THE IRREDUCIBILITY OF ITERATED TO SINGLE REVISION

Richard Booth
BoothR2@cs.cardiff.ac.uk

Jake Chandler
jake.chandler@cantab.net

Abstract After a number of decades of research into the dynamics of rational belief, the belief revision theory community remains split on the appropriate handling of sequences of changes in view, the issue of so-called iterated revision. It has long been suggested that the matter is at least partly settled by facts pertaining to the results of various single revisions of one's initial state of belief. Recent work has pushed this thesis further, offering various strong principles that ultimately result in a wholesale reduction of iterated to one-shot revision. The present paper offers grounds to hold that these principles should be significantly weakened and that the reductionist thesis should ultimately be rejected. Furthermore, the considerations provided suggest a close connection between the logic of iterated belief change and the logic of evidential relevance.

Keywords AGM - Darwiche-Pearl - evidence - iterated belief revision

Under which conditions should a particular belief be held after a sequence of revisions of one's world view? Under which conditions should it not? In spite of close to three decades of research in the influential programme that is AGM belief revision theory, it turns out that consensus remains far from being reached on this very elementary issue, which has come to be commonly known as the 'problem of iterated revision'.

In recent incarnations of the AGM model, the beliefs of a rational agent are modeled by a so-called doxastic state Ψ. Typically treated as a primitive, Ψ determines a 'belief set' $[\Psi]$, a deductively closed set of sentences, standardly drawn from a propositional, truth-functional language, that are believed by the agent to be true.[1]

The operation of belief revision, consisting in the adjustment of Ψ to accommodate the acquisition of a new belief is represented by a revision function $*$. This function takes as an input a pair consisting of a prior doxastic state Ψ and sentence A and outputs the posterior doxastic state $\Psi * A$ obtained by revising Ψ by A. Two popular sets of principles of diachronic rationality, the AGM postulates (Alchourrón, Gärdenfors and Makinson 1985) and the somewhat more recent Darwiche-Pearl (DP) postulates (Darwiche and Pearl 1997), provide constraints on the behaviour of the revision function.

[1] $[\Psi]$ is also alternatively denoted in the literature by 'Bel(Ψ)' (in Darwiche & Pearl 1997) or again '$\ulcorner\Psi\urcorner$' (in Rott 2011).

The AGM postulates constrain (a) the composition of the belief set $[\Psi * A]$ resulting from a *single* revision of a doxastic state Ψ by a sentence A, on the basis of (b) the composition of $[\Psi]$. A well-known representation theorem states that the single-shot revision dispositions of an AGM-compliant agent can be represented by a complete weak preference ordering of possible worlds, with the set of maximal A-worlds corresponding to those worlds in which all and only the sentences in $[\Psi * A]$ are true.[2] One noteworthy fact, which we shall make use of later in the paper, is that single-shot revision dispositions with respect to any truth-functional combination C of sentences A and B are fully determined by the restriction of this preference relation to the members of the sets of maximal $A \wedge B$-, $A \wedge \neg B$-, $\neg A \wedge B$- and $\neg A \wedge \neg B$-worlds.[3]

It should be noted that the restrictions imposed fall short of having (b) fully determine (a). In other words, the AGM postulates are consistent with a violation of the following reductionist thesis for single revisions:

(RED1) If $[\Psi] = [\Psi']$, then $[\Psi * A] = [\Psi' * A]$

This principle would entail that the composition of the set of maximal worlds determines the entire preference ordering. There is a broad consensus, however, that RED1 should be rejected, with Hansson (1992, 531-532) offering a convincing counterexample to the claim.

One consequence of this permissiveness of the AGM postulates is that they wind up placing relatively little by way of constraints on the belief set resulting from a *sequence* of two or more successive revisions of an initial doxastic state. And indeed, it has been noted that the AGM postulates remain perfectly consistent with a number of intuitively problematic sequences of change in view (Darwiche and Pearl *op.cit.*, 5-6).

This observation provides the rationale for the introduction of the DP postulates, which, in the presence of the AGM postulates, turn out to impose certain constraints on the relation between (c) the composition of the belief set $[(\Psi * A) * B]$ resulting from a twofold revision of Ψ by A and then B and (d) the composition of the belief sets $[\Psi * C]$ resulting from single revisions of Ψ by the various truth-functional combinations C of A and B. More specifically, one can show that the precise import of these principles, in the presence of the AGM postulates, amounts to the following:

$$\text{(DP}'\text{)} \quad [(\Psi * A) * B] = \begin{cases} [\Psi * A \wedge B], & \text{if (i) } A \in [(\Psi * A) * B] \\ [\Psi * B] \cap [\Psi * A \wedge B], & \text{if (ii) } A, \neg A \notin [(\Psi * A) * B]^{4} \\ [\Psi * B], & \text{if (iii) } \neg A \in [(\Psi * A) * B] \end{cases}$$

This reformulation of the DP postulates tells us that (c) is determined by (d), *if and only if* the latter also determines (e) whether $[(\Psi * A) * B]$ includes A, $\neg A$ or neither A nor

[2]For a presentation of the theorem, see Katsuno and Mendelzon (1991), as well as the related result in Grove (1988).

