $\kappa \wedge \mu$ is consistent, and show that for any $w \in [\kappa \wedge \mu]$, it holds that $w \in [\kappa \circ \mu]$ as well. By property o_5, this is equivalent to showing that $w \in \min_{\leq_\kappa}[\mu]$. Take an arbitrary interpretation $w' \in [\mu]$. Since $w \in [\kappa]$, we can apply property o_6 to get that $w \leq_\kappa w'$. Hence $w \in \min_{\leq_\kappa}[\mu]$. ("$\Leftarrow$") Take a basic revision operator $\circ$ satisfying R_6. and the assignment α which represents it. To show that $\leq_\kappa$ satisfies property o_6, take two interpretations w_1 and w_2 such that $w_1 \in [\kappa]$. Then, by axiom R_6, we have that $\kappa \wedge \varphi_{1,2} \models \kappa \circ \varphi_{1,2}$. By property o_5, it holds that $[\kappa \circ \varphi_{1,2}] = \min_{\leq_\kappa}[\varphi_{1,2}]$ and, since $w_1 \in [\kappa \wedge \varphi_{1,2}]$, it follows that $w_1 \in \min_{\leq_\kappa}[\varphi_{1,2}]$. Thus, $w_1 \leq_\kappa w_2$.

Equivalence 1 is known and can be found in the belief revision literature, where it is usually bundled up together with Theorem 1 (see, for instance, the proof in [17]).

For equivalence 2, we show again each direction in turn. ("$\Rightarrow$") Take an assignment α satisfying property o_7 and the revision operator $\circ$ represented by it. Let us assume that $\kappa \wedge \mu$ is consistent. Take $w \in [\kappa \circ \mu]$, and suppose $w \notin [\kappa \wedge \mu]$. By property o_5, we have that $[\kappa \circ \mu] = \min_{\leq_\kappa}[\mu]$, and hence $w \in [\mu]$. Thus, the fact that $w \notin [\kappa \wedge \mu]$ implies that $w \notin [\kappa]$. But, by assumption, it holds that $[\kappa \wedge \mu] \neq \emptyset$. Thus, there exists $w' \in [\kappa \wedge \mu]$ and, by property o_7, it follows that $w' <_\kappa w$. But we also have that $w \in \min_{\leq_\kappa}[\mu]$, which implies that it cannot be the case that $w' <_\kappa w$, which is a contradiction. ("$\Leftarrow$") Take a basic revision operator $\circ$ satisfying R_7 and the preorder $\leq_\kappa$ that represents it. To show that $\leq_\kappa$ satisfies property o_7, take $w_1 \in [\kappa]$ and $w_2 \notin [\kappa]$. We then have that $\kappa \wedge \varphi_{1,2}$ is consistent and hence, by axiom R_7, that $\kappa \circ \varphi_{1,2} \models \kappa \wedge \varphi_{1,2}$. Since $[\kappa \circ \varphi_{1,2}]$ is, by axioms R_{1-2}, a non-empty subset of $[\varphi_{1,2}] = \{w_1, w_2\}$, we have that at least one of w_1 and w_2 is in $[\kappa \circ \varphi_{1,2}]$. Notice, now, that we cannot have $w_2 \in [\kappa \circ \varphi_{1,2}]$, since it would follow that $w_2 \in [\kappa \wedge \varphi_{1,2}]$ and $w_2 \in [\kappa]$, which is a contradiction. Thus, $[\kappa \circ \varphi_{1,2}] = \{w_1\}$. Since $\leq_\kappa$ represents $\circ$, we have by property o_5 that $[\kappa \circ \varphi_{1,2}] = \min_{\leq_\kappa}[\varphi_{1,2}]$. It follows that $w_1 <_\kappa w_2$.

Equivalences 3 and 4 are analogous to 1 and 2, respectively. For equivalence 5, assume first that axiom R_{10} holds, and take interpretations w and w' and a knowledge base κ' such that $w' \leq_\kappa w$ and $[\kappa'] = \{w'\}$. To show that $w' \leq_{\kappa \vee \kappa'} w$, we must show that $w' \in [(\kappa \vee \kappa') \circ \varphi_{w,w'}]$, where $\varphi_{w,w'}$ is a formula such that $[\varphi_{w,w'}] = \{w, w'\}$. This follows immediately by applying axiom R_{10}. Conversely, suppose $[\kappa'] = \{w'\}$, and take $w \in [\kappa \circ \mu]$. Then, we get that $w' \leq_\kappa w$, and we can apply property o_{10} to derive the conclusion. $\qquad\qquad\square$

Theorem 2 is better understood through an illustration of how such preorders treat models of κ. Property o_6 says that models of κ are minimal elements in κ, i.e., the agent considers possible worlds satisfying its beliefs among the most plausible possible worlds, though possibly not uniquely so (Figure 2-(a)). Property o_7 states that there are no counter-models of κ more plausible than the models of κ, but the models of κ themselves may not be equally plausible (Figure 2-(b)). Properties o_{8-9} say that models of the dual knowledge base κ are the least plausible interpretations in $\leq_\kappa$ (Figure 2-(c,d)), while property o_{10} says that if w' is more plausible than w when the initial beliefs are κ, then w' would still be more plausible than w if it were part of the initial beliefs (Figure 2-(e)).

