Bounded revision: Two-dimensional belief change between conservative and moderate revision

Hans Rott

Department of Philosophy, University of Regensburg
hans.rott@psk.uni-r.de

Abstract

In this paper I present the model of 'bounded revision' that is based on two-dimensional revision functions taking as arguments pairs consisting of an input sentence and a reference sentence. The key idea is that the input sentence is accepted as long as (and just a little longer than) the reference sentence is 'cotenable' with it. Bounded revision satisfies the AGM axioms as well as the Same Beliefs Condition (SBC) saying that the set of beliefs accepted after the revision do not depend on the reference sentence (although the posterior belief state does depend on it). Bounded revision satisfies the Darwiche-Pearl (DP) axioms. If the reference sentence is fixed to be a tautology or a contradiction, two well-known one-dimensional revision operations result; bounded revision thus naturally fills the space between conservative revision (also known as natural revision) and moderate revision (also known as lexicographic revision). I compare this approach to the two-dimensional model of 'revision by comparison' investigated by Fermé and Rott (*Artificial Intelligence* 157, 2004) that satisfies neither SBC nor the DP axioms. I conclude that two-dimensional revision operations add substantially to the expressive power of qualitative approaches that do not make use of numbers as measures of degrees of belief.

1. Introduction

Representations of belief states in terms of probability functions or ranking functions are very rich and powerful.[1] However, it is often hard to come by meaningful numbers. Qualitative belief change in the style of Alchourrón, Gärdenfors and Makinson (1985, henceforth 'AGM') and its extensions to iterated belief change in the 1990s, on the other hand, are simple and do not need numbers, but are a lot more restricted in their expressive power. Two-dimensional belief revision attempts to strike a good balance between the advantages of quantitative and AGM-style qualitative approaches.

[1]This paper is based on work first presented in Rott (2007a) and (2007b).

Fermé and Rott (2004) suggested a basically qualitative approach that is more flexible than AGM style models in that it allows a new piece of information to be accepted in various degrees or strengths.[2] The key idea of their *Revision by comparison* (henceforth, *RbC*) is that an input sentence α does not come with a number, but does not come 'naked' either. It rather comes with a reference sentence δ that typically expresses an antecedently held belief. The agent is then supposed to follow an instruction of the form

'Accept α with a strength that at least equals that of δ.'

If the reference sentence δ is strong enough, or more precisely, if it is more entrenched than the negation $\neg\alpha$ of the input sentence α, then RbC yields an AGM revision of the initial belief set. If, however, the reference sentence is weaker, than it gets lost, and what we get is not a successful revision by α but a severe withdrawal (Pagnucco and Rott 1999) of the reference sentence δ.

A drawback of the RbC approach of Fermé and Rott is that it does not satisfy the Darwiche-Pearl postulates for iterated belief change (Darwiche and Pearl 1997). These postulates have a very appealing possible worlds semantics that strongly suggests that they should be satisfied by iterated revision functions (compare Section 2.2 below).

Bounded revision is motivated by the same concerns as RbC, combined with the desire to satisfy the Darwiche-Pearl postulates. Although the reference sentence δ functions here as a measure of how firmly entrenched α should be in the agent's posterior belief state, the idea of bounded revision can be expressed more precisely by the following recipe:

'Accept α as long as δ holds along with α, and just a little more.'

In a way, δ serves as a bound for the acceptance of α. Intuitively, we can think of the function of the reference sentence δ in two ways. First, we may suppose that it is a marker delineating the shape of a sphere in a Grovean system of spheres that characterizes the reasoner's initial belief state. Second, δ can be used as a sentence that should hold throughout a range of relatively plausible situations in which α holds, without necessarily requiring that δ is a prior belief. We will argue that the latter option is preferable for bounded revision.

Thus any arbitrary sentence δ may sensibly serve as the parameter sentence for a bounded revision. However, the paradigm cases are those in which δ is cotenable with α to some extent, in the sense that a stretch of the comparatively plausible ways of making α true are all ways that make δ true as well.[3] Not only

[2] An approach similar to revision by comparison was introduced earlier in Cantwell's (1997) 'raising' operation. Cantwell has also presented an interesting dual operation that he calls 'lowering'.

[3] The notion of cotenability is due to, and made comparatively precise by, Goodman (1955,

the *most* plausible α-worlds, but also all those that are *sufficiently* plausible are moved center stage in a bounded revision by α, and it is precisely the task of the reference sentence δ to indicate what is meant by 'sufficient' (compare Fig. 1 below). The greater the stretch throughout which δ holds along with α, the firmer α gets accepted by a revision that is bounded by δ.

Usually the intended cases of belief revision are those in which the input sentence α is not believed prior to the revision. However, two-dimensional revision may well be used to increase the strength or entrenchment of a sentence that the agent has already believed to be true prior to the revision.[4]

2. Generalizing AGM to two-dimensional revisions of belief states

What is a belief state? For the purposes of this paper, we may think of belief states as entities of any type whatsoever, neural states, holistic mental states, abstract machine states, etc. We assume that the set of beliefs of a reasoner, but not necessarily his or her whole belief state is epistemically accessible. The beliefs are, so to speak, the visible tip of the iceberg that itself remains concealed from our eyes and, perhaps, from the reasoner's own eyes as well. Our hypothesis is that belief states have a rich structure that determines the development of the agent's belief set in response to any sequence of inputs. We do not exclude that it determines more, but this is what we are interested in. In Section 3, we shall specify concrete formal structures as representations of belief states that contain a lot more structure than a plain belief set, but are still abstractions from 'real' belief states. Before doing that, however, we address the problem of belief *change* in abstraction from any particular conceptualization of belief states.

2.1. One-dimensional and two-dimensional belief revision operators

A *one-dimensional* belief revision operation is a function $*$ that takes a belief state $\mathcal{B}$ and an input sentence α and returns the new belief state. $*(\mathcal{B}, \alpha)$ denotes the state $\mathcal{B}$ revised by α. A *two-dimensional* revision function is similar, except that the input is a pair of sentences $\langle \alpha, \delta \rangle$. The first sentence is the *input sentence*, the second sentence the *reference sentence*. The sentences can vary fully independently from each other, thus the name 'two-dimensional.' Usually, I will use the variables α, β and γ etc. for input sentences, and the variables δ,

p. 15). A semantic analysis of two *other* concepts of cotenability was given by Lewis (1973, pp. 57, 69–70), though Lewis didn't tell that his concepts are different from Goodman's. Unfortunately, my use of the term is yet different, but it is the best term for my purposes that I can think of.

[4] This is the main idea underlying Cantwell's approach mentioned in footnote 2.

ε, ζ etc. for reference sentences.[5] As is common in theories of one-dimensional revision functions $*$, we write $\mathcal{B} * \alpha$ for $*(\mathcal{B}, \alpha)$. For a two-dimensional revision function $*$, we write $\mathcal{B} *_\delta \alpha$ for $*(\mathcal{B}, \langle\alpha, \delta\rangle)$.

We work with a finitary propositional language $\mathcal{L}$ with finitely many propositional variables. In addition to the usual propositional connectives, we use the sentential constants $\top$ (TRUTH) and $\bot$ (FALSITY). The set of possible worlds (interpretations, models) and the set of sets of logically equivalent sentences are supposed to be finite, too.[6] We use Cn to indicate a consequence operation governing $\mathcal{L}$. We suppose throughout this paper that the logic is Tarskian, that it includes classical propositional logic, and that it satisfies the deduction theorem.[7] The only inconsistent and logically closed set of sentences in the language is the set of all sentences which we also denote by $\mathcal{L}$.

Notation: For any belief state $\mathcal{B}$, $\ulcorner\mathcal{B}\urcorner$ is the set of beliefs held by a person in belief state $\mathcal{B}$ (more exactly: the beliefs that can be ascribed to the person, or the beliefs that the person is committed to). We assume that $\ulcorner\mathcal{B}\urcorner$ is logically closed. If $\mathcal{B}$ and $\mathcal{B}'$ are two belief states, then we sometimes write $\mathcal{B} \simeq \mathcal{B}'$ as an abbreviation for $\ulcorner\mathcal{B}\urcorner = \ulcorner\mathcal{B}'\urcorner$.

The main benchmark in theory change are the famous AGM postulates for one-step belief revision (AGM 1985). We rewrite them in a new notation that makes explicit that belief revision (i) is really about the revision of belief states rather than belief sets and (ii) is more generally conceived as two-dimensional rather than one-dimensional.

(AGM1) $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ is logically closed

(AGM2) $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ contains α

(AGM3) $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ is a subset of $Cn(\ulcorner\mathcal{B}\urcorner \cup \{\alpha\})$

(AGM4) If α is consistent with $\ulcorner\mathcal{B}\urcorner$, then $\ulcorner\mathcal{B}\urcorner$ is a subset of $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$

(AGM5) If α is consistent, then $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ is consistent

[5] The terminology of input and reference sentences is taken over from Fermé and Rott (2004). The overused epithet 'two-dimensional' is certainly not an ideal name, but all the friendly alternative suggestions I have received from colleagues (thanks!) have their own disadvantages and so I stick to this name.

[6] We presuppose finiteness mainly as a matter of convenience, in order not to burden this paper with technical details distracting us from the main issues. An infinite language would not complicate things as long as we work with entrenchment relations, but when working with systems of spheres, infinity complicates the matter enormously. See, e.g., Pagnucco and Rott (1999, Section 8).

[7] By saying that the logic Cn is Tarskian, we mean that it is reflexive ($\Gamma \subseteq Cn(\Gamma)$), monotonic (if $\Gamma \subseteq \Gamma'$, then $Cn(\Gamma) \subseteq Cn(\Gamma')$), idempotent ($Cn(Cn(\Gamma)) \subseteq Cn(\Gamma)$) and compact (if $\alpha \in Cn(\Gamma)$, then $\alpha \in Cn(\Gamma')$ for some finite $\Gamma' \subseteq \Gamma$). The deduction theorem says that $\alpha \rightarrow \beta \in Cn(\Gamma)$ if and only if $\beta \in Cn(\Gamma \cup \{\alpha\})$. We also write $\Gamma \vdash \alpha$ for $\alpha \in Cn(\Gamma)$.

(AGM6) If α is logically equivalent with β, then $\ulcorner\mathcal{B} *_\delta \alpha\urcorner = \ulcorner\mathcal{B} *_\delta \beta\urcorner$

(AGM7) $\ulcorner\mathcal{B} *_\delta (\alpha \wedge \beta)\urcorner$ is a subset of $Cn(\ulcorner\mathcal{B} *_\delta \alpha\urcorner \cup \{\beta\})$

(AGM8) If β is consistent with $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$, then $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ is a subset of $\ulcorner\mathcal{B} *_\delta (\alpha \wedge \beta)\urcorner$

To get the AGM postulates for one-dimensional revision operations, just drop the subscript 'δ' from each occurrence of '$*_\delta$'.

