

# Belief Revision I: The AGM Theory

Franz Huber\*

University of Toronto

---

## Abstract

Belief revision theory studies how an ideal doxastic agent should revise her beliefs when she receives new information. In part I I will first present the AGM theory of belief revision (Alchourrón & Gärdenfors & Makinson 1985). Then I will focus on the problem of *iterated* belief revisions.

---

## 1. Introduction

Belief revision theory studies how an ideal doxastic agent should revise her beliefs when she receives new information. Sophia believes many things, among others that it will rain on Tuesday, that it will be sunny on Wednesday, and that weather forecasts are always reliable. Belief revision theory tells Sophia how to revise her beliefs when she learns that the weather forecast for Tuesday and Wednesday predicts rain. As we will see, this depends on the details of her beliefs, but under one way of filling in the details she should keep her belief that it will rain on Tuesday and give up her belief that it will be sunny on Wednesday. To state in full detail how Sophia should revise her beliefs when she learns new information we need a representation of her old beliefs and of the new information she receives.

Belief revision theory models belief as a qualitative attitude towards sentences or propositions: an ideal doxastic agent believes a proposition, or she disbelieves the proposition by believing its negation, or she suspends judgment with respect to the proposition and its negation. This is different in the theory of subjective probabilities which is sometimes referred to as *Bayesianism* (Easwaran 2011a, 2011b). Here, belief is modeled as a quantitative attitude towards a sentence or proposition: an ideal doxastic agent believes a proposition to a specific degree, her degree of belief or credence for the proposition. As we will see, though, in order to adequately model conditional beliefs and iterated belief revisions, belief revision theory has to model the ideal doxastic agent's beliefs with more than just the set of sentences or propositions she believes.

## 2. The AGM Theory of Belief Revision

The AGM theory of belief revision derives its name from the seminal paper by Alchourrón & Gärdenfors & Makinson (1985). Comprehensive overviews can be found in the work of Gärdenfors (1988), Gärdenfors and Rott (1995), and Rott (2001).

One version of the AGM theory of belief revision represents the agent's old beliefs with a set of sentences from some formal language, her *belief set*, together with an *entrenchment ordering* over these sentences that represents how strongly the agent holds the beliefs in her belief set. The entrenchment ordering represents the details of the agent's beliefs. The new information is represented by a single sentence. The theory distinguishes between the easy case, called *expansion*, where the new information does not contradict the agent's old belief set and is simply added conjunctively and the general case, called *revision*, where the new information might contradict the old belief set. The general case is difficult, because the agent has to turn her

old belief set, which is assumed to be consistent, into a new belief set that contains the new information and is consistent. Usually the general case is dealt with in two steps: in a first step, called *contraction*, the old belief set is cleared from everything that contradicts the new information. In a second step, one simply expands by adding the new information conjunctively. So the hard work is done by contraction, which turns the general case of revision into the easy case of expansion.

A formal language  $\mathcal{L}$  is defined inductively as follows.  $\mathcal{L}$  contains the contradictory sentence  $\perp$  and all elements of a given countable set of propositional variables  $PV = \{p, q, r, \dots\}$ . Furthermore, whenever  $\alpha$  and  $\beta$  are sentences of  $\mathcal{L}$ , then so are the negation of  $\alpha$ ,  $\neg\alpha$ , as well as the conjunction of  $\alpha$  and  $\beta$ ,  $\alpha \wedge \beta$ . Finally, nothing else is a sentence of  $\mathcal{L}$ .

The new information is represented by a single sentence  $\alpha$  from  $\mathcal{L}$ . The ideal doxastic agent's old beliefs are represented by a set of sentences, her belief set  $\mathcal{B} \subseteq \mathcal{L}$ , and an entrenchment ordering  $\preceq$  for  $\mathcal{B}$ . Where  $\vdash$  is the consequence relation on  $\mathcal{L}$  and  $Cn(\mathcal{B}) = \{\alpha \in \mathcal{L} : \mathcal{B} \vdash \alpha\}$  the set of consequences of  $\mathcal{B}$ , the entrenchment ordering has to satisfy the following postulates. For all  $\alpha, \beta, \gamma \in \mathcal{L}$

- $\preceq 1$  If  $\alpha \preceq \beta$  and  $\beta \preceq \gamma$ , then  $\alpha \preceq \gamma$ . Transitivity
- $\preceq 2$  If  $\beta \in Cn(\{\alpha\})$ , then  $\alpha \preceq \beta$ . Dominance
- $\preceq 3$   $\alpha \preceq \alpha \wedge \beta$  or  $\beta \preceq \alpha \wedge \beta$ . Conjunctivity
- $\preceq 4$  If  $\perp \notin Cn(\mathcal{B})$ , then:  $\alpha \notin \mathcal{B}$  just in case for all  $\beta \in \mathcal{L}$ ,  $\alpha \preceq \beta$ . Minimality
- $\preceq 5$  If for all  $\alpha$  from  $\mathcal{L}$ ,  $\alpha \preceq \beta$ , then  $\beta \in Cn(\emptyset)$ . Maximality