Together, properties o_{1-7} define what is more commonly known as a *faithful assignment*, placing all and only models of κ on the lowest level of $\leq_\kappa$. This corresponds to an agent that holds its initial beliefs to be the most plausible states of affairs [17]. Consequently, Theorem 1 plus equivalences 1-2 from Theorem 2 make up the classical representation result for belief revision operators [17]. Here we have opted for a more fine-grained approach to the placement of models of κ in $\leq_\kappa$, which allows a more diverse representation of the different types of attitudes an agent can have towards initial beliefs. Though operators that do not satisfy the classical KM postulate R_2 have been considered before [25, 4], the idea that such deviations correspond to possible epistemic attitudes and can be axiomatized is, to the best of our knowledge, new.
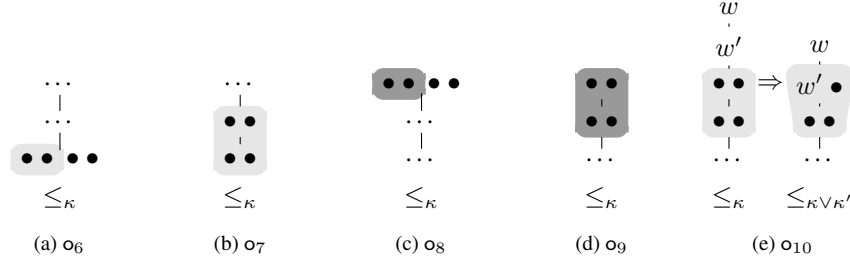
w

w' w

$\Rightarrow$ w'

$\leq_\kappa$ $\quad$ $\leq_\kappa$ $\quad$ $\leq_\kappa$ $\quad$ $\leq_\kappa$ $\quad$ $\leq_\kappa$ $\quad$ $\leq_{\kappa \vee \kappa'}$

(a) o_6 $\qquad$ (b) o_7 $\qquad$ (c) o_8 $\qquad$ (d) o_9 $\qquad$ (e) o_{10}

Figure 2: Schematic view of prototypical preorders satisfying each of the properties o_{6-10}; models of κ are in the light gray area, models of $\overline{\kappa}$ are in the dark gray area.

3.4 Indifference to Already Held Beliefs

One particular consequence of weakening the KM axiom R_2 (axioms R_{6-7} in the current context) is that the following property is not guaranteed to hold anymore:

($\mathsf{R}_{\mathsf{IDF}}$) $\kappa \circ \kappa \equiv \kappa$.

This property, called here $\mathsf{R}_{\mathsf{IDF}}$ (for *indifference to already held beliefs*), says that revising with information the agent already believes does not change the agent's epistemic state. More generally, the KM standard set of postulates implies that revising by any formula μ such that $\kappa \models \mu$ results in κ. It quickly becomes apparent that axiom R_6 implies $\mathsf{R}_{\mathsf{IDF}}$,[2] but R_7 does not. Thus, if an agent is allowed to rank models of κ unequally, then $\mathsf{R}_{\mathsf{IDF}}$ is not guaranteed to hold.

Example 5. If $\mathcal{P} = \{a, b\}$, take $\kappa = \{a \vee b\}$ and a revision operator that satisfies axiom R_7, and which orders interpretations as follows: $a \approx_\kappa b <_\kappa ab <_\kappa \emptyset$. We get that $[\kappa \circ \kappa] = \{a, b\}$, i.e., $\kappa \circ \kappa \equiv (a \leftrightarrow \neg b) \not\equiv \kappa$.

This points to a more graded view of what it means to believe κ. Thus, an agent might have a certain threshold of plausibility, along the lines of what is known in epistemology as *the Lockean thesis* [10], according to which it calibrates its beliefs: anything above the threshold counts as part of the belief κ and anything below counts as disbelief. This fits with the idea that an agent might assign different degrees of plausibility to states of affairs consistent with its belief κ: indeed, this is the point of view we endorse here, in contrast to more standard approaches, which consider that an agent assigns equal degrees of plausibility to all items of its belief. Thus, incoming information that confirms an agent's belief might have the effect of *reinforcing* parts that are given more plausibility at the expense of parts that are given less, and this is the kind of situation we take to be modeled by Example 5.