[3]See Appendix, Observation **??**. The proof is trivial, but we include it since we are not aware of its having been provided elsewhere in the literature.

[4]See Appendix, Observation **??**. The appendix also includes a statement of the DP and relevant AGM postulates.

$\neg A$. But while the AGM and DP postulates do turn out to also offer some constraints on the relation between (d) and (e),[5] these constraints fall short of having the AGM and DP postulates jointly entail the following reductionist thesis for two-fold revisions:

(RED2) If $[\Psi * C] = [\Psi' * C]$, for any truth-functional combination C of A and B, then $[(\Psi * A) * B] = [(\Psi' * A) * B]$

This principle is clearly weaker than RED1. Indeed, consider an arbitrary A and B and assume the antecedent of RED2. It follows that $[\Psi * A \wedge (B \vee \neg B)] = [\Psi' * A \wedge (B \vee \neg B)]$. On the standard assumption that revision by logically equivalent sentences yields the same posterior beliefs, this amounts to $[\Psi * A] = [\Psi' * A]$. By RED1, we then have the result that $[(\Psi * A) * B] = [(\Psi' * A) * B]$, as required. QED.

Having said that, RED2 remains an extremely substantial thesis. It has as an obvious consequence the following principle:

(RED3) If $[\Psi * C] = [\Psi' * C]$, for any C, then $[(\Psi * A) * B] = [(\Psi' * A) * B]$

which amounts, on the assumption that doxastic states are individuated by their behaviour under iterated revision, to an identification of doxastic states with preference orderings over worlds and a reduction of the problem of iterated revision to that of finding an appropriate function from preference orderings to preference orderings.[6]

Unlike RED1, however, RED2 (and hence RED3), appears to be widely endorsed in the belief revision community, albeit implicitly.[7] Indeed, a variety of competing supplementations of the AGM and DP postulates have been floated in the literature that turn out to entail this condition, ensuring that (d) determines (e) and hence (c). Such supplementations have for instance been offered by proponents of the 'lexicographic' (Nayak 1994, Nayak *et al* 2003), 'restrained' (Booth, Chopra and Meyer 2005, Booth and Meyer 2006) and 'natural' (Boutilier 1996) approaches to revision. Regarding natural revision, we have the following requirement:

(NR) (a) $A \in [(\Psi * A) * B]$ if $A \in [\Psi * B]$ or $\neg B \notin [\Psi * A]$

(b) $A, \neg A \notin [(\Psi * A) * B]$ if $A, \neg A \notin [\Psi * B]$ and $\neg B \in [\Psi * A]$

[5] These constraints take the form of two principles. The first offers a condition that is sufficient to place us in case (i) of DP′, telling us that $A \in [(\Psi * A) * B]$ if either $A \in [\Psi * B]$ or $\neg B \notin [\Psi * A]$. The second offers a condition that is sufficient to place us in case (iii) of DP′, telling us that $\neg A \in [(\Psi * A) * B]$ if both $\neg A \in \mathrm{Cn}(B)$ and $\neg B \notin \mathrm{Cn}(\varnothing)$. See Appendix, Observations **??** and **??**. These are the strongest constraints on the relation between (d) and (e) that we know to be derivable from the AGM and DP postulates.

[6] To put things a little more precisely, let us say that (i) Ψ and Ψ' are *k-equivalent* iff for any k-tuple $\langle A_1, \ldots, A_k \rangle$, $[(((\Psi * A_1) * \ldots) * A_k] = [(((\Psi' * A_1) * \ldots) * A_k]$ and that (ii) they are *equivalent* simpliciter iff there are k-equivalent for all k. RED3 is then the claim that, if two doxastic states are 1-equivalent (in other words: if they are such that their single-shot revision dispositions are representable by the same preference ordering), then they are 2-equivalent. What we are effectively noting is that this claim amounts to the following: if two doxastic states are 1-equivalent, then they are equivalent. We provide a quick proof of this in the appendix–see Observation **??**.

[7] To the best of our knowledge, the present paper offers the first explicit formulation and critical discussion of these claims.

(c) $\neg A \in [(\Psi * A) * B]$ if $\neg A \in [\Psi * B]$ and $\neg B \in [\Psi * A]$

Regarding restrained revision, the proposal is:

(RR) (a) $A \in [(\Psi * A) * B]$ if either $\neg A \notin [\Psi * B]$ or $\neg B \notin [\Psi * A]$

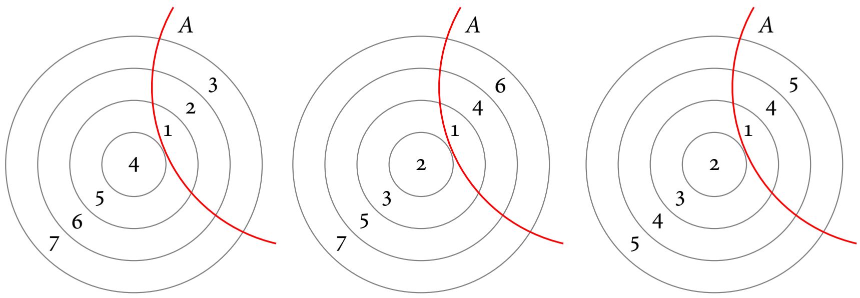
(b) $\neg A \in [(\Psi * A) * B]$ otherwise