From the point of view of the present paper, (AGM5) introduces an unnecessary loss of generality.[8] An agent may consider more sentences than just logical falsehoods as 'absolutely impossible'. And a revision by a doxastically impossible sentence may lead him or her either into an inconsistent belief set (the AGM idea) or into a refusal to change anything (an alternative idea which makes equally good sense).

It is natural to assume that two-dimensional revision functions satisfy the AGM postulates, except for (AGM5). However, Fermé and Rott's Revision by Comparison is an operation that does not satisfy them. The reason is that it has features of belief contraction (belief withdrawal, removal, subtraction, etc.) as well as belief revision. This does not hold for bounded revision which is a pure operation of revision.

(AGM6)–(AGM8) compare the results of two alternative revisions. They focus on a variation in the input sentences, the reference sentence is kept fixed. Once we work in a two-dimensional context, however, it is equally natural to ask about the behaviour with respect to the reference sentences. This was a major topic for Fermé and Rott (2004), because RbC can, in a specified class of cases, be seen as a withdrawal operation with respect to the reference sentence. Bounded revision does not show this kind of duality between input and reference sentences, and the corresponding behaviour is much simpler. First, it appears to be as unproblematic as (AGM6) to stipulate that if δ is logically equivalent with ε, then $\ulcorner\mathcal{B} *_\delta \alpha\urcorner = \ulcorner\mathcal{B} *_\varepsilon \alpha\urcorner$. A second and more interesting question for two-dimensional belief change operations is whether it is possible to vary the reference sentence without changing the set of beliefs obtained. It does not seem implausible to assume that the reference sentence δ only specifies the *extent* or *strength* with which the input sentence α is to be accepted, but does not affect the *content* of the new belief set. The central condition to be discussed is this:

(SBC) $\ulcorner\mathcal{B} *_\delta \alpha\urcorner = \ulcorner\mathcal{B} *_\varepsilon \alpha\urcorner$ for all δ and ε

[8] Recommendations how to weaken (AGM5) in a one-dimensional context are obtained by connecting pp. 149–153, 206 and 118 in Rott (2001). The relevant conditions are: ($*\emptyset1$) If $\bot \in \ulcorner\mathcal{B} * \alpha\urcorner$, then $\bot \in \ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$; and ($*\emptyset2$) If $\bot \in \ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$, then $\neg\beta \in \ulcorner\mathcal{B} * \alpha\urcorner$.

Let us call this condition the *Same Beliefs Condition.* It says that the posterior *belief set* does not depend on the reference sentence. If (SBC) is satisfied, only the *belief state* obtained is sensitive to variations of the reference sentence. We shall see that bounded revision satisfies (SBC). This is quite different from the situation with Fermé and Rott's Revision by Comparison. There the reference sentence, and in particular its relative strength as compared to the negation of the input sentence, does matter. This is again due to the fact that RbC is not a pure operation of revision but has features of belief contraction, too. Restricted versions of (SBC), however, hold for RbC.[9]

2.2. Iterations

How can we get from a belief state $\mathcal{B}$ and an input of the form α or $\langle\alpha,\beta\rangle$ to the revised belief state? A key to understanding much of traditional research in belief revision is that the *belief sets* obtained after potential second revision steps provide all evidence that we need about the structure of the *belief state* the agent is in after the first revision step.

In their seminal paper, Darwiche and Pearl (1997) introduced the following set of constraints for one-dimensional iterated belief change. They are now widely known as the *Darwiche-Pearl postulates.*

(DP^{1}1) If β implies α, then $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner = \ulcorner\mathcal{B} * \beta\urcorner$

(DP^{1}2) If β is inconsistent with α, then $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner = \ulcorner\mathcal{B} * \beta\urcorner$

(DP^{1}3) If α is in $\ulcorner\mathcal{B} * \beta\urcorner$, then α is in $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner$

(DP^{1}4) If $\neg\alpha$ is not in $\ulcorner\mathcal{B} * \beta\urcorner$, then $\neg\alpha$ is not in $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner$

Notice that these postulates make statements only about *belief sets.* But since they concern iterations, they implicitly talk about one-step changes of *belief states* as well. (More about this in Section 3.)

Darwiche and Pearl (1997) showed that these postulates correspond one by one to very appealing semantic constraints in terms of total pre-orderings of possible worlds (interpretations, models). Assume that a belief state is represented by such a pre-ordering and that the new piece of information is α. Then, as already mentioned, the first pair of postulates essentially says that a revision

[9] Here are two important special cases. In Rott (2007b) it is shown that RbC satisfies (SBC) *restricted to successful revisions* and (SBC) *restricted to revisions with reference sentences of equal entrenchment*:

(SBC^{r1}) If α is in both $\ulcorner\mathcal{B} *_\delta \alpha\urcorner$ and $\ulcorner\mathcal{B} *_\varepsilon \alpha\urcorner$, then $\mathcal{B} *_\delta \alpha \simeq \mathcal{B} *_\varepsilon \alpha$

(SBC^{r2}) If neither δ nor ε is in $\ulcorner\mathcal{B} *_{\delta\wedge\varepsilon} \bot\urcorner$, then $\mathcal{B} *_\delta \alpha \simeq \mathcal{B} *_\varepsilon \alpha$

by α should not mess up the pre-ordering *within the α-worlds*, nor should it mess up the pre-ordering *within the $\neg\alpha$-worlds*. The second pair of postulates essentially says that the relative position of an α-world with respect to a $\neg\alpha$-world must not be worse after a revision of the belief state by α.[10] I take it this eminently plausible semantics recommends that the Darwiche-Pearl postulates be obeyed by reasonable iterated belief revision operators.[11]

Let us now adapt the Darwiche-Pearl postulates so as make them applicable to the general case of two-dimensional revision functions.

(DP1) If β implies α, then $(\mathcal{B} *_\delta \alpha) *_\varepsilon \beta \;\simeq\; \mathcal{B} *_\varepsilon \beta$

(DP2) If β is inconsistent with α, then $(\mathcal{B} *_\delta \alpha) *_\varepsilon \beta \;\simeq\; \mathcal{B} *_\varepsilon \beta$

(DP3) If α is in $\ulcorner \mathcal{B} *_\varepsilon \beta \urcorner$, then α is in $\ulcorner (\mathcal{B} *_\delta \alpha) *_\varepsilon \beta \urcorner$

(DP4) If $\neg\alpha$ is not in $\ulcorner \mathcal{B} *_\varepsilon \beta \urcorner$, then $\neg\alpha$ is not in $\ulcorner (\mathcal{B} *_\delta \alpha) *_\varepsilon \beta \urcorner$

A bolder reformulation of the Darwiche-Pearl could take varying reference sentences for different occasions of the revisions by β in each of (DP1)–(DP4). However, in the presence of (SBC), the differences would vanish anyway. After essentially dropping the subscripts to '$*\beta$', however, it becomes clear that under the special conditions regarding α and β, that even δ can be chosen arbitrarily. So if (SBC) is given and we are prepared to tolerate a little sloppiness in notation, we can simply forget about the reference sentences and use the original Darwiche-Pearl formulations even in the context of two-dimensional revision functions.

We shall see that in contrast to RbC which violates (DP2),[12] bounded revision satisfies all the Darwiche-Pearl postulates.

[10] In symbols:

(DPO1) For any two α-worlds w and w', $w \leq w'$ iff $w \leq^*_\alpha w'$.

(DPO2) For any two $\neg\alpha$-worlds w and w', $w \leq w'$ iff $w \leq^*_\alpha w'$.

(DPO3) For any α-world w and $\neg\alpha$-world w', if $w < w'$, then $w <^*_\alpha w'$.

(DPO4) For any α-world w and $\neg\alpha$-world w', if $w \leq w'$, then $w \leq^*_\alpha w'$.

Note that worlds that are smaller according to $\leq$ are *more* plausible, or *closer* to the agent's beliefs than worlds that are greater according to $\leq$. Such ordering relations can also be graphically represented as Grovean systems of spheres (Grove 1988). More on this in Section 3.

[11] Note, however, that the correspondence between the DP postulates with their semantic 'counterparts' depends on the satisfaction of other conditions. Papini (2001, pp. 292–293), for instance, shows that her reverse-lexicographic belief change operator $\circ_\triangleleft$ satisfies all semantic properties, but fails to satisfy (DP1) and (DP2).

[12] This is easily understood from the semantics of RbC which collapses distinctions between some $\neg\alpha$-worlds. See Figure 3 below.

We finally specify a sufficient condition for the satisfaction of the Darwiche-Pearl postulates.

Lemma. Let the AGM postulates (except (AGM5)) and (SBC) be given, and let Φ be a condition entailing that $\ulcorner \mathcal{B} * (\alpha \wedge \beta) \urcorner$ is consistent. Then any iterated revision recipe of the form

$$(+) \qquad \mathcal{B} * \alpha * \beta \simeq \begin{cases} \mathcal{B} * (\alpha \wedge \beta) & \text{if } \Phi \\ \mathcal{B} * \beta & \text{otherwise} \end{cases}$$

satisfies the Darwiche-Pearl postulates.

A proof of this lemma can be found in the Appendix.

Notice that, given (AGM2), such a condition Φ implies that $\alpha \wedge \beta$ is consistent, and, by (AGM7), Φ is implied by (but does not in general imply) the condition that $\ulcorner \mathcal{B} * \alpha \urcorner \cup \{\beta\}$ is consistent.

To give an impression of the scope of the lemma, we give a first example of its application. Given (AGM7), the general format of (+) covers *restrained revision* as introduced by Booth and Meyer (2006, p. 142). This model of (one-dimensional) belief revision is characterized by the following condition:

$$\mathcal{B} * \alpha * \beta \simeq \begin{cases} \mathcal{B} * (\alpha \wedge \beta) & \text{if } \ulcorner \mathcal{B} * \alpha \urcorner \cup \{\beta\} \text{ is consistent} \\ & \quad \text{or } \ulcorner \mathcal{B} * \beta \urcorner \cup \{\alpha\} \text{ is consistent} \\ \mathcal{B} * \beta & \text{otherwise .} \end{cases}$$

3. Representing belief states as order relations

We will work with two different forms of representations of belief states that are sufficient to determine the set of beliefs held after any sequence of inputs.[13] The first is a *total pre-ordering of possible worlds.* Such pre-orderings can equivalently be presented in the form of Grovean *systems of spheres* (*s.o.s.*) of possible worlds (Grove 1988). This is the most graphic and easily comprehensible representation of a belief state. For the present paper, we assume that an s.o.s. \$ is a non-empty, finite set of finite sets of possible worlds such that for any sets S and S' in \$, either $S \subseteq S'$ or $S' \subseteq S$ (that is, the elements of \$ are 'nested', or form a chain with respect to set inclusion). Intuitively, the smallest (graphically, the 'innermost') sphere of \$ contains the most plausible worlds, the second smallest sphere contains the second most plausible worlds (besides the most plausible ones), and so on. Worlds not contained in any sphere are called *inaccessible* in

[13]A third representation in terms of prioritized belief bases is particularly attractive for the RbC operation; see Rott (2009).