Expansion, revision, and contraction turn belief sets and new information into belief sets. Formally they are functions from  $\wp(\mathcal{L}) \times \mathcal{L}$  into  $\wp(\mathcal{L})$ . Expansion  $\dot{+}$  turns each old belief set  $\mathcal{B} \subseteq \mathcal{L}$  and sentence  $\alpha \in \mathcal{L}$  into the new belief set  $\mathcal{B} \dot{+} \alpha = Cn(\mathcal{B} \cup \{\alpha\})$ . Revision  $*$  turns each old belief set  $\mathcal{B} \subseteq \mathcal{L}$  and sentence  $\alpha \in \mathcal{L}$  into a new belief set  $\mathcal{B} * \alpha$ , where  $*$  is required to satisfy the following postulates. For all  $\mathcal{B} \subseteq \mathcal{L}$  and all  $\alpha, \beta \in \mathcal{L}$

- \*1.  $\mathcal{B} * \alpha = Cn(\mathcal{B} * \alpha)$ . Closure
- \*2.  $\alpha \in \mathcal{B} * \alpha$ . Success
- \*3.  $\mathcal{B} * \alpha \subseteq Cn(\mathcal{B} \cup \{\alpha\})$ . Inclusion
- \*4. If  $\neg\alpha \notin Cn(\mathcal{B})$ , then  $\mathcal{B} \subseteq \mathcal{B} * \alpha$ . Preservation
- \*5. If  $Cn(\{\alpha\}) = Cn(\{\beta\})$ , then  $\mathcal{B} * \alpha = \mathcal{B} * \beta$ . Congruence
- \*6. If  $\neg\alpha \notin Cn(\emptyset)$ , then  $\perp \notin \mathcal{B} * \alpha$ . Consistency
- \*7.  $\mathcal{B} * (\alpha \wedge \beta) \subseteq Cn((\mathcal{B} * \alpha) \cup \{\beta\})$ . Conjunction 1
- \*8. If  $\neg\beta \notin \mathcal{B} * \alpha$ , then  $Cn((\mathcal{B} * \alpha) \cup \{\beta\}) \subseteq \mathcal{B} * (\alpha \wedge \beta)$ . Conjunction 2

The two-step view of revision according to which it consists in first contracting the old belief set by the negation of the new information, thus making it consistent with the new information (as well as everything implied by the new information), and then expanding it by adding the new information conjunctively is described by the *Levi identity* (Levi 1977):

$$\mathcal{B} * \alpha = Cn(\mathcal{B} \dot{-} \neg\alpha \cup \{\alpha\})$$

The Levi identity puts contraction  $\dot{-}$  center stage of the revision process. Contraction  $\dot{-}$  turns each old belief set  $\mathcal{B} \subseteq \mathcal{L}$  and sentence  $\alpha \in \mathcal{L}$  into a reduced belief set  $\mathcal{B} \dot{-} \alpha$  that is cleared from  $\alpha$  (as well as everything implying  $\alpha$ ) and required to satisfy the following postulates. For all  $\mathcal{B} \subseteq \mathcal{L}$  and all  $\alpha, \beta \in \mathcal{L}$

- |  |               |
|--|---------------|
| ÷1. $\mathcal{B} \dot{-} \alpha = Cn(\mathcal{B} \dot{-} \alpha)$ .  | Closure       |
| ÷2. If $\alpha \notin Cn(\emptyset)$ , then $\alpha \notin Cn(\mathcal{B} \dot{-} \alpha)$ .   | Success       |
| ÷3. $\mathcal{B} \dot{-} \alpha \subseteq Cn(\mathcal{B})$ .   | Inclusion     |
| ÷4. If $\alpha \notin Cn(\mathcal{B})$ , then $\mathcal{B} \dot{-} \alpha = \mathcal{B}$ .   | Vacuity       |
| ÷5. If $Cn(\{\alpha\}) = Cn(\{\beta\})$ , then $\mathcal{B} \dot{-} \alpha = \mathcal{B} \dot{-} \beta$ .  | Congruence    |
| ÷6. $Cn(\mathcal{B}) \subseteq Cn((\mathcal{B} \dot{-} \alpha) \cup \{\alpha\})$ .   | Recovery      |
| ÷7. $(\mathcal{B} \dot{-} \alpha) \cap (\mathcal{B} \dot{-} \beta) \subseteq \mathcal{B} \dot{-} (\alpha \wedge \beta)$ .                                  | Conjunction 1 |
| ÷8. If $\alpha \notin \mathcal{B} \dot{-} (\alpha \wedge \beta)$ , then $\mathcal{B} \dot{-} (\alpha \wedge \beta) \subseteq \mathcal{B} \dot{-} \alpha$ . | Conjunction 2 |

The Levi identity turns each contraction operator  $\dot{-}$  satisfying ÷1 – ÷8 into a revision operator  $*$  satisfying \*1 – \*8. The converse is true of the *Harper identity* (Harper 1976):

$$\mathcal{B} \dot{-} \alpha = \mathcal{B} * \neg\alpha \cap Cn(\mathcal{B})$$

Given a fixed belief set  $\mathcal{B}$  we can use an entrenchment ordering  $\preceq$  for  $\mathcal{B}$  to define a revision operator  $*$  as follows:

$$\mathcal{B} * \alpha = Cn(\{\beta \in \mathcal{L} : \neg\alpha < \beta\} \cup \{\alpha\}),$$

where  $\alpha < \beta$  holds just in case  $\alpha \preceq \beta$  and  $\beta \not\preceq \alpha$ .