Finally, lexicographic revision is characterised by:

(LR) (a) $A \in [(\Psi * A) * B]$ if $\neg B \notin \mathrm{Cn}(A)$

(b) $\neg A \in [(\Psi * A) * B]$ otherwise [8]

These suggestions all have well known and elegant representations in terms of preferences over possible worlds. In the case of the lexicographic approach, the posterior ordering is obtained by replacing the relation between any A-world and $\neg A$-world by a relation of strict preference for the former over the latter, whilst preserving the prior relations between all other kinds of pairs of worlds. In the case of the restrained approach, the transformation proceeds in two steps. First, any prior indifference relation between an A-world and an $\neg A$-world is replaced by a relation of strict preference for the former over the latter. Secondly, any prior relation between a world in the prior set of most preferred A-worlds and a world outside of that set is transformed into a relation of strict preference for the former over the latter. All other prior relations are preserved. Finally, the natural approach simply makes use of the second step of the restrained approach, while preserving all remaining prior relations. Figure **??**, reproduced from Rott (2009), offers a clear graphical representation of these mappings.



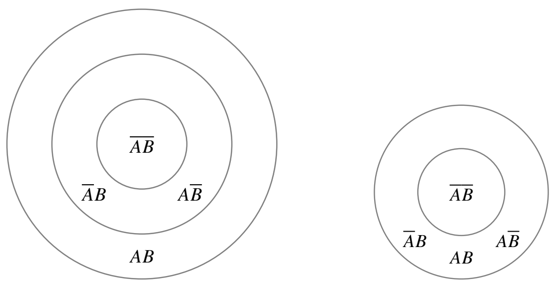
(a) Lexicographic revision **(b)** Restrained revision **(c)** Natural revision

Figure 1: Preferential representations of three reductionist proposals for iterated revision. The parallel bands represent equivalence classes of equipreferred worlds, with the order of preference increasing towards the center. The numerals indicate the ordinal position of the relevant set of worlds after revision by A.

However, technically convenient as it may be, RED2 turns out to be rather implausible on closer inspection.

[8]These are all clear strengthenings of the principles mentioned in footnote 5 above. It is easy to verify that they are equivalent, in the presence of the DP and AGM postulates, to the corresponding characteristic principles listed in Rott (2009, pp. 278–280). We provide a straightforward proof of this equivalence in the appendix–see Observation **??**.

Consider first the case in which the following conditions obtain: (a) $\neg A$ is in $[\Psi * B]$, (b) $\neg A$ is in $[\Psi]$, (c) $\neg B$ is in $[\Psi * A]$, (d) $\neg B$ is in $[\Psi]$ and (e) neither A nor B are in $[\Psi * A \vee B]$. It is easily verified that these conditions are jointly sufficient to fully determine the single shot revision dispositions of an agent with respect to all truth functional combinations of A and B, i.e. fully determine the composition of $[\Psi * C]$, for any truth functional combination C of A and B. The proof is trivial: As we noted at the very start of the paper, such dispositions are fully determined by the restriction of the preference ordering associated with Ψ to the members of the sets of maximal $A \wedge B$-, $A \wedge \neg B$-, $\neg A \wedge B$- and $\neg A \wedge \neg B$-worlds. Furthermore, conditions (a) to (e) are sufficient to determine the unique corresponding ordering depicted in Figure **??**. Indeed, from (b) and (d), it follows that the members of the set of maximal $\neg A \wedge \neg B$-worlds are strictly preferred to the members of the set of maximal $\neg A \wedge B$-worlds, the set of maximal $A \wedge \neg B$ and the set of maximal $A \wedge B$-worlds. (a) and (c) ensure that the members of the set of maximal $\neg A \wedge B$-worlds and the set of maximal $A \wedge \neg B$-worlds are strictly preferred to the members of the set of maximal $A \wedge B$-worlds. Now assume for reductio that either the members of the set of maximal $\neg A \wedge B$-worlds are strictly preferred to the members of the set of maximal $A \wedge \neg B$-worlds or that the converse holds. From the former, it would follow that B is in $[\Psi * A \vee B]$, while from the latter, it would follow that A is in that set, in both cases contrary to (e). QED.



(a) Examples 1 and 2　　**(b)** Examples 3 and 4

Figure 2: Preferential representation of the single shot revision dispositions with respect to all truth-functional combinations of A and B in Examples 1 to 4. 'AB', '$\overline{A}B$', '$A\overline{B}$' and '$\overline{A}\overline{B}$' represent the worlds in the sets of maximal $A \wedge B$-, $\neg A \wedge B$-, $A \wedge \neg B$- and $\neg A \wedge \neg B$- worlds, respectively. As in the previous figure, the most preferred worlds are situated in the innermost band.

The following pair of examples, however, suggests that such a profile of dispositions is insufficient to determine whether a sequence of revisions of Ψ by A and then by B would leave one believing that A or believing that $\neg A$:

> *Example 1:* I was told a couple of days ago that there is a party being thrown this weekend and that my friends Sam and Pam are invited. Initially unsure as to whether either would be tempted to attend, I have now heard that the venue is located far out of town, probably too far to really be worth the trip for either of them. I also believe that Sam and Pam do not get on and are unlikely to attend the same party.