What would be worrying would be a revision policy that makes an agent cycle between different viewpoints when confronted repeatedly with the same type of information: we will see that for revision operators satisfying R_7 this concern is unwarranted, but we must first introduce some new notation. We write κ^i for the knowledge base

---

[2]The converse is not true: $\mathsf{R}_{\mathsf{IDF}}$ enforces only that models of κ are equally plausible, but not where they are placed in $\leq_\kappa$.

$$
\begin{array}{ccccc}
 & & ab & & b \\
 & & | & & | \\
 & b, ab & b & & ab \\
 & | & | & & | \\
[\kappa]\ \emptyset & [\kappa]\ a & [\kappa]\ \emptyset \\
 | & | & | \\
 a & \emptyset & a \\
 \\
 \leq_{\kappa^0} & \leq_{\kappa^1} & \leq_{\kappa^2}
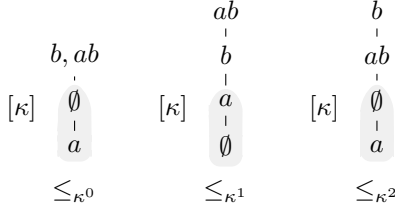\end{array}
$$

Figure 3: Repeated revision by κ cycles between $\{a\}$ and $\{\emptyset\}$.

obtained by revising κ by itself an i number of times. Thus, $\kappa^0 = \kappa$ and $\kappa^{i+1} = \kappa^i \circ \kappa$. Consider now the following property:

(R$_{\mathsf{STB}}$) There is $n \geq 1$ such that $\kappa^m \equiv \kappa^n$, for every $m \geq n$.

A revision operator $\circ$ is *stable* if it satisfies property R$_{\mathsf{STB}}$. Stability implies that repeated revision by κ ultimately settles (or stabilizes) on a set of models that does not change through subsequent revisions by κ. The following result proves relevant to the issue of stability.

Proposition 3. If a revision operator $\circ$ satisfies axioms R$_1$ and R$_7$, then $\kappa^{i+1} \models \kappa^i$.

Proof. By axiom R$_1$, we have that $\kappa \circ \kappa \models \kappa$, and thus $\kappa^1 \models \kappa^0$. Applying axiom R$_7$, we have that $(\kappa \circ \kappa) \circ \kappa \models (\kappa \circ \kappa) \wedge \kappa \models \kappa \circ \kappa$. Thus, $\kappa^2 \models \kappa^1$, and it is straightforward to see how this argument is iterated to get the conclusion. $\qquad\square$

If the operator $\circ$ also satisfies axiom R$_2$ (which, here, says that if the revision formula is consistent, then the revision result is also consistent), it follows that if κ is consistent, then κ_i is consistent, for any $i \geq 0$. Thus, combining this fact and Proposition 3, we get that repeated revision by κ leads to a chain of ever more specific knowledge bases, i.e., $\emptyset \subset \cdots \subseteq [\kappa^{i+1}] \subseteq [\kappa^i] \subseteq \cdots \subseteq [\kappa^0]$. Since a knowledge base has a finite number of models, it falls out immediately from this that there must be a point at which further revision by κ does not change anything.

Corollary 4. A basic revision operator $\circ$ satisfying axiom R$_7$ is stable.

Unfortunately, axioms R$_{8-9}$ do not guarantee stability. Since these axioms require only that the agent places the models of $\overline{\kappa}$ as the least plausible interpretations, it becomes possible that an agent's plausibility ranking does not hold on to a core set of interpretations through successive revisions by κ.

Example 6. If $\mathcal{P} = \{a, b\}$, take $\kappa = \{\neg b\}$ and a revision operator satisfying R$_{8-9}$, which orders interpretations as shown in Figure 3. We have that $[\kappa^0] = [\kappa] = \{\emptyset, a\}$, $[\kappa^1] = [\kappa \circ \kappa] = \{a\}$, and $[\kappa^2] = [\kappa^1 \circ \kappa] = \{\emptyset\}$. By R$_3$, subsequent revisions by κ cycle between $\{a\}$ and $\{\emptyset\}$ ($[\kappa^3] = \{a\}$, $[\kappa^4] = \{\emptyset\}$), thus never settling on a stable answer.

The issue of stability suggests another dimension along which revision operators can be analyzed, with Corollary 4 and Example 6 showing that a revision operator does not

satisfy it trivially. Example 6, in particular, shows that there is interplay between $\leq_\kappa$ and $\leq_{\kappa'}$, if $\kappa' \models \kappa$, which is relevant to the question of whether an operator is stable. This interplay is reminiscent of topics like iterated revision and kinetic consistency [7, 23], but pursuing it further would take us too far afield of the aims of the current work.