The idea behind this equation is the following: when the agent revises  $*$  her old belief set  $\mathcal{B}$  by the new information  $\alpha$  she first has to clear  $\mathcal{B}$  from  $\neg\alpha$  as well as everything else that is as entrenched as, or less entrenched than,  $\neg\alpha$ . For instance,  $\mathcal{B}$  also has to be cleared from everything implying  $\neg\alpha$ . However, all sentences  $\beta$  from the agent's old belief set  $\mathcal{B}$  that are more entrenched than  $\neg\alpha$  can be preserved. This gives us the preserved belief set  $\{\beta \in \mathcal{L} : \neg\alpha < \beta\}$ . Then one adds the new information  $\alpha$  to obtain  $\{\beta \in \mathcal{L} : \neg\alpha < \beta\} \cup \{\alpha\}$ . Finally, one adds everything that is implied by the preserved belief set together with the new information.

Then one can prove

**Theorem 1** *Let  $\mathcal{L}$  be a formal language and let  $\mathcal{B} \subseteq \mathcal{L}$  be a belief set. For each entrenchment ordering  $\preceq$  for  $\mathcal{B}$  satisfying  $\preceq 1 - \preceq 5$  there is a revision operator  $*$  satisfying \*1 – \*8 such that  $\mathcal{B} * \alpha = Cn(\{\beta \in \mathcal{L} : \neg\alpha < \beta\} \cup \{\alpha\})$ . For each revision operator  $*$  satisfying \*1 – \*8 there is an entrenchment ordering  $\preceq$  for  $\mathcal{B}$  satisfying  $\preceq 1 - \preceq 5$  such that  $\mathcal{B} * \alpha = Cn(\{\beta \in \mathcal{L} : \neg\alpha < \beta\} \cup \{\alpha\})$ .*

See Gärdenfors (1988) and Gärdenfors and Makinson (1988).

There is a different way of representing postulates \*1 – \*8 for revision operators  $*$  due to Grove (1988). Similar to Lewis' (1973) theory of counterfactuals it uses systems of spheres defined on a set of possible worlds instead of entrenchment orderings defined on a formal language. A set of possible worlds can be thought of as a set of complete or maximally specific ways the world might be. One approach, used by Grove (1988), is to identify possible worlds with maximally consistent sets of sentences from  $\mathcal{L}$ , i.e. sets of sentences that are consistent, but that become inconsistent as soon as a single new sentence is added. Another approach is to take possible worlds as primitive. For present purposes we do not have to take a stance on this issue and can assume that we are given a set of possible worlds  $W_{\mathcal{L}}$  relative to which we interpret the sentences from  $\mathcal{L}$ .

$\llbracket \alpha \rrbracket = \{w \in W_{\mathcal{L}} : w \models \alpha\}$  is the proposition expressed by the sentence  $\alpha$ , i.e. the set of possible worlds in which  $\alpha$  is true. If we follow Grove (1988) and construct  $W_{\mathcal{L}}$  as the set

of maximally consistent sets of sentences from  $\mathcal{L}$ , then  $w \models \alpha$  holds just in case  $\alpha \in w$ . If possible worlds are taken as primitive,  $\models$  has to be analyzed differently. Either way there may be propositions  $B \subseteq W_{\mathcal{L}}$  that are not expressed by any sentence  $\alpha$  from  $\mathcal{L}$ .  $\llbracket B \rrbracket = \{w \in W_{\mathcal{L}} : w \models \alpha \text{ for all } \alpha \in B\}$  is the proposition expressed by the set of sentences  $B$ . If we follow Grove (1988) and construct  $W_{\mathcal{L}}$  as the set of maximally consistent sets of sentences from  $\mathcal{L}$ , then for each proposition  $B \subseteq W_{\mathcal{L}}$  there is a set of sentences from  $\mathcal{L}$ ,  $t(B)$ , such that  $B = \llbracket t(B) \rrbracket$  (it is  $\bigcap \{A \subseteq \mathcal{L} : A \text{ is maximally consistent and } A \in B\}$ ). This means that every proposition  $B \subseteq W_{\mathcal{L}}$  can be expressed by a set of sentences from  $\mathcal{L}$ . If possible worlds are taken as primitive this has to be stipulated. We assume this stipulation to be made.