Example 2: As in Example 1, save that I believe that Sam and Pam have never met and know nothing about each other.

Let 'A' stand for the proposition that Pam will attend the party and 'B' stand for the proposition that Sam will do so. The following is true in both Example 1 and Example 2: First of all, I initially believe that neither Sam nor Pam will attend ($\neg A$ and $\neg B$ are both in $[\Psi]$). Secondly, were I to come to believe of either Pam or Sam that they will attend, my initial grounds for thinking that the other will not would still hold: the venue would be no closer to town and the person attending would presumably need some special reason to want to go the extra mile ($\neg A$ is in $[\Psi * B]$ and $\neg B$ is in $[\Psi * A]$). Finally, were I to come to believe that at least one of the two will attend, I would not feel in a position to determine which (neither A nor B are in $[\Psi * A \vee B]$). So in both examples, my beliefs satisfy conditions (a) to (e). Additionally, in Example 2, we have A in $[(\Psi * A) * B]$. Indeed, in this situation, my coming to believe that Sam is attending has no bearing on my previously acquired belief that Pam will be there. In Example 1, however, we may well find ourselves with $\neg A$ in $[(\Psi * A) * B]$. Indeed, it is at the very least rationally permissible for my belief regarding Sam and Pam's mutual dislike to be one that I would be extremely reticent to give up, in particular for it to be one that would survive successive revisions by A and then by B. In such case, my ultimately coming to believe that Sam is attending would make me change my mind about Pam: I would judge that Pam will have made other plans after all. This, however would be inconsistent with RED2.

For a second type of counterexample with a somewhat similar flavour, consider now the case in which conditions (b), (d) and (e) above hold, i.e. $\neg A$ and $\neg B$ are both in $[\Psi]$ and neither A nor B are in $[\Psi * A \vee B]$, but (a) and (c) do not, so that $\neg A$ is not in $[\Psi * B]$, nor is $\neg B$ in $[\Psi * A]$. Again, this suffices to determine a profile of single-shot revision dispositions, depicted in Figure **??**. Indeed, as we saw above, (b) and (d) entail that the members of the set of maximal $\neg A \wedge \neg B$-worlds are strictly preferred to the members of the set of maximal $\neg A \wedge B$-worlds, the set of maximal $A \wedge \neg B$ and the set of maximal $A \wedge B$-worlds. We also saw that (e) entails that the members of the set of maximal $\neg A \wedge B$-worlds and the members of the set of maximal $A \wedge \neg B$-worlds are equipreferred. So assume for reductio that the members of the set of maximal $A \wedge B$-worlds are either strictly preferred to the members of the set of maximal $\neg A \wedge B$-worlds and the set of maximal $A \wedge \neg B$-worlds or strictly dispreferred to them. In the first case, we would have both A and B in $[\Psi * A \vee B]$, contrary to (e). In the second case, we would have $\neg A$ in $[\Psi * B]$ and $\neg B$ in $[\Psi * A]$, contrary to our assumption that (a) and (c) are false. QED.

But here too, a pair of cases shows that the relevant profile of single shot dispositions is insufficient to determine the outcome of a sequence of revisions of Ψ by A and then by B. This time, however, we have one case having such a sequence lead to a belief that A and the other having it lead to a suspension of judgment as to whether or not A:

Example 3: I was told a couple of days ago that a party was being thrown this weekend and that my friends Sam and Pam were invited. Initially unsure as to whether either of them would be tempted to attend, I have now heard that the party has been cancelled. Furthermore, I believe that Pam's relationship with Sam is somewhat unpredictable and that they avoid each other at all costs when they are going through a rough patch. I currently have no idea of the state of their relationship.

Example 4: As in Example 3, save that I believe that Sam and Pam have never met and know nothing about each other.

The following holds true in both Example 3 and Example 4: Since I think that the party has been cancelled, I initially believe, as I did in Examples 1 and 2, that neither Sam nor Pam will attend ($\neg A$ and $\neg B$ are both in $[\Psi]$). In contrast to what was the case in Examples 1 and 2, however, my coming to believe either of Pam or of Sam that they will attend *would* affect my prior grounds for thinking that the other will not. Indeed, such a change in view would require the belief that the party is not cancelled after all. Furthermore, my initial understanding that the party was cancelled constituted my sole grounds to presume of either invitee that they would not be attending. Hence, in coming to believe of one invitee that he or she will attend, I must leave open the possibility that the other will do so too ($\neg A$ is not in $[\Psi * B]$, nor is $\neg B$ in $[\Psi * A]$). Finally, as in Examples 1 and 2, were I to come to believe that at least one of the two will attend, I would not feel in a position to determine whether solely the one would show up, solely the other, or both together (neither A nor B are in $[\Psi * A \vee B]$). So my doxastic states in Examples 3 and 4 both satisfy conditions (b), (d) and (e) and violate (a) and (c). In Example 4, it is also the case, as with Example 2, that A is in $[(\Psi * A) * B]$. The situation is different, however, in Example 3, where it is perfectly possible to wind up with neither A nor $\neg A$ in $[(\Psi * A) * B]$. Whether or not this is the case will hinge on the extent to which my belief regarding the volatile nature of their relationship is deeply entrenched. It is certainly at least rationally permissible that it should survive the successive revisions by A and then by B. But if it does, then my coming to believe that Sam is attending will lead me to suspend judgment as to whether or not Pam is coming along too. Again, RED2 fails.