\$. The set of sentences true at all the worlds contained in the innermost sphere encode the beliefs held true by an agent in belief state \$. The agent believes α if and only if α is true throughout the innermost sphere. The agent's set of beliefs, or more terminologically, his or her *belief set* is denoted by $\ulcorner \$ \urcorner$.

The s.o.s. presentation is generally to be preferred to an equivalent total pre-ordering of possible worlds, because it is much easier to visualize. However, the semantic conditions corresponding to the Darwiche-Pearl postulates, written as constraints on the change of s.o.s.s, are rather less intuitive than those for orderings (compare footnote 10). As I am not aware that the Darwiche-Pearl postulates have been represented as conditions for the change of systems of spheres elsewhere, let us give them in this form here. We need a few preparatory definitions. If \$ is a system of spheres and α a sentence then $\$ \cap [\alpha]$ is short for $\{S \cap [\alpha] : S \in \$ \text{ and } S \cap [\alpha] \neq \emptyset\}$. If X is a set of worlds, let $C_\$(X)$ denote the cover of X in \$, i.e., the minimal sphere S in \$ such that $X \subseteq S$. Here now are the semantic constraints corresponding to the (one-dimensional) Darwiche-Pearl postulates in s.o.s. language, with $\$^*_\alpha$ denoting the belief state \$ revised by α.[14]

(DPS1) $\$^*_\alpha \cap [\alpha] = \$ \cap [\alpha]$

(DPS2) $\$^*_\alpha \cap [\neg\alpha] = \$ \cap [\neg\alpha]$

(DPS3) For every S in \$, $C_{\$^*_\alpha}(S \cap [\alpha]) \subseteq S \cup [\alpha]$.

(DPS4) For every S' in $\$^*_\alpha$, $C_\$(S' \cap [\neg\alpha]) \subseteq S' \cup [\neg\alpha]$.

Our second way of representing a belief state is by a total pre-ordering $\leq$ of sentences, usually called *entrenchment relation* (Gärdenfors and Makinson 1988, Rott 2001; 2003a). Such an ordering can roughly be thought of as reflecting the degree of belief or the comparative retractability of the respective sentences. These degrees are required to respect logical structure in two ways. First, if α implies β, then the entrenchment of α cannot be higher as that of β (*dominance*). Second, the conjunction $\alpha \wedge \beta$ is not less entrenched than the weaker of α and β (*conjunctiveness*). In the first respect entrenchments behave like probabilities, in the second, they are quite different. The set of sentences that are more than minimally entrenched are the beliefs held true by an agent in belief state $\leq$. This is his or her *belief set* and denoted by $\ulcorner \leq \urcorner$.[15]

[14] A short proof of the equivalence of these DP sphere postulates with the original DP ordering postulates is given in the Appendix.

[15] We assume in this paper that the entrenchment relation is non-trivial in the sense that it does not relate all sentences of the language, or equivalently, in the sense that $\top$ is strictly more entrenched than $\bot$. The belief set $\ulcorner \leq \urcorner$ associated with a non-trivial entrenchment ordering $\leq$ is *always* consistent. Sometimes, if the agent has in fact inconsistent beliefs, one must think of her entrenchment ordering $\leq$ as supporting the set $\{\alpha : \bot \leq \alpha\}$ which is the set of all sentences, by the dominance condition for $\leq$.

What does it mean to say that a system of spheres or an entrenchment ordering represents a belief state? This is not a trivial question. As we said in the introduction, a formal structure like an s.o.s. $\$$ or an entrenchment relation $\leq$ is still an abstraction from a real belief state, and what the belief state really is may be inscrutable to us. But we can say that $\$$ or $\leq$ *represents a belief state* if it reproduces just those aspects of belief states we may hope to have access to, and this is the development of the agent's beliefs. In particular, the structures $\$$ and $\leq$ should encode all the information that is necessary to derive the resulting belief sets for all iterated belief changes, provided that a specific recipe of using $\$$ and $\leq$ to construct a single revision step is given. For the representation of a belief state $\mathcal{B}$ by a system of spheres $\$$, this means that in one-dimensional belief revision

$$\ulcorner(((\mathcal{B} * \alpha) * \beta) * \gamma) * \ldots\urcorner = \ulcorner(((\$^*_\alpha)^*_\beta)^*_\gamma)^* \ldots\urcorner$$

and in two-dimensional belief revision

$$\ulcorner(((\mathcal{B} *_\delta \alpha) *_\varepsilon \beta) *_\zeta \gamma) * \ldots\urcorner = \ulcorner(((\$^*_{\alpha,\delta})^*_{\beta,\varepsilon})^*_{\gamma,\zeta})^* \ldots\urcorner$$

for all finite sequences of inputs $\langle\alpha, \beta, \gamma, \ldots\rangle$ or $\langle\langle\alpha, \delta\rangle, \langle\beta, \varepsilon\rangle, \langle\gamma, \zeta\rangle, \ldots\rangle$, respectively. The definitions concerning the representation of a belief state $\mathcal{B}$ by an entrenchment relation $\leq$ are similar.

It is well-known from the belief revision literature beginning with AGM that an ordering representation of a belief *state* $\mathcal{B}$ determines a one-dimensional revision function specifying, for each potential input sentence α, the belief *set* that results from revising $\mathcal{B}$ by α. Conversely, given such a one-dimensional belief *set* revision function satisfying certain rationality postulates, one can (re-)construct an ordering that can be taken to represent the belief *state*.[16] In order to make this paper self-contained, we list some of the relevant bridge principles.

For the connection between systems of spheres and revised belief sets, we can make use of the following transitions (see Grove 1988):

(From $\$$ to $\ulcorner*\urcorner$)	β is in $\ulcorner\mathcal{B}*\alpha\urcorner$ if and only if there is a sphere in $\$$ containing some α-worlds and all α-worlds in this sphere are β-worlds, or there is no sphere in $\$$ containing any α-worlds.
(From $\ulcorner*\urcorner$ to $\$$)	A set S of possible worlds is a sphere in $\$$ if and only if there is a sentence α such that $S = \{w \in W :$ for some β, w satisfies all sentences in $\ulcorner\mathcal{B} * (\alpha \vee \beta)\urcorner\}$.

[16] See AGM (1985), Gärdenfors (1988) and Rott (2001). The additional information encoded in two-dimensional belief change operations is not needed for the (re-)construction of the belief state. Note also that the connections to be presented in the rest of this section appeal to revised belief *sets* only, not to the full structure of any revised belief *states*.

For the connection between entrenchments and revised belief sets, we can use the following transitions (cf. Gärdenfors and Makinson 1988, Lindström and Rabinowicz 1991, Rott 1991):

(From $\leq$ to $\ulcorner * \urcorner$) β is in $\ulcorner \mathcal{B} * \alpha \urcorner$ if and only if $\neg\alpha < \alpha \rightarrow \beta$ or $\top \leq \neg\alpha$.

(From $\ulcorner * \urcorner$ to $\leq$) $\alpha \leq \beta$ if and only if α is not in $\ulcorner \mathcal{B} * \neg(\alpha \wedge \beta) \urcorner$ or $\ulcorner \mathcal{B} * \neg(\alpha \wedge \beta) \urcorner$ is inconsistent.

There is a certain asymmetry in the intuitive plausibilities of these recipes. For systems of spheres, the construction of $*$ in terms of \$ is much more transparent then the (re-)construction of \$ from $*$. For entrenchment relations, the situation is just the reverse: The (re-)construction of $\leq$ from $*$ is much more convincing than the construction of $*$ in terms of $\leq$. But because the two directions fit together perfectly both for systems of spheres and for entrenchments, both pairs of recipes are almost universally accepted in the one-dimensional setting.

Additional support comes from the result that the s.o.s. modelling and the entrenchment modelling are equivalent in quite a strong sense. One can easily complete a triangle by defining direct links between entrenchment relations and s.o.s.s in such a way that the linked structures generate exactly the same revision function. The relevant transitions are as follows (see, for instance, Pagnucco and Rott 1999):

(From \$ to $\leq$) $\alpha \leq \beta$ if and only if for all spheres S in \$, if α is true throughout S, then β is true throughout S as well.

(From $\leq$ to \$) A non-empty set S of possible worlds is a sphere in \$ if and only if there is a sentence α such that $S = \{w \in W : w$ satisfies all sentences β such that $\alpha \leq \beta\}$.[17]

When there is no danger of confusion, we will sometimes allow ourselves to say that an s.o.s. or an entrenchment relation *is* a belief state rather than saying that it is an abstraction from, or a representation of, a belief state.

4. Bounded revision as an operation on systems of spheres

In the last section we have seen that given the revised belief *sets* $\ulcorner \mathcal{B} * \alpha \urcorner$ for all inputs α, one can (re-)construct an ordering representation of the belief *state* $\mathcal{B}$. Similarly, given the revised belief sets $\ulcorner (\mathcal{B} *_\delta \alpha) * \beta \urcorner$ for all inputs β, one can reconstruct a representation of the belief state $\mathcal{B} *_\delta \alpha$. That is to say that

[17] I neglect the problem of adding the empty set to systems of spheres. Cf. footnote 33.

equations for the two-fold revision of belief sets in effect specify transitions from formal representations of $\mathcal{B}$ to formal representations of $\mathcal{B} *_\delta \alpha$.

In this way bounded revision functions can be viewed as functions applying to formal representations of belief states – not just to belief sets which for many purposes contain too little information, and not to full belief states themselves may ultimately be inscrutable. So far everything has been very abstract. In this and the next section, we give concrete and direct constructions for the transitions of bounded revision, as applying to systems of spheres and entrenchment relations respectively. We begin with the representation of belief states in terms of systems of spheres.

Let \$ be an s.o.s. Let α intersect \$, i.e., let there be at least one sphere in \$ that has a non-empty intersection with $[\alpha]$, where $[\alpha]$ denotes the set of possible worlds in which α is true. Let $S_{\alpha,\delta}$ be the smallest sphere S in \$ such that $S \cap [\alpha] \not\subseteq [\delta]$; if there is no such sphere, take $S_{\alpha,\delta}$ to be the largest sphere in \$. Let $S_{\alpha;\delta}$ be the largest sphere S in \$ such that $S \cap [\alpha] \subseteq [\delta]$; if there is no such sphere, put $S_{\alpha;\delta} = \emptyset$. Except for the limiting cases, $S_{\alpha,\delta}$ and $S_{\alpha;\delta}$ are neighbouring spheres, the former is just a little larger than the latter.

Now let $\$^*_{\alpha,\delta}$ denote the system of spheres that results from revising the prior s.o.s. \$ by an input sentence α, bounded by reference sentence δ. Our official definition of bounded revision as an operation on systems of spheres applies to the case in which α intersects \$.