4 Concrete Operators with Varying Attitudes

We now ask: what is a natural way to construct operators with such varying attitudes towards initial information? Our answer builds on ideas found in belief merging [19, 20], i.e., it defines rankings on interpretations by appeal to *two* main ingredients. The first is a *distance* $d \colon \mathcal{U} \times \mathcal{U} \to \mathbb{R}_{\geq 0}$ between interpretations, defined such that $d(w_1, w_2) = 0$ iff $w_1 = w_2$ and $d(w_1, w_2) = d(w_2, w_1)$. Given $w \in \mathcal{U}$ and $\kappa \in 2^{\mathsf{Prop}}_{\mathsf{cons}}$ such that $[\kappa] = \{w_1, \ldots, w_n\}$, the *vector of distances from w to κ* is $d(w, \kappa) = (d(w, w_1), \ldots, d(w, w_n))$. For brevity we will omit commas and simply write $d(w, \kappa)$ as a string of numbers. We recall two prominent examples of distance: *drastic distance* d_D works by the all-or-nothing rule: $d_\mathsf{D}(w_1, w_2) = 0$ if $w_1 = w_2$, and 1 otherwise; *Hamming distance* d_H, counts the number of atoms on which two interpretations differ. The second ingredient is an *aggregation function* (more precisely, a family of functions) $f \colon \mathbb{R}^n \to \mathbb{R}$, for $n \in \mathbb{N}$, mapping a distance vector $d(w, \kappa)$ to a number, and used to compare distance vectors. We write $\overrightarrow{d(w, \kappa)}$ and $\overleftarrow{d(w, \kappa)}$ for the vectors of distances from w to κ ordered in ascending order and descending order, respectively. The lexicographic order between two vectors is denoted by $\leq_{\mathsf{lex}}$. The minimal and maximal elements of $d(w, \kappa)$ are $\min d(w, \kappa)$ and $\max d(w, \kappa)$, respectively, and $\sum d(w, \kappa) = \sum_{w_i \in [\kappa]} d(w, w_i)$. The *centrality of w with respect to κ* is $\mathrm{cen}(w, \kappa) = \max d(w, \kappa) - \min d(w, \kappa)$. The *displacement of w with respect to κ* is $\mathrm{dis}(w, \kappa) = \min d(w, \kappa) - \min d(w^*, \kappa)$, where w^* is an interpretation such that $\min d(w^*, \kappa)$ is minimal among all the interpretations w' for which $\mathrm{cen}(w', \kappa) = \mathrm{cen}(w, \kappa)$. Finally, the *agreeability index of w with respect to κ* is $\mathrm{agr}(w, \kappa) = \min\{\min d(w, \kappa), \ \mathrm{cen}(w, \kappa) + \mathrm{dis}(w, \kappa)\}$, while the *disagreeability index of w with respect to κ* is $\mathrm{dagr}(w, \kappa) = n - \mathrm{agr}(w, \overline{\kappa})$, where $n = |\mathcal{P}|$.

Given a distance d between interpretations and aggregation function f, we write $\leq_\kappa^{d, f}$ for the ranking generated using d and f, and $\circ^{d, f}$ for the revision operator represented by the assignment generated using d and f, i.e., $[\kappa \circ^{d, f} \mu] = \min_{\leq_\kappa^{d, f}}[\mu]$. We then define the following types of rankings:

$$
\begin{array}{llll}
w_1 \leq_\kappa^{d, \min} w_2 & \text{if} & \min d(w_1, \kappa) \leq \min d(w_2, \kappa), \\
w_1 \leq_\kappa^{d, \mathrm{lmin}} w_2 & \text{if} & \overrightarrow{d(w_1, \kappa)} \leq_{\mathsf{lex}} \overrightarrow{d(w_2, \kappa)}, \\
w_1 \leq_\kappa^{d, \mathrm{agr}} w_2 & \text{if} & \mathrm{agr}(w_1, \kappa) \leq \mathrm{agr}(w_2, \kappa), \\
w_1 \leq_\kappa^{d, \max} w_2 & \text{if} & \max d(w_1, \kappa) \leq \max d(w_2, \kappa), \\
w_1 \leq_\kappa^{d, \mathrm{lmax}} w_2 & \text{if} & \overleftarrow{d(w_1, \kappa)} \leq_{\mathsf{lex}} \overleftarrow{d(w_2, \kappa)}, \\
w_1 \leq_\kappa^{d, \mathrm{dagr}} w_2 & \text{if} & \mathrm{dagr}(w_1, \kappa) \leq \mathrm{dagr}(w_2, \kappa), \\
w_1 \leq_\kappa^{d, \mathrm{sum}} w_2 & \text{if} & \sum d(w_1, \kappa) \leq \sum d(w_2, \kappa).
\end{array}
$$

Example 7. If $\mathcal{P} = \{a, b, c\}$, take $\kappa = \{(\neg(a \wedge b) \wedge \neg c) \vee (a \wedge b \wedge c)\}$, for which we get that $[\kappa] = \{\emptyset, a, b, abc\}$. For the interpretation $w = \emptyset$, we get that