The root cause of the potential divergence of behaviour under iterated revision between the cases in each pair appears to be a difference in the evidential relations that are perceived to hold between A and B. In Examples 2 and 4, A and B are initially taken to be entirely evidentially independent and, in particular, B is not taken to undermine A, i.e. to provide a reason to not believe it. In contrast, in Example 1, B is taken to have a bearing on the question of whether or not A, indeed, it is taken to provide a sufficient reason to believe $\neg A$. In Example 3, B is also deemed evidentially pertinent to A, this being taken to provide a sufficient reason to suspend judgment as to whether or not A. What these pairs of cases demonstrate is that the dispositions that one has to revise one's beliefs upon single revisions by different truth-functional combinations of A and B are insufficient to determine one's attitude to such evidential matters.

While our examples constitute a direct challenge to RED2, the above evidential diagnosis of the underlying issue suggests that the weaker RED3 should also probably be rejected. Indeed, consider a case in which an agent has an extremely restricted conceptual repertoire, limited to combinations of two atomic sentences, A and B, and his or her single shot revision dispositions are captured by one or the other of the preference orderings depicted in Figure **??**, ordering **??** for example. It seems perfectly open to him or her to take B to provide a reason to believe $\neg A$, and hence believe $\neg A$ upon revising his or her initial beliefs by A and then by B. But it seems equally permissible for him or her to take A and B to be evidentially independent, thus believing A at the end of this sequence of revisions. (And indeed, if one or the other of these attitudes were not rationally acceptable, which one would it be and why?) But if such leeway is granted, then RED3 cannot hold: the doxastic state of an agent must be represented by a structure that is richer than a mere preference ordering over possible worlds.[9,10,11]

To sum things up succinctly, then, a set of counterexamples shows that progress on the key issue of iterated revision has been hampered over the past couple of decades by a tacit commitment to an implausibly strong reductionist claim, in the form of RED2. These counterexamples arguably also motivate the rejection somewhat weaker principle in the form of RED3, suggesting that a preference ordering over the set of possible worlds provides insufficient structure to represent an agent's commitments to policies of iterated revision. Furthermore, what appears to be a credible diagnosis of the situation suggests that future research on the logic of belief dynamics may be well advised to

[9]Could one not, in response to this, insist that a restricted version of RED3 nevertheless holds for agents with more extensive conceptual resources? In principle, sure. But the resulting picture would strike us as being unappealingly disunified, with doxastic states being representable by preference orderings in some cases but only by richer structures in others.

[10]Robert Stalnaker has also recently voiced suspicions regarding RED3. However, the grounds that he offers for doubting the principle are insufficiently strong. Finding fault with the first two Darwiche-Pearl postulates, he ipso facto rejects any reductionist proposal that satisfies them, including the three proposals that we consider here. But RED3 does not logically require either of of the postulates that he criticises and the aforementioned proposals do not exhaust the space of reductiivist options. See Stalnaker (2009).

[11]We should perhaps mention in passing another potential line of argument from the failure of RED2 to the failure of RED3: Recall that RED2 asserts that the belief sets resulting from revising a doxastic state by the different truth-functional combinations of A and B jointly determine the belief sets resulting from sequentially revising that state by A and then by B. But this strong 'determination' thesis entails an altogether far weaker 'consistency', or again 'irrelevance', principle, namely:

(IR) If $[\Psi * C] = [\Psi' * C]$, for any truth functional combination C of A and B, then there exists a Ψ'', such that $[\Psi' * C] = [\Psi'' * C]$, for any C, and $[(\Psi * A) * B] = [(\Psi'' * A) * B]$

The entailment is obvious: let $\Psi'' = \Psi'$. IR, in effect, tells us that, holding fixed one's single-shot revision dispositions with respect to sentences that *are* truth functional combinations of A and B, one's single-shot revision dispositions with respect to sentences that *are not* truth functional combinations of A and B are irrelevant to the composition of the belief set resulting from a sequential revision of one's doxastic state by A and then by B. It does not strike us as being a unreasonable requirement to impose. It is also one that is perfectly consistent with the examples that we consider. However, given the latter, it is easy to see that IR entails that RED3 must fail too. See Appendix, Observation **??**.

attend to the issue of the logic of evidential relevance.