(BoundRevSS)

$$\boxed{\$^*_{\alpha,\delta} = \{S \cap [\alpha] : S \in \$, S \cap [\alpha] \neq \emptyset \text{ and } S \subseteq S_{\alpha,\delta}\} \cup \{S \cup (S_{\alpha,\delta} \cap [\alpha]) : S \in \$\}}$$

If α does not intersect \$, let us simply define $\$^*_{\alpha,\delta} = \$ \cup \{\emptyset\}$. It would be an almost equally good idea to have exactly the same definitions with $S_{\alpha;\delta}$ uniformly substituted for $S_{\alpha,\delta}$; denote the resulting s.o.s. by $\$^*_{\alpha;\delta}$. Intuitively, the best α-worlds are moved to the center, as long as δ holds along with α, and in one of the variants even a little longer. Thus an initial section of the best α-worlds get promoted and become the best worlds *simpliciter*, while the remaining α-worlds stay where they were. Figures 1 and 2 illustrate what happens to an s.o.s. when it gets revised according to the two variants. The numbers used in these figures are there just to indicate the relative plausibilities of (regions of) possible worlds. '1' designates the most plausible worlds, '2' the second most plausible worlds, and so on. '∞' designates the doxastically impossible or inaccessible worlds.

We shall not further pursue the alternative method using $\$^*_{\alpha;\delta}$ in the present pa-

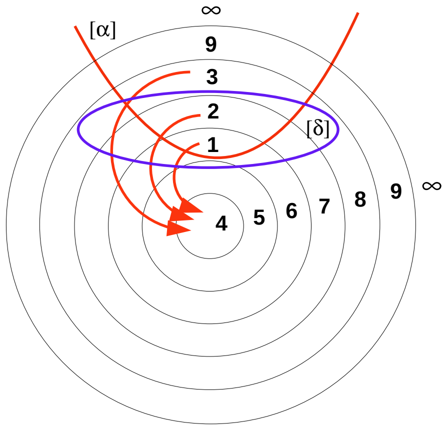

Fig. 1: Bounded revision using $S_{\alpha,\delta}$: Moving α-worlds to the center, as long as *and just a little longer than* they also satisfy δ

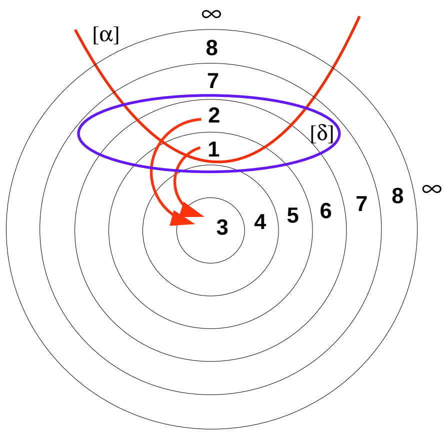

Fig. 2: Variant of bounded revision using $S_{\alpha;\delta}$: Moving α-worlds to the center, as long as they also satisfy δ

per.[18] This is for three reasons. First, this recipe violates the success condition (AGM2) if there is no S in \$ such that $S \cap [\alpha] \subseteq \delta$. Second, while we get that α covers *more* spheres in the posterior s.o.s. than δ if we use the recipe using $S_{\alpha,\delta}$, we get no such relation for the recipe using $S_{\alpha;\delta}$. Third, (BoundRevSS) is the recipe that will allow us to reconstruct both conservative and moderate revision as limiting cases.[19]

Another (less important) design decision concerns the question how to deal with

[18]But see the definition of $\leq^*_{\alpha;\delta}$ and footnote 28 below.

[19]These are my names (Rott 2003b). The operations I denote by these names are more well-known as *natural revision* (Boutilier 1993) and *lexicographic revision* (Nayak 1994 and others), respectively.

inaccessible worlds. I suggest that once a world is inaccessible, it should not be made accessible by an ordinary operation of belief revision. The condition that inaccessibility be preserved is violated, for example, by purely lexicographic revisions that give absolute priority to the most recent information, i.e., that give preference to previously inaccessible α-worlds. But this does not seem desirable. Suppose a world in which humans have seven heads is doxastically inaccessible. Then a revision by, say, the proposition that Yulia Tymoshenko is the winner of the 2010 presidential elections in Ukraine should not make a world with humans having seven heads and Tymoshenko being the winner of the 2010 elections more plausible than a world in which the opposite is true. This is why we shall stick to the preservation of inaccessibility.[20]

Figure 1 highlights the fact that bounded revision satisfies the Same Beliefs Condition (SBC). If the input α is doxastically possible, then the new belief set will invariably, i.e., for any arbitrary reference sentence, be determined by the set of most plausible α-worlds, just as the Grovean interpretation of AGM would have it. If the input is doxastically impossible, then the new belief set will always be the inconsistent set, again irrespective of the reference sentence.

Once the semantics for bounded revision in terms of total pre-orderings of possible worlds is understood, it can be read off from the semantic pictures that bounded revision satisfies the Darwiche-Pearl postulates (DP1)–(DP4). All distinctions within the α-worlds and all distinctions within the $\neg\alpha$-worlds are preserved, and no α-world suffers a loss in preferential status vis-à-vis any $\neg\alpha$-world.[21] Again this contrasts with the RbC operation of Fermé and Rott. Figure 3 shows what happens to an s.o.s. when it gets changed by RbC. RbC collapses plausibility distinctions among some $\neg\alpha$-worlds and thus violates (DP2).

Figure 3 also brings out the fact that RbC tends to decrease the number of spheres in an s.o.s. (in the example of Fig. 3 from 6 spheres to 4 spheres), thus making plausibility distinctions coarser. This contrasts with bounded revision that tends to increase the number of spheres (in the example of Fig. 1 from 6 spheres to 9 spheres) and thus in a way introduces finer plausibility distinctions. In this respect, the two methods complement each other.

Now let us have a look at the limiting cases concerning the reference sentence

[20] Technically, this is guaranteed by excluding inaccessible worlds from $S_{\alpha,\delta}$. – Here is a consequence of this design decision. Let β be doxastically possible, and $\alpha \wedge \beta$ be consistent but doxastically impossible (a proposition is *doxastically possible* if there is an accessible world at which it is true). Then $(\mathcal{B} *_\top \alpha) * \beta \simeq \mathcal{B} * \beta$, and $\ulcorner(\mathcal{B} *_\top \alpha) * \beta\urcorner$ contains $\neg\alpha$ and does not contain α. Cf. below the remarks on moderate revision for which the treatment of inaccessible worlds is most relevant.

[21] The interested reader may wish to verify directly that (BoundRevSS) satisfies the systems of spheres conditions (DPS1)–(DPS4) corresponding to Darwiche and Pearl's postulates.

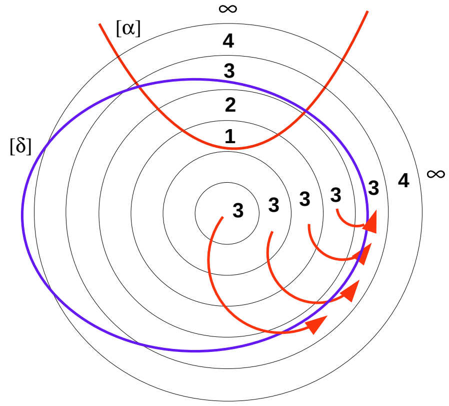

Fig. 3: Revision by comparison: Moving the best $\neg\alpha \wedge \delta$-worlds to the closest $\neg\delta$-permitting sphere, deleting all distinctions

for (BoundRevSS), and suppose that δ is (i) never or (ii) always cotenable with α. The first case surely obtains when δ is $\bot$, the second case when δ is $\top$.[22] If δ is $\bot$, then $S_{\alpha,\delta}$ is the smallest sphere S in \$ such that $S \cap [\alpha] \neq \emptyset$; let us denote this sphere by S_α. If there is no sphere S such that $S \cap [\alpha] \neq \emptyset$, we let S_α be the largest sphere S in \$ and denote it by S_{max}. What we get is *natural revision* (Boutilier 1993) or *conservative revision* (Rott 2003b):

$$\$^*_{\alpha,\bot} \;=\; \{S_\alpha \cap [\alpha]\} \;\cup\; \{S \cup (S_\alpha \cap [\alpha]) : S \in \$\}$$

If δ is $\top$, then $S_{\alpha,\delta}$ is the S_{max}. What we get in this case is *lexicographic revision* (Nayak 1994, Nayak, Pagnucco and Peppas 2003) or *moderate revision* (Rott 2003b):

$$\$^*_{\alpha,\top} \;=\; \{S \cap [\alpha] : S \in \$ \text{ and } S \cap [\alpha] \neq \emptyset\} \;\cup\; \{S \cup (S_{max} \cap [\alpha]) : S \in \$\}$$

Fig. 4 gives pictures of these (essentially one-dimensional) limiting cases of bounded revision.

5. Bounded revision as an operation on entrenchment relations

We now turn to the construction for bounded revision conceived as an iterable revision function that operates on representations of belief states by entrenchment relations. Let $\leq$ be an entrenchment relation, and let $\leq^*_{\alpha,\delta}$ be the entrenchment relation that results from revising $\leq$ by the input sentence α,

[22] The same effects are achieved if we let δ be $\neg\alpha$ and α, respectively. – I am neglecting the case of a doxastically impossible α for a while.

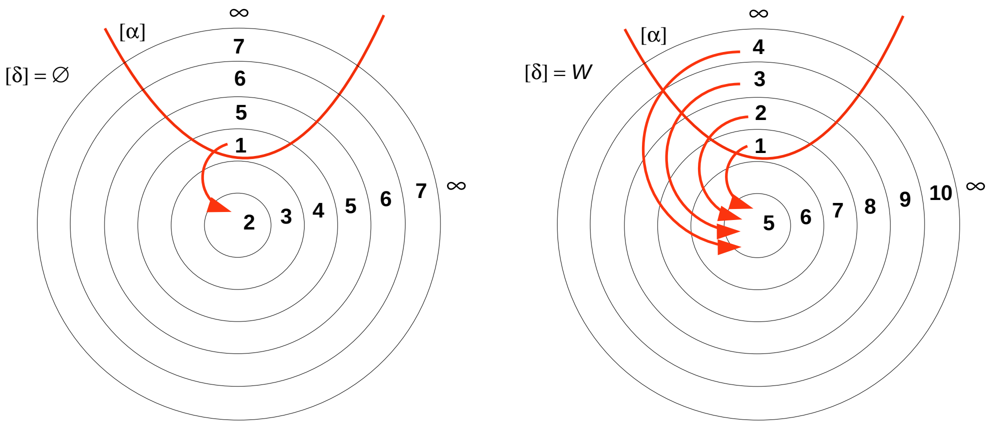

Fig. 4: Bounded revision using $S_{\alpha,\delta}$ with $\delta = \bot$ and $\delta = \top$: conservative and moderate revision

bounded by the reference sentence δ. As mentioned in the introduction, we allow arbitrary sentences to take the role of δ, keeping in mind that δ cannot in general be interpreted as a marker delineating a degree of belief relative to the prior belief state $\leq$. In fact δ need not even be a belief at all. Here is the revised entrenchment ordering, defined by comparing any two sentences β and γ:

(BoundRevEnt)

$$\boxed{\beta \leq^*_{\alpha,\delta} \gamma \text{ iff } \begin{cases} \alpha \to \beta \leq \alpha \to \gamma & \text{if } \alpha \to (\beta \wedge \gamma) \leq \alpha \to \delta \\ & \quad \text{and } \alpha \to (\beta \wedge \gamma) < \top \\ \beta \leq \gamma & \text{otherwise} \end{cases}}$$

It is not easy to understand this recipe. What does the condition used for the case distinction express, i.e., what is the meaning of $\alpha \to (\beta \wedge \gamma) \leq \alpha \to \delta$? The condition basically requires, translated into the model of systems of spheres, that at least one of $[\beta]$ and $[\gamma]$ has a non-empty intersection with $S_{\alpha,\delta} \cap [\alpha]$. (BoundRevEnt) says that in this case, it is precisely this intersection that decides, or these intersections that decide, the new entrenchment ordering between β and γ. If neither of them intersects $S_{\alpha,\delta} \cap [\alpha]$, the previous ordering is decisive. (This gloss neglects the limiting case $\top \leq \alpha \to (\beta \wedge \gamma)$.)