	$\emptyset$	a	b	abc	$\overrightarrow{d_{\mathsf{H}}(w,\kappa)}$	$\overleftarrow{d_{\mathsf{H}}(w,\kappa)}$	min	max	$\sum$
$\emptyset$	0	1	1	3	(0113)	(3110)	0	3	5
a	1	0	2	2	(0122)	(2210)	0	2	5
b	1	2	0	2	(0122)	(2210)	0	2	5
c	1	2	2	2	(1222)	(2221)	1	2	7
ab	2	1	1	1	(1112)	(2111)	1	2	5
ac	2	1	3	1	(1123)	(3211)	1	3	7
bc	2	3	1	1	(1123)	(3211)	1	3	7
abc	3	2	2	0	(0223)	(3220)	0	3	7

Table 1: Table of Hamming distances for κ from Example 7

$\overrightarrow{d_{\mathsf{H}}(w,\kappa)} = (0113)$, $\overleftarrow{d_{\mathsf{H}}(w,\kappa)} = (3110)$, $\min d_{\mathsf{H}}(w,\kappa) = 0$, $\max d_{\mathsf{H}}(w,\kappa) = 3$ and $\sum d_{\mathsf{H}}(w,\kappa) = 5$. The distances and aggregated distances for each interpretation are depicted in Table 1. Notice how the models of κ are distributed when the interpretations are ranked according to the different aggregation functions used: we have $\emptyset \approx_{\kappa}^{\mathsf{H},\min} a$, since $\min d_{\mathsf{H}}(\emptyset,\kappa) = \min d_{\mathsf{H}}(a,\kappa) = 0$, but $\emptyset <_{\kappa}^{\mathsf{H},\mathrm{lmin}} a$, since $(0113) \leq_{\mathsf{lex}} (0122)$. Also, we have that $c <_{\kappa}^{\mathsf{H},\max} abc$, $c <_{\kappa}^{\mathsf{H},\mathrm{lmax}} abc$ and $ab <_{\kappa}^{\mathsf{H},\mathrm{sum}} abc$, i.e., models of κ are not minimal in $\leq_{\kappa}^{\mathsf{H},\max}$, $\leq_{\kappa}^{\mathsf{H},\mathrm{lmax}}$ and $\leq_{\kappa}^{\mathsf{H},\mathrm{sum}}$. In particular, $\leq_{\kappa}^{\mathsf{H},\max}$ makes the models of $\overline{\kappa}$ (i.e., abc, bc, ac and $\emptyset$) the least plausible interpretations.

The agreement and disagreement operators ($\circ^{d,\mathrm{agr}}$ and $\circ^{d,\mathrm{dagr}}$) are simpler than they appear: the idea behind $\circ^{d,\mathrm{agr}}$ is to allow interpretations other than the models of κ as the minimal elements of the preorder $\leq_{\kappa}$. Notice that the score of an interpretation in $\leq_{\kappa}^{d,\mathrm{agr}}$ is 0 if it is either a model of κ, or it is equidistant from every model of κ (i.e., its centrality is 0) and it is the 'closest' interpretation to κ with this property. The disagreement operator $\circ^{d,\mathrm{dagr}}$ works in similar fashion, by making models of $\overline{\kappa}$ and interpretations minimally equidistant to them the least plausible interpretations in $\leq_{\kappa}^{d,\mathrm{dagr}}$.

Example 8. If $\mathcal{P} = \{a,b,c\}$, take κ such that $[\kappa] = \{a,b,c\}$, and notice that $d_{\mathsf{H}}(\emptyset,\kappa) = (111)$ and $d_{\mathsf{H}}(abc,\kappa) = (222)$, i.e., they are both equidistant to κ, hence their centrality is 0. However, $\emptyset$ is closer to κ than abc (its displacement is 0, compared to abc's displacement of 1), and $\mathrm{agr}(\emptyset,\kappa) = 0$. Thus, what $\leq_{\kappa}^{\mathsf{H},\mathrm{agr}}$ does is to give a minimal score to models of κ and to the minimally equidistant interpretation $\emptyset$. By contrast, $\leq_{\kappa}^{\mathsf{H},\mathrm{dagr}}$ gives a maximal score to the models of $\overline{\kappa}$ and to the maximally equidistant interpretation abc.

All operators proposed generate a total preorder $\leq_{\kappa}$ over interpretations, but differ in how they arrange models of κ in $\leq_{\kappa}$: this corresponds to the different attitudes an agent can have towards κ prior to any revision. The operator $\circ^{\mathsf{H},\min}$, known as Dalal's operator [6], considers all models of κ as the most plausible elements in $\leq_{\kappa}$ and is the only operator for which $\kappa \circ \mu$ is equivalent to $\kappa \wedge \mu$ when $\kappa \wedge \mu$ is consistent. Similarly, $\circ^{\mathsf{H},\mathrm{lmin}}$ also ranks models of κ as more plausible than any other interpretation, but discriminates among models of κ. The operators $\leq_{\kappa}^{\mathsf{H},\max}$ and $\leq_{\kappa}^{\mathsf{H},\mathrm{lmax}}$ push away models of $\overline{\kappa}$, under the assumption that they are the most implausible possible worlds. They difference between them is that $\leq_{\kappa}^{\mathsf{H},\max}$ considers models of $\overline{\kappa}$ equally implausible, whereas $\leq_{\kappa}^{\mathsf{H},\mathrm{lmax}}$ uses the more fine-grained lexicographic approach. The operator