Let  $B \subseteq W_{\mathcal{L}}$  be a proposition and let  $\mathbf{S} \subseteq \wp(W_{\mathcal{L}})$  be a set of propositions.  $\mathbf{S}$  is a *system of spheres* in  $W_{\mathcal{L}}$  that is centered on  $B$  if and only if for all  $A, C \subseteq W_{\mathcal{L}}$  and all  $\alpha \in \mathcal{L}$

- S1 If  $A, C \in \mathbf{S}$ , then  $A \subseteq C$  or  $C \subseteq A$ . S is nested
- S2  $B \in \mathbf{S}$ , and: if  $A \in \mathbf{S}$ , then  $B \subseteq A$ . S is centered on B
- S3  $W_{\mathcal{L}} \in \mathbf{S}$ .
- S4 If  $\llbracket \alpha \rrbracket \cap D \neq \emptyset$  for some  $D \in \mathbf{S}$ , then there is  $D^* \in \mathbf{S}$  such that  $\llbracket \alpha \rrbracket \cap D^* \neq \emptyset$ , and  $D^* \subseteq E$  for all  $E \in \mathbf{S}$  with  $\llbracket \alpha \rrbracket \cap E \neq \emptyset$ .<sup>1</sup>

In combination with S3 requirement S4 says that for each consistent sentence  $\alpha$  there is a smallest sphere  $D^* \in \mathbf{S}$  intersecting  $\llbracket \alpha \rrbracket$ . Let  $c_{\mathbf{S}}(\alpha) := \llbracket \alpha \rrbracket \cap D^*$  and define  $c_{\mathbf{S}}(\alpha) = \emptyset$  if  $\alpha$  is inconsistent.  $c_{\mathbf{S}}(\alpha)$  is the set of  $\alpha$ -worlds that are “closest” to the center  $B$ , where closeness is spelt out in terms of the system of spheres  $\mathbf{S}$ . If  $\alpha$  is consistent with (a set of sentences expressing) the center  $B$ , then  $c_{\mathbf{S}}(\alpha)$  is just the intersection of the center  $B$  with the set of  $\alpha$ -worlds,  $\llbracket \alpha \rrbracket \cap B$ . Otherwise one has to leave the center and move to the first sphere  $D^*$  that intersects with the proposition expressed by  $\alpha$  and take their intersection,  $\llbracket \alpha \rrbracket \cap D^*$ , as  $c_{\mathbf{S}}(\alpha)$ .

Given a fixed belief set  $\mathcal{B}$  we can use a system of spheres  $\mathbf{S}$  centered on  $\llbracket \mathcal{B} \rrbracket$  to define a revision operator  $*$  as follows:

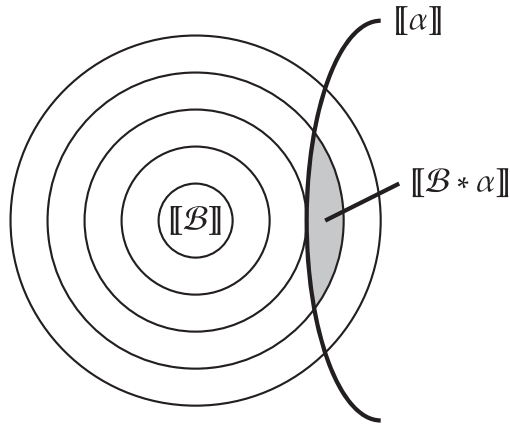
$$\mathcal{B} * \alpha = t(c_{\mathbf{S}}(\alpha))$$

The idea is that what you end up believing after revising  $*$  your old beliefs  $\mathcal{B}$  with the new information  $\alpha$  is (a set of sentences expressing) the proposition  $c_{\mathbf{S}}(\alpha)$  that contains the  $\alpha$ -worlds that are closest when the proposition expressed by your old beliefs,  $\llbracket \mathcal{B} \rrbracket$ , is the center. Expansion is the special case where the proposition expressed by the new information intersects the proposition expressed by the old beliefs,  $\llbracket \alpha \rrbracket \cap \llbracket \mathcal{B} \rrbracket \neq \emptyset$ . In the general case of revision this intersection may be empty so that one has to leave the innermost sphere  $\mathcal{B}$  and move to the smallest sphere  $D^*$  that intersects with  $\llbracket \alpha \rrbracket$ .

Grove (1988) proves

**Theorem 2** *Let  $\mathcal{L}$  be a formal language, let  $\mathcal{B} \subseteq \mathcal{L}$  be a belief set, and let  $W_{\mathcal{L}}$  be a set of possible worlds relative to which the sentences from  $\mathcal{L}$  are interpreted. For each system of spheres  $\mathbf{S}$  in  $W_{\mathcal{L}}$  that is centered on  $\llbracket \mathcal{B} \rrbracket$  and satisfies S1–S4 there is a revision operator  $*$  satisfying  $*1 - *8$  such that  $\mathcal{B} * \alpha = t(c_{\mathbf{S}}(\alpha))$ . For each revision operator  $*$  satisfying  $*1 - *8$  there is a system of spheres  $\mathbf{S}$  in  $W_{\mathcal{L}}$  that is centered on  $\llbracket \mathcal{B} \rrbracket$  and satisfies S1–S4 such that  $\mathcal{B} * \alpha = t(c_{\mathbf{S}}(\alpha))$ .*