(BoundRevEnt) can even be applied when α is doxastically impossible, i.e., when $\neg\alpha$ is as firmly entrenched than the tautology $\top$. In this case the lower line of (BoundRevEnt) determines that the entrenchment relation should not change at all. Joined with the usual definition of the belief set $\ulcorner \leq \urcorner$ as $\{\gamma : \bot < \gamma\}$, we get from (BoundRevEnt), after some simplification, $\ulcorner \leq^*_{\alpha,\delta} \urcorner = \{\gamma : \neg\alpha < \alpha \to \gamma\}$

if $\neg\alpha < \top$, and $\ulcorner \leq^*_{\alpha,\delta} \urcorner = \{\gamma : \bot < \gamma\}$ if $\top \leq \neg\alpha$.[23] The latter case leads to a violation of the success condition (AGM2). For $\top \leq \neg\alpha$ entails that $\alpha \leq \bot$.[24] But this problem should not be blamed on (BoundRevEnt). If we insist that a revision by an impossible input α results in the inconsistent belief set, we need to exploit the stopgap mentioned in footnote 15 and redefine for such cases $\ulcorner \leq^*_{\alpha,\delta} \urcorner$ as $\{\gamma : \bot \leq^*_{\alpha,\delta} \gamma\}$.

It is not difficult to check that in the principal case when $\alpha \rightarrow \delta$ is less entrenched than $\top$, (BoundRevEnt) yields the posterior ordering $\delta <^*_{\alpha,\delta} \alpha$, that is, $\delta \leq^*_{\alpha,\delta} \alpha$ but not $\alpha \leq^*_{\alpha,\delta} \delta$. So α surpasses δ in terms of entrenchment after the revision has been made. But it does so only to the slightest possible degree. There is no sentence ϕ for which $\delta <^*_{\alpha,\delta} \phi <^*_{\alpha,\delta} \alpha$. One could rightly say that (BoundRevEnt) literally defines a kind of "revision by comparison", in that it implements a reasonable way of minimally accepting the condition $\delta < \alpha$.

Let us have a brief look at the following variant of (BoundRevEnt):

$$\beta \leq^*_{\alpha;\delta} \gamma \ \text{ iff } \ \begin{cases} \alpha \rightarrow \beta \leq \alpha \rightarrow \gamma & \text{if } \alpha \rightarrow (\beta \wedge \gamma) < \alpha \rightarrow \delta \\ \beta \leq \gamma & \text{otherwise} \end{cases}$$

This variant is obviously quite similar to our official definition. If fact, the definition of $\leq^*_{\alpha;\delta}$ corresponds to the operation depicted in Figure 2 in exactly the same way that (BoundRevEnt) corresponds to Figure 1, where the correspondence is based on the bridge ideas between systems of spheres and entrenchment orderings mentioned at the end of section 3.[25] This variant of (BoundRevEnt) fails to validate $\delta <^*_{\alpha;\delta} \alpha$.

We still need to put to record that the two definitions lead from entrenchment relations $\leq$ to entrenchment relations $\leq^*_{\alpha,\delta}$ and $\leq^*_{\alpha;\delta}$. We state without proof the following

Lemma. (BoundRevEnt) and its variant using $\leq^*_{\alpha;\delta}$ define entrenchment relations, i.e. total pre-orderings that satisfy dominance and conjunctiveness.

[23]Notice by the way that both of these sets are independent of δ, which makes it clear again the (SBC) holds for bounded revision.

[24]This partial violation of 'success' is related to the partial violation of the Triangle property in Rott (2003b, p. 120).

[25]There is a residual difference between representations by s.o.s.s and representations by entrenchments that is difficult to fix. Both (BoundRevEnt) and its variant violate the success condition (AGM2) if $\top \leq \neg\alpha$ and the belief set $\ulcorner \leq^*_{\alpha,\delta} \urcorner$ is defined as usual. What should happen after a revision by a doxastically impossible input is that the belief set comes out inconsistent. But, as mentioned before, the belief set associated with any non-trivial entrenchment relation is consistent. While we can change s.o.s.s by adding $\emptyset$ to $\$$ without changing the corresponding ordering of worlds, there is no analogous trick for entrenchment relations. Compare what was said about the stopgap redefinition of $\ulcorner \leq^*_{\alpha,\delta} \urcorner$ above.

We now turn to the limiting cases in which the reference sentence is a logical falsehood or a logical truth. First, let δ be $\bot$. Then (BoundRevEnt) reduces to:

$$\beta \leq^*_{\alpha,\bot} \gamma \text{ iff } \begin{cases} \alpha \rightarrow \beta \leq \alpha \rightarrow \gamma & \text{if } \alpha \rightarrow (\beta \wedge \gamma) \leq \neg\alpha \text{ and } \alpha \rightarrow (\beta \wedge \gamma) < \top \\ \beta \leq \gamma & \text{otherwise} \end{cases}$$

As already noted, this is conservative revision.[26] For the second limiting case, let δ be $\top$. In this case (BoundRevEnt) reduces to

$$\beta \leq^*_{\alpha,\top} \gamma \text{ iff } \begin{cases} \alpha \rightarrow \beta \leq \alpha \rightarrow \gamma & \text{if } \alpha \rightarrow (\beta \wedge \gamma) < \top \\ \beta \leq \gamma & \text{otherwise} \end{cases}$$

As noted before, this is moderate revision.[27]

6. Bounded revision as captured by a postulate for iterated belief change

Bounded revision is a change operation the non-iterated part of which is fully taken care of by (SBC) and the AGM postulates (without AGM5). So it remains to address the problem of repeated revisions. The general iteration condition for the two-dimensional operation of *bounded revision* is this:

(BoundRevIter)

$$\boxed{\mathcal{B} *_\delta \alpha * \beta \;\; \simeq \;\; \begin{cases} \mathcal{B} * (\alpha \wedge \beta) & \text{if } \ulcorner \mathcal{B} * (\alpha \wedge (\delta \rightarrow \beta)) \urcorner \text{ is consistent with } \beta \\ \mathcal{B} * \beta & \text{otherwise} \end{cases}}$$

Since we presuppose that bounded revision satisfies the Same Beliefs Condition (SBC), we may omit the final subscripts of a revision sequence as long as we are only interested in the identity of the resulting belief *sets*. Unfortunately, (BoundRevIter) still is not very transparent. The rationale for it is, of course, that it corresponds to the modellings in terms of systems of spheres and entrenchments.[28]

We are now going to consider the limiting cases. Let us this time also have an extra look what happens if the input sentence α is replaced by $\top$ or $\bot$. For the case $\alpha = \top$, condition (BoundRevIter) gives $\mathcal{B} *_\delta \top * \beta \;\simeq\; \mathcal{B} * \beta$

[26] Notice that $\alpha \rightarrow (\beta \wedge \gamma) \leq \neg\alpha$ means, as in the usual AGM paradigm, the same as $\beta \wedge \gamma \notin \ulcorner \mathcal{B} * \alpha \urcorner$. If this condition is satisfied, $\alpha \rightarrow \beta \leq \alpha \rightarrow \gamma$ can be simplified to $\alpha \rightarrow \beta \leq \neg\alpha$.

[27] Had we presumed that only logical truths get top entrenchment (an assumption corresponding to (AGM5)), then $\alpha \rightarrow (\beta \wedge \gamma) < \top$ would have meant the same as $\beta \wedge \gamma \notin Cn(\alpha)$.

[28] There is an alternative definition that uses the condition $\ulcorner \mathcal{B} *_\delta (\alpha \wedge (\delta \rightarrow \beta)) \urcorner \vdash \delta$ for the case distinction instead of the condition used in (BoundRevIter). This corresponds to the alternative options mentioned in Sections 4 and 5.

because $\mathcal{B} * (\top \wedge \beta) = \mathcal{B} * \beta$, by (AGM6). But similarly, if $\alpha = \bot$, condition (BoundRevIter) gives $\mathcal{B} {*_\delta} \bot {*} \beta \ \simeq \ \mathcal{B} {*} \beta$, in this case because the upper line where β is consistent with $\mathcal{B} *_\delta \bot$ never applies, by (AGM1) and (AGM2). Notice that revising by an inconsistency does not lead anyone applying bounded revision into "epistemic hell". There is nothing hellish about revising into inconsistency. On the contrary, such revisions are very easily repaired in a subsequent revision step.[29] A bounded revision by a contradiction is just as harmless as a bounded revision by a tautology. The reader may check for herself that (BoundRevSS) and (BoundRevEnt) leave the belief state representation unchanged if α is either $\top$ or $\bot$.

Let us now consider the limiting cases with respect to the reference sentence. It is not surprising any more that if we fix the parameter sentence as $\bot$ or as $\top$, then bounded revision boils down to the operations of conservative revision and moderate revision, respectively. For the first limiting case, let δ be $\bot$. Then (BoundRevIter) reduces to

$$\mathcal{B} *_\bot \alpha * \beta \ \simeq \ \begin{cases} \mathcal{B} * (\alpha \wedge \beta) & \text{if } \ulcorner \mathcal{B} * \alpha \urcorner \text{ is consistent with } \beta \\ \mathcal{B} * \beta & \text{otherwise} \end{cases}$$

which characterizes *conservative revision*. The upper line follows already from the AGM postulates (AGM3), (AGM4), (AGM7) and (AGM8) for one-step revisions. So the postulate for iterated bounded revision in this case comes down to saying that $\mathcal{B} *_\bot \alpha * \beta \ \simeq \ \mathcal{B} * \beta$ whenever β is inconsistent with $\ulcorner \mathcal{B} * \alpha \urcorner$.