APPENDIX

In the proofs that follow, we shall be appealing to the following principles, which are not defined in the main body of the paper. They are to be read as holding for all doxastic states Ψ and all sentences A, B, C:

(AGM2) $A \in [\Psi * A]$

(AGM3) If $B \in [\Psi * A]$, then $B \in \mathrm{Cn}([\Psi] \cup \{A\})$

(AGM4) If $\neg A \notin [\Psi]$ and $B \in \mathrm{Cn}([\Psi] \cup \{A\})$, then $B \in [\Psi * A]$

(AGM5) If A is consistent, then $[\Psi * A]$ is also consistent

(AGM7) $[\Psi * A \wedge B] \subseteq \mathrm{Cn}([\Psi * A] \cup \{B\})$

(AGM8) If $\neg B \notin [\Psi * A]$, then $\mathrm{Cn}([\Psi * A] \cup \{B\}) \subseteq [\Psi * A \wedge B]$

(DP1) If $C \in \mathrm{Cn}(A)$, then $B \in [\Psi * A]$ iff $B \in [(\Psi * C) * A]$

(DP2) If $\neg C \in \mathrm{Cn}(A)$, then $B \in [\Psi * A]$ iff $B \in [(\Psi * C) * A]$

(DP3) If $B \in [\Psi * A]$, then $B \in [(\Psi * B) * A]$

(DP4) If $\neg B \notin [\Psi * A]$, then $\neg B \notin [(\Psi * B) * A]$

Observation 1. *An agent's single-shot revision dispositions with respect to any truth-functional combination C of sentences A and B are fully determined by the restriction of the preference relation to the members of the sets of maximal $A \wedge B$-, $A \wedge \neg B$-, $\neg A \wedge B$- and $\neg A \wedge \neg B$-worlds.*

Proof of Observation **??**. Let $\geq$ be a complete weak ordering of the set W of possible worlds and $>$ its strict part. Let $[\![\psi]\!]$ denote $\{w \in W : w \models \psi\}$ and, where $S \subseteq W$, $\max(S)$ denote $\{x \in S : \text{For all } y \in S, x \geq y\}$. Since C is a truth-functional combination of A and B, the set $[\![C]\!]$ will be equal to the union of one or more of the cells of the partition $\mathcal{P} = \{[\![A \wedge B]\!], [\![A \wedge \neg B]\!], [\![\neg A \wedge B]\!], [\![\neg A \wedge \neg B]\!]\}$ of W. We now show that $x \in \max([\![C]\!])$ iff $x \in \max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\})$, the nature of this last set being determined, as required, by the restriction of the preference relation to the members of the sets of maximal $A \wedge B$-, $A \wedge \neg B$-, $\neg A \wedge B$- and $\neg A \wedge \neg B$-worlds.

Regarding the left-to-right direction: Assume that $x \in \max([\![C]\!])$. Now assume for reductio that $x \notin \max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\})$. Now either (i) $x \in \bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\}$ or (ii) $x \notin \bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\}$. Assume (i). Then there exists a y in $\max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\})$, such that $y > x$. Since $\max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\}) \subseteq [\![C]\!]$, we also have $y \in [\![C]\!]$, contradicting our initial assumption. So assume (ii). Since $\mathcal{P}$ partitions $[\![C]\!]$, there exists an $S \in \mathcal{P}$ such that $x \in S$. Given (ii), we know that $x \notin \max(S)$. So there exists a $y \in \max(S)$ such that $y > x$. Since $S \subseteq [\![C]\!]$, we also have $y \in [\![C]\!]$, again contradicting our initial assumption.

Regarding the right-to-left direction: Assume that $x \in \max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\})$. Assume for reductio that $x \notin \max([\![C]\!])$ and hence that there exists a $y \in [\![C]\!]$ such that $y > x$. Since $\mathcal{P}$ partitions $[\![C]\!]$, there exists an $S \in \mathcal{P}$ such that $y \in S$ and a $z \in \max(S)$ such that $z \geq y$. But since $x \in \max(\bigcup\{\max(S) : S \in \mathcal{P} \text{ and } S \subseteq [\![C]\!]\})$, we have $x \geq z$ and hence, by transitivity of $\geq$, $x \geq y$. Contradiction. □

Observation 2. *In the presence of the AGM postulates, the DP postulates are jointly equivalent to DP′.*

Proof of Observation ??. Regarding the left-to-right direction: We consider three cases:

(1) Suppose $A \in [(\Psi * A) * B]$. Then, by AGM7 and AGM8, it follows that $[(\Psi * A) * B] = [(\Psi * A) * A \wedge B]$. But from DP1, we know that $[(\Psi * A) * A \wedge B] = [\Psi * A \wedge B]$. Hence $[(\Psi * A) * B] = [\Psi * A \wedge B]$, as required.

(2) Suppose $A, \neg A \notin [(\Psi * A) * B]$. By AGM8, this gives $[(\Psi * A) * B] \subseteq [(\Psi * A) * \neg A \wedge B] \cap [(\Psi * A) * A \wedge B]$ while the converse inclusion to this also holds by AGM7. Hence $[(\Psi * A) * B] = [(\Psi * A) * \neg A \wedge B] \cap [(\Psi * A) * A \wedge B]$. Applying DP1 and DP2 to the right-hand side yields $[(\Psi * A) * B] = [\Psi * \neg A \wedge B] \cap [\Psi * A \wedge B]$. We now split into two cases: (i) $\neg A \in [\Psi * B]$ and (ii) $\neg A \notin [\Psi * B]$. Assume (i). It follows that $[\Psi * \neg A \wedge B] = [\Psi * B]$ and we recover the desired conclusion. Assume (ii). then we also have $A \notin [\Psi * B]$ from DP3 and the fact that $A \notin [(\Psi * A) * B]$. Hence, by AGM8 and AGM7, we have $[\Psi * B] = [\Psi * \neg A \wedge B] \cap [\Psi * A \wedge B]$. Hence $[(\Psi * A) * B] = [\Psi * B]$. But since $[\Psi * B] \subseteq [\Psi * A \wedge B]$, we have $[(\Psi * A) * B] = [\Psi * B] \cap [\Psi * A \wedge B]$, as required.