$\circ^{\mathsf{H,agr}}$ makes models of κ the most plausible elements in $\leq_\kappa$ but also allows other interpretations on that position, in particular certain interpretations that are equidistant to κ as per Example 8. The intuition is that an interpretation equally distanced from models of κ is like a compromise point of view, with good chances of being correct if it is close to κ. The operator $\circ^{\mathsf{H,dagr}}$ is the dual of $\circ^{\mathsf{H,agr}}$ and, finally, operator $\circ^{\mathsf{H,sum}}$ evokes utilitarian approaches by choosing interpretations that minimize the sum of the distances to each model of κ.

Quick reflection shows that operators obtained with drastic distance d_D collapse into two main categories. To get a grasp on this fact, consider first the *drastic revision operator* $\circ^{\mathsf{dr}}$ defined, for $\kappa \in 2^{\mathsf{Prop}}_{\mathsf{cons}}$ and $\mu \in \mathsf{Prop}$, as $\kappa \circ^{\mathsf{dr}} \mu = \kappa \wedge \mu$, if $\kappa \wedge \mu$ is consistent, and μ otherwise, and the *forgetful revision operator* $\circ^{\mathsf{fg}}$ defined as $\kappa \circ^{\mathsf{fg}} \mu = \mu$.

Proposition 5. For any knowledge base κ and formula μ, it holds that $\kappa \circ^{\mathsf{D,min}} \mu \equiv \kappa \circ^{\mathsf{D,lmin}} \mu \equiv \kappa \circ^{\mathsf{D,lmax}} \mu \equiv \kappa \circ^{\mathsf{D,sum}} \mu \equiv \kappa \circ^{\mathsf{dr}} \mu$. Moreover, $\kappa \circ^{\mathsf{D,agr}} \mu \equiv \kappa \circ^{\mathsf{fg}} \mu$ and $\kappa \circ^{\mathsf{D,max}} \mu \equiv \kappa \circ^{\mathsf{D,dagr}} \mu \equiv \begin{cases} \kappa \circ^{\mathsf{dr}} \mu, & \text{if } \kappa \text{ is complete,} \\ \kappa \circ^{\mathsf{fg}} \mu, & \text{otherwise.} \end{cases}$

Proof. If κ is a complete knowledge base and $[\kappa] = \{w_0\}$, then $d_\mathsf{D}(w, \kappa) = d_\mathsf{D}(w, w_0)$, for any interpretation w. In other words, $\overrightarrow{d_\mathsf{D}(w, \kappa)}$ yields (0) if $w = w_0$ and (1) otherwise. Thus, $\max d_\mathsf{D}(w, \kappa) = 0$ if $w = w_0$, and $\max d_\mathsf{D}(w, \kappa) = 1$, otherwise. It is then straightforward to see that if $w_0 \in [\mu]$, then $[\kappa \circ^{\mathsf{D,max}} \mu] = \{w_0\} = [\kappa \wedge \mu]$, and if $w_0 \notin [\mu]$, then $[\kappa \circ^{\mathsf{D,max}} \mu] = [\mu]$. If κ is not complete, then $\overrightarrow{d_\mathsf{D}(w, \kappa)}$ yields $(01 \ldots 1)$ if $w \in [\kappa]$, and $(1 \ldots 1)$ otherwise. It is now straightforward to see that the remaining statements of Proposition 5 hold. $\square$

With Hamming distance the landscape is more diverse, as the different attitudes the operators assume towards models of κ lead to genuinely different revision strategies. Nonetheless, certain relationships between the operators still hold, with lexicographic operators being the most discriminating, in the sense that they pick formulas with fewer models.

Proposition 6. For any $\kappa \in 2^{\mathsf{Prop}}_{\mathsf{cons}}$ and $\mu \in \mathsf{Prop}$, we have that $\kappa \circ^{\mathsf{H,lmin}} \mu \models \kappa \circ^{\mathsf{H,min}} \mu \models \kappa \circ^{\mathsf{H,agr}} \mu$ and $\kappa \circ^{\mathsf{H,lmax}} \mu \models \kappa \circ^{\mathsf{H,max}} \mu \models \kappa \circ^{\mathsf{H,dagr}} \mu$.

All operators generate total preorders over interpretations, so by Theorem 1 they all satisfy axioms R_{1-5}. Satisfaction with respect to the newly introduced postulates is clarified below.