We can picture a system of spheres centered on  $\llbracket \mathcal{B} \rrbracket$  as an “onion” around  $\llbracket \mathcal{B} \rrbracket$ . The shaded area  $\llbracket \mathcal{B} * \alpha \rrbracket$  is the strongest proposition the agent believes after revising her old belief set  $\mathcal{B}$  by the new information  $\alpha$ :



In the AGM theory of belief revision the ideal doxastic agent’s old beliefs are represented by her belief set  $\mathcal{B}$  together with her entrenchment ordering  $\preceq$  for  $\mathcal{B}$ . The latter ordering guides the revision process in that it specifies which elements of the old belief set are given up and which are kept when new information  $\delta$  is received. The result of revising the old belief set by new information is a new belief set  $\mathcal{B} * \delta$ . Sophia’s old belief set  $\mathcal{B}$  includes the beliefs that it will rain on Tuesday, that it will be sunny on Wednesday, and that weather forecasts are always reliable. Suppose her belief  $\alpha$  that it will be sunny on Wednesday is less entrenched than her belief  $\beta$  that it will rain on Tuesday, which in turn is less entrenched than her belief  $\gamma$  that weather forecasts are always reliable,  $\alpha < \beta < \gamma$ .

On Monday Sophia learns that the weather forecast for Tuesday and Wednesday predicts rain,  $\delta$ . Consequently she has to give up her belief  $\alpha$  that it will be sunny on Wednesday or her belief  $\gamma$  that weather forecasts are always reliable. The reason is that it follows from  $\delta$  that one of those two beliefs is false,  $\neg\alpha \vee \neg\gamma \in Cn(\{\delta\})$ . This implies that  $\alpha \wedge \gamma \preceq \neg\delta$ . Since  $\beta \in Cn(\{\gamma, \delta\})$  there is no need for Sophia to give up her belief  $\beta$  that it will rain on Tuesday, and we assume  $\neg\delta < \beta$  so that Sophia’s entrenchment ordering looks as follows:  $\alpha \sim \alpha \wedge \gamma \preceq \neg\delta < \beta < \gamma$ , where  $x \sim y$  holds just in case  $x \preceq y$  and  $y \preceq x$ . Thus Sophia’s new belief set is  $\mathcal{B} * \delta = Cn(\{x : \neg\delta < x\} \cup \{\delta\}) = Cn(\{\beta, \gamma, \delta, \neg\alpha\})$ .

To Sophia’s surprise it is sunny on Tuesday after all. Therefore Sophia wants to revise her newly acquired belief set  $\mathcal{B} * \delta$  a second time by  $\neg\beta$  to correct her belief  $\beta$  that it will rain on Tuesday. In addition Sophia has to give up her belief  $\delta$  that the weather forecast for Tuesday and Wednesday predicts rain (she might have misheard the weather forecast, say) or her belief  $\gamma$  that weather forecasts are always reliable (she might have been too gullible, say). The reason is that it follows from  $\neg\beta$  that one of those two beliefs is false,  $\neg\delta \vee \neg\gamma \in Cn(\neg\beta)$ . Unfortunately AGM belief revision theory is of no help here. While Sophia could use her entrenchment ordering to revise her old belief set  $\mathcal{B}$  to a new belief set  $\mathcal{B} * \delta$ , the entrenchment ordering itself has not been updated and so is silent as to whether  $\delta$  is more entrenched than  $\gamma$  (she was too gullible, say) or  $\gamma$  is more entrenched than  $\delta$  (she misheard the weather forecast, say) or  $\gamma$  is as entrenched as  $\delta$  (she was too gullible and misheard the weather forecast, say). This, however, is exactly the kind of information that Sophia would need to revise her beliefs a second time.

### 3. Iterated Belief Revision

More generally, the problem is that Sophia's beliefs are represented as a belief set plus an entrenchment ordering before the revision process, but as a belief set without an entrenchment ordering after the revision process. To handle *iterated* belief revisions the ideal doxastic agent's beliefs have to be represented in the same way before and after the revision process. Gärdenfors and Rott (1995: 37) call this the "principle of categorical matching".

Nayak (1994), Boutilier (1996), Darwiche and Pearl (1997), Segerberg (1998) and Fermé (2000), Rott (2003), Rott (2006), and others do exactly this. They augment the AGM postulates by further postulates specifying how the ideal doxastic agent should revise her entrenchment ordering in addition to her belief set when she receives new information. On their accounts, the agent's beliefs are represented as a belief set plus an entrenchment ordering both before and after the revision process and both of these two elements are revised when new information is received.

Let us have a closer look at the proposal by Darwiche and Pearl (1997). In addition to \*1 – \*8 they propose the following four postulates for iterated belief revision. For all  $\mathcal{B} \subseteq \mathcal{L}$  and all  $\alpha, \beta \in \mathcal{L}$

- \*9. If  $\beta \in Cn(\{\alpha\})$ , then  $(\mathcal{B} * \beta) * \alpha = \mathcal{B} * \alpha$ .
- \*10. If  $\neg\beta \in Cn(\{\alpha\})$ , then  $(\mathcal{B} * \beta) * \alpha = \mathcal{B} * \alpha$ .
- \*11. If  $\beta \in \mathcal{B} * \alpha$ , then  $\beta \in (\mathcal{B} * \beta) * \alpha$ .
- \*12. If  $\neg\beta \notin \mathcal{B} * \alpha$ , then  $\neg\beta \notin (\mathcal{B} * \beta) * \alpha$ .