(3) Suppose $\neg A \in [(\Psi * A) * B]$. Then, by AGM7 and AGM8, $[(\Psi * A) * B] = [(\Psi * A) * \neg A \wedge B]$. By DP2, we have $[(\Psi * A) * \neg A \wedge B] = [\Psi * \neg A \wedge B]$ and so $[(\Psi * A) * B] = [\Psi * \neg A \wedge B]$. From $\neg A \in [(\Psi * A) * B]$ and DP4 we know that $\neg A \in [\Psi * B]$. Hence, by AGM7 and AGM8, we have $[\Psi * B] = [\Psi * \neg A \wedge B]$ and so $[(\Psi * A) * B] = [\Psi * B]$, as required.

Regarding the right-to-left direction:

(1) Regarding DP1: If $A \in Cn(B)$, then we must be in Case (i) of DP′ and so $[(\Psi * A) * B] = [\Psi * A \wedge B]$. Since $A \wedge B$ and B are logically equivalent, we know that $[\Psi * A \wedge B] = [\Psi * B]$ and so $[(\Psi * A) * B] = [\Psi * B]$.

(2) Regarding DP2: If $\neg A \in Cn(B)$, then we must be in Case (iii) of DP′, which immediately gives us $[(\Psi * A) * B] = [\Psi * B]$.

(3) Regarding DP3: We prove the contrapositive. Assume $A \notin [(\Psi * A) * B]$. If $\neg A \in [(\Psi * A) * B]$, then $[(\Psi * A) * B] = [\Psi * B]$ from Clause (iii) of DP′ and so $A \notin [\Psi * B]$, as required. If $\neg A \notin [(\Psi * A) * B]$, then $[(\Psi * A) * B] = [\Psi * A \wedge B] \cap [\Psi * B]$ by Clause (ii) of DP′. We know that $A \in [\Psi * A \wedge B]$, so if $A \notin [(\Psi * A) * B]$, then we must have $A \notin [\Psi * B]$, again as required.

(4) Regarding DP4: If $\neg A \in [(\Psi * A) * B]$, then we must be in Case (iii) of DP′, so $[(\Psi * A) * B] = [\Psi * B]$ and hence $\neg A \in [\Psi * B]$. □

Observation 3. *AGM2, AGM4 and DP3 jointly entail that $A \in [(\Psi * A) * B]$ if $A \in [\Psi * B]$ or $\neg B \notin [\Psi * A]$.*

Proof of Observation ??. Assume that $\neg B \notin [\Psi * A]$. By AGM2, $A \in [\Psi * A]$. It then follows by AGM4 that $A \in [(\Psi * A) * B]$. Assume that $A \in [\Psi * B]$. It follows by DP3 that $A \in [(\Psi * A) * B]$. □

Observation 4. *AGM2 entails that $\neg A \in [(\Psi * A) * B]$ if $\neg A \in Cn(B)$ and $\neg B \notin Cn(\varnothing)$.*

Proof of Observation ??. Trivial: Assume that $\neg A \in \mathrm{Cn}(B)$ and that $\neg B \notin \mathrm{Cn}(\varnothing)$. From the latter, by AGM2, we have $B \in [(\Psi * A) * B]$. By deductive closure of belief sets, and the fact that $\neg A \in \mathrm{Cn}(B)$, it then follows from this that $\neg A \in [(\Psi * A) * B]$. □

Observation 5. *The following two statements are equivalent:*

(1) If two doxastic states are 1-equivalent, then they are 2-equivalent

(2) If two doxastic states are 1-equivalent, then they are equivalent

Proof of Observation ??. Clearly (2) entails (1), from the definitions of equivalence and k-equivalence. It remains to show that (1) entails (2). So suppose that (a) holds and that Ψ and Ψ' are 1-equivalent. We will show by induction on k that Ψ and Ψ' are k-equivalent for all k. The base case, $k = 1$, holds by assumption. Regarding the inductive step, assume that Ψ and Ψ' are k-equivalent. We need to show that they are are $(k+1)$-equivalent, i.e. that for any $(k+1)$-tuple $\langle A_1, A_2, \ldots, A_k, A_{k+1}\rangle$, $[((((\Psi * A_1) * A_2) * \ldots) * A_k) * A_{k+1}] = [((((\Psi' * A_1) * A_2) * \ldots) * A_k) * A_{k+1}]$. Since Ψ and Ψ' are k-equivalent, we know that $[(((\Psi * A_1) * A_2) * \ldots) * A_{k-1}]$ and $[(((\Psi' * A_1) * A_2) * \ldots) * A_{k-1}]$ are 1-equivalent. Hence, by (1), they are also 2-equivalent and so $[((((\Psi * A_1) * A_2) * \ldots) * A_k) * A_{k+1}] = [((((\Psi' * A_1) * A_2) * \ldots) * A_k) * A_{k+1}]$, as required. □