Proposition 7. For $d \in \{\mathsf{D}, \mathsf{H}\}$ and $f \in \{\min, \mathrm{lmin}, \max, \mathrm{lmax}, \mathrm{agr}, \mathrm{dagr}, \mathrm{sum}\}$, the operators $\circ^{d,f}$ satisfy postulates $\mathsf{R}_{6-10, \mathsf{IDF}, \mathsf{STB}}$ as per Table 2.

Proof. It is already known that $\leq_\kappa^{d,\min}$ (known as Dalal's operator [6, 16]) satisfies axioms R_{5-6}. To see why the operator $\circ^{\mathsf{H,lmin}}$ satisfies R_7, notice that if $w_1 \in [\kappa]$ and $w_2 \notin [\kappa]$, then the first element in $d(w_1, \kappa)$ is 0, while the first element in $d(w_2, \kappa)$ is strictly greater than 0. This implies that $\overrightarrow{d(w_1, \kappa)} <_{\mathsf{lex}} \overrightarrow{d(w_2, \kappa)}$, which in turn implies that $w_1 <_\kappa^{\mathsf{H,lmin}} w_2$. Hence property o_7 is satisfied, which implies that axiom R_7 is

15

	R_6	R_7	R_8	R_9	R_{10}	R_{IDF}	R_{STB}
$\circ^{H,\min}$	✓	✓	×	×	✓	✓	✓
$\circ^{H,\mathrm{lmin}}$	×	✓	×	×	✓	×	✓
$\circ^{H,\mathrm{agr}}$	✓	×	×	×	✓	✓	✓
$\circ^{H,\max}$	×	×	✓	✓	×	×	✓
$\circ^{H,\mathrm{lmax}}$	×	×	×	✓	×	×	✓
$\circ^{H,\mathrm{dagr}}$	×	×	✓	×	×	✓	✓
$\circ^{H,\mathrm{sum}}$	×	×	×	×	✓	×	✓
$\circ^{\mathrm{dr}}$	✓	✓	×	×	✓	✓	✓
$\circ^{\mathrm{fg}}$	✓	×	✓	×	✓	✓	✓

Table 2: Satisfaction of axioms

satisfied. Additionally, it cannot be the case that $w_3 <^{H,\mathrm{lmin}}_\kappa w_1$, for any $w_3 \notin [\kappa]$, which shows that property o_6 (and hence axiom R_6) is satisfied. For $\circ^{H,\mathrm{lmin}}$ and axiom R_6, take $[\kappa] = \{a, b, ab\}$ and $[\mu] = \{a, b, ab\}$. We get that $[\kappa \circ^{H,\mathrm{lmin}} \mu] = \{a, b\}$. The operator $\circ^{H,\mathrm{agr}}$ satisfies axiom R_6 because it makes all models of κ, and potentially other interpretations as well (which is the reason why it does not satisfy axiom R_7), as the equally most plausible interpretations in $\leq^{H,\mathrm{agr}}_\kappa$. Since all these operator place the models of κ on the lowest levels of $\leq_\kappa$, they all satisfy axiom R_{10}.

To see why postulates R_{8-9} are not satisfied by $\circ^{d,\min}$, $\circ^{d,\mathrm{lmin}}$ or $\circ^{d,\mathrm{dagr}}$, notice that these operators do not make models of $\overline{\kappa}$ as the least plausible interpretations in $\leq_\kappa$. Thus, if $\kappa = a \vee b$, then $\overline{\kappa}$ shares some models with κ, yet these models (along with all other models of $a \vee b$) will be among the most plausible interpretations in $\leq^{d,\min}_\kappa$, $\leq^{d,\mathrm{lmin}}_\kappa$ and $\leq^{d,\mathrm{agr}}_\kappa$. The one exception is the forgetful operator $\circ^{\mathrm{fg}}$, which satisfies R_8 trivially.

The case for $\circ^{H,\max}$, $\circ^{H,\mathrm{lmax}}$ and $\circ^{H,\mathrm{dagr}}$ is analogous to the one for $\circ^{H,\min}$, $\circ^{H,\mathrm{lmin}}$ and $\circ^{H,\mathrm{agr}}$, as they can be seen as duals of each other. For the operator $\circ^{H,\mathrm{sum}}$, take $[\kappa] = \{a, b, c\}$ and $[\mu] = \{\emptyset, a, b, c\}$. We get that $[\kappa \circ^{H,\mathrm{sum}} \mu] = \{\emptyset\}$, as $\emptyset$ minimizes the sum of the Hamming distances to the models of κ: this is a counter-example to axioms R_{6-7}. For R_{8-9}, take $[\mu'] = \{\emptyset, ab, ac, bc\}$. For $\circ^{H,\mathrm{sum}}$ and R_{10}, notice that adding w' to $[\kappa]$ creates a new column for w' in the table of distances, in which the distance corresponding to w' is 0, i.e., the score assigned to w' in $\leq^{H,\mathrm{sum}}_{\kappa \vee \kappa'}$ does not increase with respect to $\leq^{H,\mathrm{sum}}_\kappa$. Satisfaction of R_{IDF} and R_{STB} is straightforward, keeping in mind how the various operators arrange the models of κ in the generated preorders. □

5 Conclusion

We have looked at the classical revision axioms from the point of view of what they assume about an agent's attitude towards its initial beliefs, and argued that this attitude is embedded in a specific axiom (the KM postulate corresponding to *Inclusion* and *Vacuity*). By varying this axiom we were able to put forward and characterize a wide range of revision operators, and refine previously entangled intuitions in the process.