For the second limiting case, let δ be $\top$. Then, given the AGM postulates minus (AGM5), (BoundRevIter) reduces to

$$\mathcal{B} *_\top \alpha * \beta \ \simeq \ \begin{cases} \mathcal{B} * (\alpha \wedge \beta) & \text{if } \ulcorner \mathcal{B} * (\alpha \wedge \beta) \urcorner \text{ is consistent} \\ \mathcal{B} * \beta & \text{otherwise} \end{cases}$$

which characterizes *moderate revision*. Had we also embraced the consistency postulate (AGM5), the upper line could be made conditional on the simpler requirement that $\alpha \wedge \beta$ be consistent. If $\ulcorner \mathcal{B} * (\alpha \wedge \beta) \urcorner$ is consistent, so is $\alpha \wedge \beta$. The reverse direction does not hold in our more general setting, because we allow for the existence of inaccessible $\alpha \wedge \beta$-worlds.

Using three different routes, we have confirmed the two well-known unary revision functions as limiting cases of bounded revision. Both conservative and moderate revision, however, seem to be defective. Conservative revision accepts

[29] The view that inconsistency is "epistemic hell" is discussed by Gärdenfors (1988), Olsson (2003) and Levi (2003).

the new piece of evidence (as it should, according to (AGM2)), but it accords it only the lowest possible entrenchment, so that new evidence gets immediately lost if any contradiction with the next piece of evidence arises. It thus violates a requirement of 'temporal coherence' (see Rott 2003b, p. 137). A particularly unwelcome consequence of this is nicely illustrated by Darwiche and Pearl's (1997, pp. 10, 21) notorious 'red bird example'. Moderate revision in a way suffers from the opposite defect by accepting the new information very firmly, and arguably too firmly. After moderate revision, *all* α-worlds (except for the inaccessible ones) are preferred to *all* $\neg\alpha$-worlds. While conservative revision is too conservative, moderate revision seems too radical. Bounded revision has the advantage of systematically covering the whole range between these two extremes.

The revision operation specified by the AGM axioms (excluding (AGM5)), (SBC) and the iteration axiom (BoundRevIter) satisfies the Darwiche-Pearl postulates. This can be proved by considering the s.o.s. semantics for bounded revision. But it can also directly be proved on the level of the postulates. In order to do that, we just have to verify that the definition (BoundRevIter) is of the form (+) of Section 2.2. Suppose that $\ulcorner\mathcal{B} * (\alpha \wedge (\delta \rightarrow \beta))\urcorner$ is consistent with β. We need to check that $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$ is consistent. But by (AGM6), $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$ is identical with $\ulcorner\mathcal{B} * ((\alpha \wedge (\delta \rightarrow \beta)) \wedge \beta)\urcorner$, and the latter set is a subset of $Cn(\ulcorner\mathcal{B} * (\alpha \wedge (\delta \rightarrow \beta))\urcorner \cup \{\beta\})$, by (AGM7). Since this latter set was supposed to be consistent, we are done.

Assuming that belief states $\mathcal{B}$ are represented by systems of spheres \$ or by entrenchment relations $\leq$, we can now supply the following two characterization theorems. Remember that we are working in a finitistic framework and understand by 'the AGM-axioms' only the postulates (AGM1)–(AGM4) and (AGM6)–(AGM8).

Theorem 1. (i) The two-dimensional revision function $*$ determined by (BoundRevSS) satisfies the AGM-axioms, (SBC) and (BoundRevIter).

(ii) If the two-dimensional revision function $*$ satisfies the AGM-axioms, (SBC) and (BoundRevIter), then there is, for each iterated revision process determined by $*$, a system of spheres \$ such that at each state in this process, the set of beliefs accepted is identical with the set of beliefs determined by the corresponding system of spheres as transformed according to (BoundRevSS):

$$\ulcorner\mathcal{B} *_\delta \alpha\urcorner = \ulcorner\$^*_{\alpha,\delta}\urcorner$$

$$\ulcorner(\mathcal{B} *_\delta \alpha) *_\varepsilon \beta\urcorner = \ulcorner(\$^*_{\alpha,\delta})^*_{\beta,\varepsilon}\urcorner$$

$$\ulcorner((\mathcal{B} *_\delta \alpha) *_\varepsilon \beta) *_\zeta \gamma\urcorner = \ulcorner((\$^*_{\alpha,\delta})^*_{\beta,\varepsilon})^*_{\gamma,\zeta}\urcorner$$

... and so on.

Theorem 2. (i) The two-dimensional revision function $*$ determined by (BoundRevEnt) satisfies the AGM-axioms, (SBC) and (BoundRevIter).

(ii) If the two-dimensional revision function $*$ satisfies the AGM-axioms, (SBC) and (BoundRevIter), then there is, for each iterated revision process determined by $*$, an entrenchment relation $\leq$ such that at each state in this process, the set of beliefs accepted is identical with the set of beliefs determined by the corresponding entrenchment relation as transformed according to (BoundRevEnt):

$$\ulcorner \mathcal{B} *_\delta \alpha \urcorner = \ulcorner \leq^*_{\alpha,\delta} \urcorner$$

$$\ulcorner (\mathcal{B} *_\delta \alpha) *_\varepsilon \beta \urcorner = \ulcorner (\leq^*_{\alpha,\delta})^*_{\beta,\varepsilon} \urcorner$$

$$\ulcorner ((\mathcal{B} *_\delta \alpha) *_\varepsilon \beta) *_\zeta \gamma \urcorner = \ulcorner ((\leq^*_{\alpha,\delta})^*_{\beta,\varepsilon})^*_{\gamma,\zeta} \urcorner$$

... and so on.

A sketch of the proof of the 'completeness part' (ii) of Theorem 2 can be found in the appendix.

7. Conclusion

We have discussed a two-dimensional operation of belief revision which lies 'between' quantitative and qualitative approaches in that it does not use numbers and is yet able to specify the *extent* or *degree* to which a new piece of information is to be accepted. It does so by specifying a reference sentence with the idea that the input has to be accepted *as far as*, and just a little *further than*, the reference sentence holds along with ('is cotenable with') the input sentence. As a result, the input sentence is accepted just a little more *strongly* than the reference sentence in the belief state reached after the revision has been performed. The acceptance of the input sentence may be said to be bounded by the reference sentence.

In these respects bounded revision is similar to the operations of *raising* and *lowering* of Cantwell (1997) and of *revision by comparison* (*RbC*) of Fermé and Rott (2004). But there are substantial differences. Since raising and lowering were not suggested as operations of belief *revision*, we summarize only the differences between bounded revision and RbC.[30]

[30]The idea of RbC was illustrated by Figure 3. In order to give the reader an impression of the complexity of the recipes involved, here are the formal definitions using s.o.s.s and entrenchments:

(1) While bounded revision follows basically an *as-long-as strategy* ('accept α as long as δ holds along with it (and just a little longer)'), RbC opts for an *at-least-as strategy* ('accept α as least as strongly as δ').

(2) In bounded revision, the input sentence is believed in the posterior belief state *just a little more strongly* than the reference sentence. It is believed *at least as strongly* as the reference sentence in RbC.

(3) While bounded revision tends to *refine* orderings of possible worlds and beliefs (the number of spheres in the agent's s.o.s. and the number of layers in her entrenchment relation increase), RbC has just the opposite effect and tends to *coarsen* orderings of possible worlds and beliefs (the number of spheres and entrenchment layers decreases).

(4) Bounded revision is invariably *successful* in the sense that the input sentence always gets accepted, independently of which reference sentence is used. RbC, in contrast, is successful only in a severely restricted form. If the reference sentence is not more entrenched than the negation of the input sentence,[31] then the former gets lost rather than the latter gets accepted.

(5) Bounded revision satisfies the *Same Beliefs Condition* (SBC) unconditionally, while RbC satisfies it only provided that either the revision is 'successful' (SBC^{r1}) or the strength of the index sentences δ and ε is the same (SBC^{r2}). The explanation for both (4) and (5) is, of course, that in contrast to bounded revision, RbC embodies not only an operation of belief revision, but sometimes also an operation of belief contraction.[32]

(6) Bounded revision satisfies the *Darwiche-Pearl postulates.* In contrast, RbC violates these postulates, since it wipes out relevant distinctions between some comparatively plausible worlds at which the input sentence is false and the reference sentence is true.

(7) Both models fill out a whole space of possibilities between interesting one-dimensional belief change functions as limiting cases. But these limiting cases are quite different. Taking a logical truth as the reference sentence gives *ir-*

(RbC-SS) $\quad \$^*_{\alpha,\delta} = \{S \cap [\alpha] : S \in \$, S \cap [\alpha] \neq \emptyset \text{ and } S \subseteq [\delta]\} \cup \{S : S \in \$ \text{ and } S \not\subseteq [\delta]\}$

(RbC-Ent) $\quad \beta \leq^*_{\alpha,\delta} \gamma \text{ iff } \begin{cases} \delta \wedge (\alpha \to \beta) \leq (\alpha \to \gamma) & \text{if } \beta \leq \delta \\ \beta \leq \gamma & \text{otherwise} \end{cases}$

To the best of my knowledge, a result analogous to Theorem 1 is valid, but has not been proved anywhere so far. Results analogous to Theorem 2 have been provided by Fermé and Rott (2004, Theorems 10–13).

[31] And, strictly speaking, if in addition neither $\top \leq \delta$ nor $\delta \leq \bot < \alpha$, where δ is the reference sentence and α the input sentence. See Fermé and Rott (2004), pp. 17–18 and condition (Q11).

[32] In more precise terminology: an operation of severe withdrawal (Pagnucco and Rott 1999). The conditions (SBC^{r1}) and (SBC^{r2}) were mentioned in footnote 9.

revocable revision (see Segerberg 1998 and Rott 2006) for RbC, while it gives *moderate revision* for bounded revision. Taking a logical falsity as the reference sentence generates *conservative revision* for bounded revision, but does not produce any change for RbC. Fixing a logical falsity as the input sentence produces a *severe withdrawal* of the reference sentence in RbC, but does not produce any sustained changes of the belief state in bounded revision.[33]

(8) Different presentations of the two operations give quite different views of their *complexity*. The iteration axiom (BoundRevIter) for bounded revision is simple, but the iteration axiom for RbC is terribly complicated.[34] On the other hand, this picture changes completely if we look at revisions of belief states in terms of changes of prioritized data bases in the style of Rott (2009). Here the recipe for bounded revision is fairly complex, while that for RbC is surprisingly simple. Space does not permit us here to explain how the shifting of priorities represented by a prioritized data base works. The meaning of the following strings are given in the paper just mentioned. We can only convey a superficial impression of the complexities involved. The revision operations are here taken to transform a prioritized data base $\overrightarrow{h} = h_1 \prec \ldots \prec h_n$ into revised data bases $(\overrightarrow{h})^*_{\alpha,\delta}$ of the following forms:

(BoundRevPDB) $\quad \overrightarrow{h} \prec\!. \ \alpha \prec\!. \ \overrightarrow{h_{\leq(\alpha\to\delta)} \vee \alpha} \prec\!. \ \overrightarrow{h_{>(\alpha\to\delta)}}.$

(RbC-PDB) $\quad \overrightarrow{h_{<\delta}} \prec\!. \ h_{=\delta} \wedge \alpha \prec\!. \ \overrightarrow{h_{>\delta}}.$

While bounded revision needs to employ rather complicated disjunctions and material conditionals, RbC turns out to be absolutely straightforward. The input sentence α is inserted at the highest level in the prioritized data base at which the reference sentence is derivable.[35]

There is a whole list of questions concerning further methods of two-dimensional belief change. Is it possible to use bounded revision with an at-least-as strategy?[36] Can we use it with a posterior of '$\delta \leq \alpha$' in the place of '$\delta < \alpha$'? Can we

[33]In terms of systems of spheres, revising by an inconsistency according to (BoundRevSS) only adds the empty sphere as the new innermost sphere and thus generates an inconsistent belief set (without involving any change in the corresponding ordering of possible worlds). In terms of entrenchments, revising by an inconsistency according to (BoundRevEnt) introduces no change in the ordering of sentences, but the belief set obtained cannot be determined as the sentences *more* entrenched than $\perp$, but as those *at least as entrenched* as $\perp$ – and those are all sentences of the language. Cf. footnotes 15 and 25 above.