In order to perspicuously represent these four postulates it is helpful to consider the following reformulation of a system of spheres  $\mathbf{S}$  in  $W_{\mathcal{L}}$  centered on  $B$ . Let  $B \subseteq W_{\mathcal{L}}$  be a proposition and let  $\leq$  be a binary relation on  $W_{\mathcal{L}}$ .  $\leq$  is a plausibility ordering on  $W_{\mathcal{L}}$  with center  $B$  just in case for all  $w, w', w'' \in W_{\mathcal{L}}$  and  $A \subseteq W_{\mathcal{L}}$ :

- $\leq 1$   $w \leq w'$  or  $w' \leq w$ .  $\leq$  is connected
- $\leq 2$  If  $w \leq w'$  and  $w' \leq w''$ , then  $w \leq w''$ .  $\leq$  is transitive
- $\leq 3$   $w \in B$  if and only if  $\forall w^* : w \leq w^*$ .
- $\leq 4$  If  $A \neq \emptyset$ , then  $\{v \in A : v \leq w^* \forall w^* \in A\} \neq \emptyset$ .

So a system of spheres centered on  $B$  can be understood as a plausibility ordering with the center  $B$  being the set of most plausible possible worlds.

In terms of such a plausibility ordering the problem with the original AGM approach is the following. Before the revision process the agent's beliefs are represented as a belief set  $\mathcal{B}$  plus a plausibility ordering  $\leq_{\mathcal{B}}$  with center  $\llbracket \mathcal{B} \rrbracket$ . After the revision process the agent's beliefs are represented as a belief set  $\mathcal{B} * \alpha$ , but without a corresponding plausibility ordering  $\leq_{\mathcal{B} * \alpha}$ . The principle of categorical matching urges us to represent the agent's beliefs as a belief set plus a plausibility ordering both before and after the revision process. In these terms postulates \*9 – \*12 say the following:

- $\leq 5$ . If  $w, w' \in \llbracket \alpha \rrbracket$ , then  $w \leq_{\mathcal{B}'} w'$  just in case  $w \leq_{\mathcal{B} * \alpha} w'$ .
- $\leq 6$ . If  $w, w' \notin \llbracket \alpha \rrbracket$ , then  $w \leq_{\mathcal{B}'} w'$  just in case  $w \leq_{\mathcal{B} * \alpha} w'$ .
- $\leq 7$ . If  $w \in \llbracket \alpha \rrbracket$  and  $w' \notin \llbracket \alpha \rrbracket$ , and if  $w <_{\mathcal{B}'} w'$ , then  $w <_{\mathcal{B} * \alpha} w'$ .
- $\leq 8$ . If  $w \in \llbracket \alpha \rrbracket$  and  $w' \notin \llbracket \alpha \rrbracket$ , and if  $w \leq_{\mathcal{B}'} w'$ , then  $w \leq_{\mathcal{B} * \alpha} w'$ .

Here  $w < w'$  holds just in case  $w \leq w'$  and  $w' \not\leq w$ .  
Grove (1988) proves

**Theorem 3** *Let  $\mathcal{L}$  be a formal language, let  $\mathcal{B} \subseteq \mathcal{L}$  be a belief set, and let  $W_{\mathcal{L}}$  be a set of possible worlds relative to which the sentences from  $\mathcal{L}$  are interpreted. For each plausibility ordering  $\leq$  on  $W_{\mathcal{L}}$  with center  $\llbracket \mathcal{B} \rrbracket$  that satisfies  $\leq 1 - \leq 4$  there is a revision operator  $*$  satisfying  $* 1 - * 8$  such that  $\mathcal{B} * \alpha = t(\{v \in \llbracket \alpha \rrbracket : v \leq w \forall w \in \llbracket \alpha \rrbracket\})$ . For each revision operator  $*$  that satisfies  $* 1 - * 8$  there is a plausibility ordering  $\leq$  on  $W_{\mathcal{L}}$  with center  $\llbracket \mathcal{B} \rrbracket$  satisfying  $\leq 1 - \leq 4$  such that  $\mathcal{B} * \alpha = t(\{v \in \llbracket \alpha \rrbracket : v \leq w \forall w \in \llbracket \alpha \rrbracket\})$ .*