Analysis of the new operators also uncovered the principles of indifference to already held beliefs (R_{IDF}) and stability (R_{STB}). Further work is needed to link these notions to the other axioms, to map out their interplay and to provide them with semantic characterizations.

At the same time, the more fine grained view on the types of attitudes an agent can have towards its initial beliefs raises the question of what these attitudes are good for, i.e., whether they can be used for tasks such as learning [18, 3]. The idea here is to view revision as part of an ongoing process by which the agent continuously refines its representation of the outside world, with the aim of settling on stable, correct information. Such a task, we think, provides a natural benchmark for revision operators, and it has the potential to connect belief revision to other topics of importance to the field of AI.

Acknowledgments

References

[1] 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.

[2] Kenneth J. Arrow. Rational choice functions and orderings. *Economica*, 26(102):121–127, 1959.

[3] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. Belief revision as a truth-tracking process. In *Proc. of TARK 2011*, pages 187–190, 2011.

[4] Salem Benferhat, Sylvain Lagrue, and Odile Papini. Revision of Partially Ordered Information: Axiomatization, Semantics and Iteration. In *Proc. of IJCAI 2005*, pages 376–381, 2005.

[5] Giacomo Bonanno. Rational choice and AGM belief revision. *Artif. Intell.*, 173(12-13):1194–1203, 2009.

[6] Mukesh Dalal. Investigations into a Theory of Knowledge Base Revision. In *Proc. of IJCAI 1988*, pages 475–479, 1988.

[7] Adnan Darwiche and Judea Pearl. On the Logic of Iterated Belief Revision. *Artif. Intell.*, 89(1-2):1–29, 1997.

[8] Ulle Endriss. Judgment Aggregation. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, pages 399–426. Cambridge University Press, 2016.

[9] Eduardo Fermé and Sven Ove Hansson. *Belief Change: Introduction and Overview*. Springer, 2018.

[10] Richard Foley. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford University Press, 1993.

[11] Simon French. *Decision Theory: An Introduction to the Mathematics of Rationality*. Ellis Horwood Limited, 1988.

[12] Peter Gärdenfors. *Knowledge in Flux: Modelling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, 1988.

[13] Adrian Haret and Stefan Woltran. Belief revision operators with varying attitudes towards initial beliefs. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 1726–1733. ijcai.org, 2019.

[14] Andreas Herzig and Omar Rifi. Propositional Belief Base Update and Minimal Change. *Artif. Intell.*, 115(1):107–138, 1999.

[15] Daniel Kahneman, Paul Slovic, and Amos Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.

[16] Hirofumi Katsuno and Alberto O. Mendelzon. On the Difference between Updating a Knowledge Base and Revising It. In *Proc. of KR 1991*, pages 387–394, 1991.

[17] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional Knowledge Base Revision and Minimal Change. *Artif. Intell.*, 52(3):263–294, 1992.

[18] Kevin T. Kelly. The Learning Power of Belief Revision. In *Proc. of TARK 1998*, pages 111–124, 1998.

[19] Sébastien Konieczny, Jérôme Lang, and Pierre Marquis. DA^2 merging operators. *Artif. Intell.*, 157(1-2):49–79, 2004.

[20] Sébastien Konieczny and Ramón Pino Pérez. Logic Based Merging. *J. Philosophical Logic*, 40(2):239–270, 2011.

[21] Hector J Levesque. Making believers out of computers. *Artif. Intell.*, 30(1):81–108, 1986.

[22] Hervé Moulin. *Axioms of Cooperative Decision Making*. Number 15 in Econometric Society Monographs. Cambridge University Press, 1991.

[23] Pavlos Peppas and Mary-Anne Williams. Kinetic Consistency and Relevance in Belief Revision. In *Proc. of JELIA 2016*, pages 401–414, 2016.

[24] Hans Rott. *Change, choice and inference: A study of belief revision and nonmonotonic reasoning*. Number 42 in Oxford Logic Guides. Oxford University Press, 2001.

[25] Mark Ryan. Belief Revision and Ordered Theory Presentations. In *Logic, Action, and Information*, pages 129–151, 1996.

[26] Amartya K. Sen. *Collective Choice and Social Welfare*. NY: North-Holland, 2nd edition, 1984.