[34]See condition (IT) in Fermé and Rott (2004, p. 24). As mentioned in Rott (2007b), it is not known yet whether this condition can be replaced by one or more simple iteration axioms.

[35]Eduardo Fermé and I were not aware of this representation when we worked on our joint paper on RbC. Had seen this, several parts of that paper would have been a lot less cumbersome.

[36]The answer to this question is 'yes.'

equip bounded revision with an integrated contraction mechanism like RbC? Or, what may amount to the same thing, can we retain the main idea of RbC and at the same time satisfy the Darwiche-Pearl postulates? Is it possible to combine RbC with an as-long-as strategy? Questions such as these indicate the richness contained in the idea of two-dimensional belief revision.

The eight factors listed above are certainly not independent of each other. But taken together they articulate the idea that bounded revision and RbC are complements in various interesting ways. Bounded revision and RbC are two implementations of a single very general idea, that of renouncing the use of numbers and working with reference sentences instead. Belief change can thus be interpreted as a sort of doxastic preference change with inputs of the form '$\delta \leq \alpha$' or '$\delta < \alpha$'. There are some good reasons why just these two operations were chosen as objects of study, but they are definitely not the only reasonable ways to go two-dimensional. In the very first paper (I think) that presented the idea of two-dimensional belief change, John Cantwell (1997) introduced a method of lowering that we have not discussed at all in the present paper. And two further possibilities of giving concrete shape to the idea of two-dimensional revision spring to one's mind.

- Keep the idea of bounded revision, but don't let the range of α-worlds be torn apart. Rather keep the "distances" between them constant by shifting all of them uniformly against the $\neg\alpha$-worlds. On closer inspection, this method is philosophically not well justified. As pointed out above, the numbers appearing in Figures 1–4 *should not* be interpreted as indicating distances, and they *should not* be used for arithmetical operations. But of course that *can* be so interpreted and used. The method is then similar to Spohn's (1988) distance-preserving shifting that is specified in terms of meaningful ordinals right from the start. But the amount of shifting is not specified by an ordinal, as in Spohn's case, but by a reference sentence, as in bounded revision. So this is a genuinely two-dimensional revision operator, and numbers are not needed.[37]

  Such an approach is related in spirit to the family of improvement operators of Konieczny and Pino Pérez (2008) and Konieczny, Medina Grespan and Pino Pérez (2010). Their approach is non-numerical and one-dimensional. A combination of ideas is possible by choosing a specific improvement operator from Konieczny et al. and letting a reference sentence specify, e.g., by its degree of entrenchment, how many times the improvement operator is to be applied.

[37] Hild and Spohn (2008) demonstrate in their deep paper how much, or how little, it requires for a person changing her beliefs repeatedly to be ascribed the implicit possession of ordinal numbers.

- Take the input as being given by a constraint $\alpha < \beta$, and think of it as generating a very simple s.o.s. or a very simple entrenchment relation. Then perform the lexicographic revision of the current belief state by this s.o.s. or entrenchment relation in the style of Nayak (1994).

Going two-dimensional gives leeway for new approaches that abstain from stipulating numbers as meaningfully measuring degrees of belief. I expect that many interesting discoveries can be made about a great diversity of two-dimensional belief change operations. I hope to have indicated that one *can* work without numbers and advance to more elaborate forms of reasoning than the ones reported by Gordon (2004).[38] This paper is meant as an invitation to cooperate and explore a rich diversity of possibilities of two-dimensional belief change. As research in belief revision progresses, an increasing number of potentially rational models for revising one's belief states emerges. What we will need in order to put these promising models to practice successfully is a general methodology telling us when to apply which operations of belief change.

Acknowledgments

I thank Giacomo Bonanno, Jim Delgrande, Igor Douven, Abhaya Nayak, Toni Rønnow-Rasmussen, Thomas Strothotte, the staff of Schloss Dagstuhl and the wonderful members of the Formal Epistemology Project at the KU Leuven who all helped providing me, in their different ways, with the opportunity and stimulation to write this paper. I am also grateful to three anonymous referees of this journal for valuable comments.

Appendix: Proofs

Proof of the lemma of Section 2.2

Since we suppose that (SBC) is satisfied, we may safely work in a one-dimensional context. We need to show that revision functions obeying (+) satisfy (DP1) – (DP4).

(DP1) Suppose that β implies α. Then $\alpha \wedge \beta$ is equivalent with β, so by (AGM6), both lines of (+) entail that $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner = \ulcorner\mathcal{B} * \beta\urcorner$, as desired.

(DP2) Suppose that β is inconsistent with α. Then, by condition Φ and (AGM2), the lower line of (+) applies, so $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner = \ulcorner\mathcal{B} * \beta\urcorner$, as desired.

[38] Still more ambitious cognitive activities without numerals and numbers are discussed by Field (1980) and Hellman (1989).

(DP3) Suppose for reductio that α is in $\ulcorner\mathcal{B} * \beta\urcorner$, but not in $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner$. So $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner \neq \ulcorner\mathcal{B} * \beta\urcorner$, so the upper line of (+) must apply. But by (AGM2) and (AGM1), α is in $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$, and we get a contradiction.

(DP4) Suppose for reductio that $\neg\alpha$ is not in $\ulcorner\mathcal{B} * \beta\urcorner$, but in $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner$. So $\ulcorner(\mathcal{B} * \alpha) * \beta\urcorner \neq \ulcorner\mathcal{B} * \beta\urcorner$, so the upper line of (+) must apply and $\neg\alpha$ is in $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$. But by (AGM2) and (AGM1), α is in $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$ as well, so $\ulcorner\mathcal{B} * (\alpha \wedge \beta)\urcorner$ is inconsistent, and according to condition Φ the upper line must not apply. Again we have got a contradiction. QED

Proof of the equivalence of the (DPS) postulates with the (DPO) formulations (Section 3)

For the proof, we use the bridge principle mentioned in footnote 33. Let $\preceq'$ be short for $\preceq^*_\alpha$

(DPS1) and (DPS2) are trivial.

For DPS3, we need to show that it is equivalent to

(DPO3) For any α-world w and $\neg\alpha$-world w', if $w <_{\mathcal{B}} w'$, then $w <_{\mathcal{B}*\alpha} w'$.

(DPS3) implies (DPO3). Let $w \in [\alpha]$, $w' \in [\neg\alpha]$ and $w \prec w'$. The latter means, by the bridge principle, that there is a sphere S in \$ such that w but not w' is in S. Take this sphere S. We know that $w \in S \cap [\alpha] \subseteq C_{\$'}(S \cap [\alpha])$. Now $w' \notin S \cup [\alpha]$, since $w' \notin S$. So by (DPS3), $w' \notin C_{\$'}(S \cap [\alpha])$. So $C_{\$'}(S \cap [\alpha])$ separates w and w' in $\$'$, i.e., $w \prec' w'$, as desired.

(DPO3) implies (DPS3). Let $w \in C_{\$'}(S \cap [\alpha])$. In our finite setting this is equivalent to saying that $w \in \bigcup\{C_{\$'}(\{w'\}) : w' \in S \cap [\alpha]\}$. This in turn means, by the bridge principle, that $w \preceq' w'$ for some $w' \in S \cap [\alpha]$. Suppose that $w \notin [\alpha]$, that is $w \in [\neg\alpha]$. For the claim of (DPS3), we need to show that $w \in S$. Suppose for reductio that $w \notin S$. Then $w' \prec w$, separated by S. Since $w' \in [\alpha]$ and $w \in [\neg\alpha]$, it follows by (DPO3) that $w' \prec' w$, and we have a contradiction.

(DPO4) is equivalent with: For any α-world w and $\neg\alpha$-world w', if $w' <' w$, then $w' < w$. So it is clear that this case is analogous to the case of (DP3), with \$ and $\$'$ and α and $\neg\alpha$ changing their roles. QED

Sketch of proof for the completeness part (ii) of Observation 2

We derive (BoundRevIter) from (BoundRevEnt) and the bridge principles of Section 3 that connect one-step revisions and entrenchment relations. This will show that if the initial entrenchment relation, obtained from one-step belief revision though (From $\ulcorner * \urcorner$ to $\leq$), develops in accordance with (BoundRevEnt), then it generates, through (From $\leq$ to $\ulcorner * \urcorner$), exactly the development of $\ulcorner \mathcal{B} \urcorner$ according to (BoundRevIter). This is what the completeness part of the observation claims. Notice that the bridge principles connecting two-dimensional one-step revisions and entrenchment relations do not depend on the reference sentences, due to the (SBC) condition that $\ulcorner \mathcal{B} *_\delta \alpha \urcorner = \ulcorner \mathcal{B} *_\varepsilon \alpha \urcorner$ for all δ and ε.

$$\phi \in \ulcorner \mathcal{B} *_\delta \alpha *_\varepsilon \beta \urcorner \qquad \text{iff (From } \leq \text{ to } \ulcorner * \urcorner\text{)}$$

$$\text{(a) } \neg\beta <^*_{\alpha,\delta} \beta \to \phi \quad \text{or} \quad \text{(b) } \top \leq^*_{\alpha,\delta} \neg\beta$$