Darwiche and Pearl (1997) prove

**Theorem 4** *Let  $\mathcal{L}$  be a formal language, let  $\mathcal{B} \subseteq \mathcal{L}$  be a belief set, and let  $W_{\mathcal{L}}$  be a set of possible worlds relative to which the sentences from  $\mathcal{L}$  are interpreted. Suppose  $*$  satisfies  $* 1 - * 8$  and  $\leq_{\mathcal{B}}$  is the corresponding plausibility ordering on  $W_{\mathcal{L}}$  with center  $\llbracket \mathcal{B} \rrbracket$  such that  $\leq 1 - \leq 4$  and  $\mathcal{B} * \alpha = t(\{v \in \llbracket \alpha \rrbracket : v \leq w \forall w \in \llbracket \alpha \rrbracket\})$ . Then  $*$  satisfies  $* 9 - * 12$  just in case  $\leq_{\mathcal{B}}$  satisfies  $\leq 5 - \leq 8$ .*

These approaches to iterated belief revision, of which Darwiche and Pearl (1997) is just one among many, have in common that the agent's beliefs are represented as a belief set plus an entrenchment ordering/system of spheres/plausibility ordering both before and after the revision process. They also have in common that the new information is represented as a single sentence or a single proposition. The latter is also true for the approach by Jin and Thielscher (2007) discussed below, but different for what Rott (2009) calls "two-dimensional" belief revision operators (see Cantwell 1997 and Fermé and Rott 2004 and Rott 2007).

In "one-dimensional" belief revision the new information comes as a "naked" (Rott 2007) sentence or proposition, and it is the job of the various belief revision methods – lexicographic revision (Nayak 1994), natural revision (Boutilier 1996), irrevocable revision (Seegerberg 1998, Fermé 2000), irrefutable revision (Rott 2006), and others – to say where exactly in the new entrenchment ordering/system of spheres/plausibility ordering that sentence or proposition should be placed. In two-dimensional belief revision it is the new information itself that carries this information, at least to some extent. Here the new information does not merely say that the input sentence  $\alpha$  should be accepted.<sup>2</sup> It additionally specifies how firmly  $\alpha$  should be accepted by requiring  $\alpha$  to be (at least) as entrenched as some "reference sentence"  $\beta$  so that the new information is of the form  $\beta \preceq \alpha$ .

Let us return to our example. On Monday Sophia learns that the weather forecast for Tuesday and Wednesday predicts rain,  $\delta$ . In one-dimensional belief revision she picks one of the iterated belief revision methods and revises her old belief set  $\mathcal{B}$  and entrenchment ordering  $\preceq_{\mathcal{B}}$  to obtain a new belief set  $\mathcal{B} * \delta$  and entrenchment ordering  $\preceq_{\mathcal{B} * \delta}$ . Different methods return different outputs, but on all of them she ends believing that it will rain on Tuesday,  $\beta$ . On Tuesday she sees that is sunny,  $\neg\beta$ . In one-dimensional belief revision Sophia proceeds as before. In two-dimensional belief revision Sophia does not merely receive the qualitative information  $\neg\beta$ , but the comparative information  $\neg\gamma \preceq \neg\beta$ , indicating that she trusts her sight at least as much as the weather forecast. While there still are several belief revision methods to choose from (Rott 2009) – among others reflecting the fact that Sophia can respect  $\neg\gamma \preceq \neg\beta$  by lowering the doxastic status of  $\neg\gamma$  or by raising the doxastic status of  $\neg\beta$  – the new information now is more specific and leaves less room to be filled by the revision method. It is then only a small step to equip Sophia with the quantitative, numerical information that  $\neg\beta$  is entrenched to a specific degree. In that case the new

information determines where  $\neg\beta$  is located in the new entrenchment ordering. The revision method merely has to incorporate this into Sophia's old beliefs in a coherent way. We will take this step in part II.

Before doing so let us briefly return to the qualitative approaches to iterated belief revision. As Jin and Thielscher (2007) point out the accounts by Darwiche and Pearl (1997) and others are still too permissive in the sense that beliefs are given up unnecessarily. As a consequence they additionally require the ideal doxastic agent to consider the new information  $\beta$  independent of a sentence  $\alpha$  after revision by  $\beta$  if she did so before. While the idea behind Jin and Thielscher's (2007) proposal seems to be correct, their actual proposal is too strong. The reason is that their notion of dependence is too strong in the sense that too many sentences are rendered independent. According to them a believed sentence  $\alpha$  is independent of another sentence  $\beta$  if the believed sentence  $\alpha$  is still believed after revision by the negation of the other sentence,  $\neg\beta$ . But I can receive new information  $\neg\beta$ , say, that the captain of my favorite football will not be fit for the match, whose negation  $\beta$  is positively relevant to, and so *not* independent of, a belief of mine  $\alpha$ , say, that my favorite football team will win the match, without making me give up that belief of mine altogether. More generally, the ways two sentences can depend on each other are many and varied, and the qualitative and comparative notions of AGM belief revision theory seem to be too coarse-grained to capture those. Hild and Spohn (2008) argue axiomatically that, in order to adequately represent all dependencies, and to handle indefinitely iterated belief revisions, one has to go all the way to quantitative, numerical ranking functions.

### Acknowledgements

I am grateful to Benjamin Bewersdorf, Jennifer Nagel, Hans Rott, Sherrilyn Roush, Wolfgang Spohn, Jonathan Weisberg, and an anonymous referee for helpful comments on earlier versions of these two papers. Robert Lehnert helped me with the design of the graphics. My research was in part supported by the German Research Foundation through its Emmy Noether Program.

### Short Biography

Franz Huber is an assistant professor in the Department of Philosophy at the University of Toronto. His research interests are in formal epistemology, general philosophy of science, and philosophical logic.

### Notes

\*Correspondence: Franz Huber, Department of Philosophy, University of Toronto, Canada. Email: franz.huber@utoronto.ca

<sup>1</sup> In one respect Grove's (1988) notion of a system of spheres is more general than Lewis' (1973) notion as it allows  $\mathbf{S}$  to be centered on arbitrary propositions  $B \subseteq W_{\mathcal{L}}$ , whereas Lewis (1973: 14f) requires the center  $B$  to contain the actual world (and nothing but the actual world). In another respect Grove's (1988) notion is less general than that of Lewis' (1973) as requirement  $\mathbf{S4}$  effectively makes the limit assumption, which Lewis (1973: 19f) famously rejects.

<sup>2</sup> For present purposes we can ignore the case of "non-prioritized" belief revision, where the new information need not be accepted. See Hansson, Fermé, Cantwell, Falappa (2001).

### Works Cited

- Alchourrón, Carlos E & Gärdenfors, Peter & Makinson, David. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic* 50 (1985): 510–530.
- Boutilier, Craig. Iterated Revision and Minimal Change of Conditional Beliefs. *Journal of Philosophical Logic* 25 (1996): 263–305.



- Cantwell, John. On the Logic of Small Changes in Hypertheories. *Theoria* 63 (1997): 54–89.
- Darwiche, Adnan & Pearl, Judea. On the Logic of Iterated Belief Revision. *Artificial Intelligence* 89 (1997): 1–29.
- Easwaran, Kenny. Bayesianism I: Introduction and Arguments in Favor. *Philosophy Compass* 6 (2011a): 312–320.
- . Bayesianism II: Applications and Criticisms. *Philosophy Compass* 6 (2011b): 321–332.
- Fermé, Eduardo. Irrevocable Belief Revision and Epistemic Entrenchment. *Logic Journal of IGPL* 8 (2000): 645–652.
- Fermé, Eduardo & Rott, Hans. Revision by Comparison. *Artificial Intelligence* 157 (2004): 5–47.
- Gärdenfors, Peter. *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. Cambridge, MA: MIT Press, 1988.
- Gärdenfors, Peter & Makinson, David. Revisions of Knowledge Systems Using Epistemic Entrenchment. *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*. San Francisco: Morgan Kaufmann, (1988): 83–95.
- Gärdenfors, Peter & Rott, Hans. Belief Revision. *Epistemic and Temporal Reasoning. Handbook of Logic in Artificial Intelligence and Logic Programming: Volume 4*. Ed. D.M. Gabbay & C.J. Hogger & J.A. Robinson. Oxford: Clarendon Press, (1995): 35–132.
- Grove, Adam. Two Modellings for Theory Change. *Journal of Philosophical Logic* 17 (1988): 157–170.
- Hansson, Sven O & Fermé, Eduardo & Cantwell, John & Falappa, Marcelo A. Credibility-Limited Revision. *Journal of Symbolic Logic* 66 (2001): 1581–1596.
- Harper, William L. Rational Conceptual Change. *PSA 1976* 2 (1976): 462–494.
- Hild, Matthias & Spohn, Wolfgang. The Measurement of Ranks and the Laws of Iterated Contraction. *Artificial Intelligence* 172 (2008): 1195–1218.
- Jin, Yi & Thielscher, Michael. Iterated Belief Revision, Revised. *Artificial Intelligence* 171 (2007): 1–18.
- Levi, Isaac. Subjunctives, Dispositions and Chances. *Synthese* 34 (1977): 423–455.
- Lewis, David K. *Counterfactuals*. Cambridge, MA: Harvard University Press, 1973.
- Nayak, Abhaya C. Iterated Belief Change Based on Epistemic Entrenchment. *Erkenntnis* 41 (1994): 353–390.
- Rott, Hans. *Change, Choice, and Inference. A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford: Oxford University Press, 2001.
- . Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence. *Journal of Logic and Computation* 13 (2003): 111–145.
- . Revision by Comparison as a Unifying Framework: Severe Withdrawal, Irrevocable Revision and Irrefutable Revision. *Theoretical Computer Science* 355 (2006): 228–242.
- . Two-dimensional Belief Change: An Advertisement. In G. Bonanno & J. Delgrande & J. Lang & H. Rott (eds.), *Formal Models of Belief Change in Rational Agents. Dagstuhl Seminar Proceedings* 07351 (2007).
- . Shifting Priorities: Simple Representations for Twenty-seven Iterated Theory Change Operators. *Towards Mathematical Philosophy. Trends in Logic* 28. Ed. D. Makinson & J. Malinowski & H. Wansing. Dordrecht: Springer, (2009): 269–296.
- Seegerberg, Krister. Irrevocable Belief Revision in Dynamic Doxastic Logic. *Notre Dame Journal of Formal Logic* 39 (1998): 287–306.