We consider (a) first. Applying in the second step (BoundRevEnt), we get for (a)

$$\left\{ \begin{array}{ll} \alpha\to\neg\beta < \alpha\to(\beta\to\phi) & \text{if } \alpha\to(\neg\beta\wedge(\beta\to\phi)) \leq \alpha\to\delta \\ & \text{and } \alpha\to(\neg\beta\wedge(\beta\to\phi)) < \top \\ \neg\beta < \beta\to\phi & \text{otherwise} \end{array} \right\} \text{ iff (logic)}$$

$$\left\{ \begin{array}{ll} \alpha\to\neg\beta < \alpha\to(\beta\to\phi) & \text{if } \alpha\to\neg\beta \leq \alpha\to\delta \\ & \text{and } \alpha\to\neg\beta < \top \\ \neg\beta < \beta\to\phi & \text{otherwise} \end{array} \right\} \text{ iff (From } \ulcorner * \urcorner \text{ to } \leq\text{)}$$

$$\left\{ \begin{array}{l} \alpha\to(\beta\to\phi) \in \ulcorner \mathcal{B} *_\zeta \neg((\alpha\to\neg\beta)\wedge(\alpha\to(\beta\to\phi)))\urcorner \neq \mathcal{L} \\ \qquad\qquad \text{if } \alpha\to\neg\beta \notin \ulcorner \mathcal{B} *_\delta \neg((\alpha\to\neg\beta)\wedge(\alpha\to\delta))\urcorner \\ \qquad\qquad \text{or } \ulcorner \mathcal{B} *_\delta \neg((\alpha\to\neg\beta)\wedge(\alpha\to\delta))\urcorner = \mathcal{L}, \\ \qquad\qquad \text{and } \ulcorner \mathcal{B} *_\delta \neg((\alpha\to\neg\beta))\urcorner \neq \mathcal{L} \\ \beta\to\phi \in \ulcorner \mathcal{B} *_\zeta \neg(\neg\beta\wedge(\beta\to\phi))\urcorner \neq \mathcal{L} \quad \text{otherwise} \end{array} \right\} \text{ iff (logic)}$$

$$\left\{ \begin{array}{l} \alpha\to(\beta\to\phi) \in \ulcorner \mathcal{B} *_\zeta (\alpha\wedge\beta)\urcorner \\ \qquad\qquad \text{if } \alpha\to\neg\beta \notin \ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner \\ \qquad\qquad \text{or } \ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner = \mathcal{L}, \\ \qquad\qquad \text{and } \ulcorner \mathcal{B} *_\delta (\alpha\wedge\beta)\urcorner \neq \mathcal{L} \\ \beta\to\phi \in \ulcorner \mathcal{B} *_\zeta \beta\urcorner \neq \mathcal{L} \quad \text{otherwise} \end{array} \right\} \text{ iff (AGM1,AGM2)}$$

$$\left\{ \begin{array}{ll} \phi \in \ulcorner \mathcal{B} *_\zeta (\alpha\wedge\beta)\urcorner & \text{if } \neg\beta \notin \ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner \text{ or } \ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner = \mathcal{L}, \\ & \text{and } \ulcorner \mathcal{B} *_\delta (\alpha\wedge\beta)\urcorner \neq \mathcal{L} \\ \phi \in \ulcorner \mathcal{B} *_\zeta \beta\urcorner \neq \mathcal{L} & \text{otherwise} \end{array} \right.$$

Except for some limiting cases, we thus get a confirmation of (BoundRevIter). To deal with the limiting cases satisfactorily, we need to follow the recommendation mentioned in footnote 8 and stipulate that ($*\emptyset 1$) be valid. Then the condition that $\ulcorner \mathcal{B} *_\delta (\alpha\wedge\beta)\urcorner$ is consistent implies that $\ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner$ is consistent, too. On the other hand, using (AGM6)–(AGM8), we can see that $\neg\beta \notin \ulcorner \mathcal{B} *_\delta (\alpha\wedge(\delta\to\beta))\urcorner$ implies that $\ulcorner \mathcal{B} *_\delta (\alpha\wedge\beta)\urcorner$ is consistent. So the final condition for (a) as a whole reduces to:

$$\begin{cases} \phi \in \ulcorner \mathcal{B} *_\zeta (\alpha \wedge \beta) \urcorner & \text{if } \neg\beta \notin \ulcorner \mathcal{B} *_\delta (\alpha \wedge (\delta \to \beta)) \urcorner \\ \phi \in \ulcorner \mathcal{B} *_\zeta \beta \urcorner \neq \mathcal{L} & \text{otherwise} \end{cases}$$

Now we consider condition (b), $\top \leq^*_{\alpha,\delta} \neg\beta$. Applying (BoundRevEnt) to this case, we get

$$\left.\begin{cases} \alpha \to \top \leq \alpha \to \neg\beta & \text{if } \alpha \to (\top \wedge \neg\beta) \leq \alpha \to \delta \\ & \text{and } \alpha \to (\top \wedge \neg\beta) < \top \\ \top \leq \neg\beta & \text{otherwise} \end{cases}\right\} \text{ iff (logic)}$$

$$\left.\begin{cases} \top \leq \neg(\alpha \wedge \beta) & \text{if } \neg(\alpha \wedge \beta) \leq \alpha \to \delta \\ & \text{and } \neg(\alpha \wedge \beta) < \top \\ \top \leq \neg\beta & \text{if } \alpha \to \delta < \neg(\alpha \wedge \beta) \text{ or } \top \leq \neg(\alpha \wedge \beta) \end{cases}\right\}$$

But the upper line is inconsistent with its condition of application, so we remain with the lower line. Since $\top \leq \neg\beta$ implies $\top \leq \neg(\alpha \wedge \beta)$, this the lower line reduces to

$$\top \leq \neg\beta \quad \text{iff (From } \ulcorner * \urcorner \text{ to } \leq)$$

$$\top \notin \ulcorner \mathcal{B} *_\zeta \neg(\top \wedge \neg\beta) \urcorner \text{ or } \ulcorner \mathcal{B} *_\zeta \neg(\top \wedge \neg\beta) \urcorner = \mathcal{L} \quad \text{iff}$$

$$\ulcorner \mathcal{B} *_\zeta \beta \urcorner = \mathcal{L}$$

Putting together the two conditions for (a) and (b), we finally get that $\phi \in \ulcorner \mathcal{B} *_\delta \alpha *_\varepsilon \beta \urcorner$ if and only if

$$\begin{cases} \phi \in \ulcorner \mathcal{B} *_\zeta (\alpha \wedge \beta) \urcorner & \text{if } \neg\beta \notin \ulcorner \mathcal{B} *_\delta (\alpha \wedge (\delta \to \beta)) \urcorner \\ \phi \in \ulcorner \mathcal{B} *_\zeta \beta \urcorner & \text{otherwise} \end{cases}$$

which is exactly (BoundRevIter). QED

References

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

Booth, R. and T. Meyer (2006). Admissible and restrained revision. *Journal of Artificial Intelligence Research 26*, 127–151.

Boutilier, C. (1993). Revision sequences and nested conditionals. In R. Bajcsy (Ed.), *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI'93)*, pp. 519–525.

Cantwell, J. (1997). On the logic of small changes in hypertheories. *Theoria 63*, 54–89.

Darwiche, A. and J. Pearl (1997). On the logic of iterated belief revision. *Artificial Intelligence 89*, 1–29.

Fermé, E. and H. Rott (2004). Revision by comparison. *Artificial Intelligence 157*, 5–47.

Field, H. (1980). *Science without numbers: A defence of nominalism.* Princeton, N.J./Oxford: Princeton University Press/Blackwell.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Cambridge, Mass.: Bradford Books, MIT Press.

Gärdenfors, P. and D. Makinson (1988). Revisions of knowledge systems using epistemic entrenchment. In M. Vardi (Ed.), *TARK'88: Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pp. 83–95. Los Altos, CA: Morgan Kaufmann.

Goodman, N. (1955). *Fact, Fiction, and Forecast.* Cambridge, Mass.: Harvard University Press. (Fourth edition 1983).

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science 306 (issue 5695)*, 496–499.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic 17*, 157–170.

Hellman, G. (1989). *Mathematics without numbers: Towards a modal-structural interpretation.* Oxford: Clarendon Press.

Hild, M. and W. Spohn (2008). The measurement of ranks and the laws of iterated contraction. *Artificial Intelligence 172*, 1195–1218.

Konieczny, S., M. M. Grespan, and R. P. Pérez (2010). Taxonomy of improvement operators and the problem of minimal change. In F. Lin, U. Sattler, and M. Truszczynski (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010.* AAAI Press.

Konieczny, S. and R. P. Pérez (2008). Improvement operators. In G. Brewka and J. Lang (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pp. 177–187. AAAI Press.

Levi, I. (2003). Contracting from epistemic hell is routine. *Synthese 135*, 141–164.

Lewis, D. (1973). *Counterfactuals.* Oxford: Blackwell. Second edition 1986.

Lindström, S. and W. Rabinowicz (1991). Epistemic entrenchment with incomparabilities and relational belief revision. In A. Fuhrmann and M. Morreau (Eds.), *The Logic of Theory Change*, LNAI 465, pp. 93–126. Berlin: Springer.

Nayak, A. (1994). Iterated belief change based on epistemic entrenchment. *Erkenntnis 41*, 353–390.

Nayak, A., M. Pagnucco, and P. Peppas (2003). Dynamic belief revision operators. *Artificial Intelligence 146*, 193–228.

Olsson, E. J. (2003). Avoiding epistemic hell: Levi on pragmatism and consistency. *Synthese 135*, 119–140.

Pagnucco, M. and H. Rott (1999). Severe withdrawal – and recovery. *Journal of Philosophical Logic 28*, 501–547. (correct, complete reprint with original pagination in the February 2000 issue; see publisher's 'Erratum', *Journal of Philosophical Logic* 29, 2000, p. 121.).

Papini, O. (2001). Iterated revision operations stemming from the history of an agent's observations. In M.-A. Williams and H. Rott (Eds.), *Frontiers in Belief Revision*, pp. 279–301. Dordrecht: Kluwer.

Rott, H. (1991). A nonmonotonic conditional logic for belief revision I. In A. Fuhrmann and M. Morreau (Eds.), *The Logic of Theory Change*, Volume 465 of *LNAI*, pp. 135–181. Berlin: Springer-Verlag.

Rott, H. (2001). *Change, Choice and Inference: A Study in Belief Revision and Nonmonotonic Reasoning*. Oxford: Oxford University Press.

Rott, H. (2003a). Basic entrenchment. *Studia Logica 73*, 257–280.

Rott, H. (2003b). Coherence and conservatism in the dynamics of belief. Part II: Iterated belief change without dispositional coherence. *Journal of Logic and Computation 13*, 111–145.

Rott, H. (2006). Revision by comparison as a unifying framework: Severe withdrawal, irrevocable revision and irrefutable revision. *Theoretical Computer Science 355*, 228–242.

Rott, H. (2007a). Bounded revision: Two-dimensional belief change between conservatism and moderation. In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson (Eds.), *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Lunds Universitet. Available at http://www.fil.lu.se/hommageawlodek.

Rott, H. (2007b). Two-dimensional belief change – an advertisement. In G. Bonanno, J. P. Delgrande, J. Lang, and H. Rott (Eds.), *Formal Models of Belief Change in Rational Agents*. IBFI, Schloss Dagstuhl. Available at http://drops.dagstuhl.de/opus/volltexte/2007/1240.

Rott, H. (2009). Shifting priorities: Simple representations for twenty-seven iterated theory change operators. In D. Makinson, J. Malinowski, and H. Wansing (Eds.), *Towards Mathematical Philosophy*, Trends in Logic, pp. 269–296. Berlin: Springer Verlag.

Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic 39*, 287–306.

Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In W. Harper and B. Skyrms (Eds.), *Causation in Decision, Belief Change, and Statistics*, pp. 105–134. Dordrecht: Kluwer.