Observation 6. *In the presence of the AGM and DP postulates, NR, RR and LR are respectively equivalent to*

$$\text{(NR}') \quad [(\Psi * A) * B] = \begin{cases} [\Psi * A \wedge B], & \text{if } \neg B \notin [\Psi * A] \\ [\Psi * B], & \text{otherwise} \end{cases}$$

$$\text{(RR}') \quad [(\Psi * A) * B] = \begin{cases} [\Psi * A \wedge B], & \text{if } \neg A \notin [\Psi * B] \text{ or } \neg B \notin [\Psi * A] \\ [\Psi * B], & \text{otherwise} \end{cases}$$

$$\text{(LR}') \quad [(\Psi * A) * B] = \begin{cases} [\Psi * A \wedge B], & \text{if } K * A \wedge B \text{ is consistent} \\ [\Psi * B], & \text{otherwise} \end{cases}$$

Proof of Observation ??. Regarding the equivalence between NR and NR′:

- From NR to NR′: Assume $\neg B \notin [\Psi * A]$. It follows by NR that $A \in [(\Psi * A) * B]$. By DP′, we then recover $[(\Psi * A) * B] = [\Psi * A \wedge B]$, as required. Assume $\neg B \in [\Psi * A]$. Now either (i) $\neg A \notin [\Psi * B]$, or (ii) $\neg A \in [\Psi * B]$. Assume (i). Then, by NR, $\neg A \in [(\Psi * A) * B]$ and hence, by DP′, $[(\Psi * A) * B] = [\Psi * B]$, as required. Assume (ii). On the one hand, it follows by AGM8 that $\mathrm{Cn}([\Psi * B] \cup \{A\}) \subseteq [\Psi * A \wedge B]$ and hence that (iii) $[\Psi * B] \subseteq [\Psi * A \wedge B]$. On the other hand, it it follows by NR that $A, \neg A \notin [(\Psi * A) * B]$ and hence, by DP′, that (iv) $[(\Psi * A) * B] = [\Psi * A \wedge B] \cap [\Psi * B]$. By (iii) and (iv), we have $[(\Psi * A) * B] = [\Psi * B]$, as required.

- From NR′ to NR: Assume that $A \in [\Psi * B]$ or $\neg B \notin [\Psi * A]$. From the latter, by NR′, we recover $[(\Psi * A) * B] = [\Psi * A \wedge B]$ and hence, by AGM2 and closure of belief sets, $A \in [(\Psi * A) * B]$, as required. Assume that $\neg B \in [\Psi * A]$. From this, by NR′, we recover $[(\Psi * A) * B] = [\Psi * B]$. Assume further that $A, \neg A \notin [\Psi * B]$. It follows that $A, \neg A \notin [(\Psi * A) * B]$, as required. Finally, alternatively, assume that $\neg A \notin [\Psi * B]$. It follows that $\neg A \in [(\Psi * A) * B]$, again as required.

Regarding the equivalence between RR and RR′:

- From RR to RR′: Assume that either $\neg A \notin [\Psi * B]$ or $\neg B \notin [\Psi * A]$. By RR, we have $A \in [(\Psi * A) * B]$ and hence, by DP′, $[(\Psi * A) * B] = [\Psi * A \wedge B]$, as required. Assume instead that $\neg A \in [\Psi * B]$ and $\neg B \in [\Psi * A]$. By RR, we have $\neg A \in [(\Psi * A) * B]$ and hence, by DP′, $[(\Psi * A) * B] = [\Psi * B]$, as required.

- From RR′ to RR: Assume that either $\neg A \notin [\Psi * B]$ or $\neg B \notin [\Psi * A]$. By RR′ it follows that $[(\Psi * A) * B] = [\Psi * A \wedge B]$ and hence, by AGM2 and closure of belief sets, that $A \in [(\Psi * A) * B]$, as required. So assume instead that $\neg A \in [\Psi * B]$ and $\neg B \in [\Psi * A]$. It follows by RR′ that $[(\Psi * A) * B] = [\Psi * B]$ and hence, since $\neg A \in [\Psi * B]$, that $\neg A \in [(\Psi * A) * B]$, as required.

Regarding the equivalence between LR and LR′:

- From LR to LR′: Assume that $[\Psi * A \wedge B]$ is consistent. By AGM2, it follows that $\neg B \notin \mathrm{Cn}(A)$ and hence, by LR, that $A \in [(\Psi * A) * B]$. By DP′, we then recover the required result that $[(\Psi * A) * B] = [\Psi * A \wedge B]$. Assume instead that $[\Psi * A \wedge B]$ is inconsistent. By AGM5, it follows that $\neg B \in \mathrm{Cn}(A)$. By LR, we therefore have $\neg A \in [(\Psi * A) * B]$ and hence, by DP′, $[(\Psi * A) * B] = [\Psi * B]$, as required.

- From LR′ to LR: Assume that $\neg B \notin \mathrm{Cn}(A)$. It follows, by AGM5, that $[\Psi * A \wedge B]$ is consistent. and hence, by LR′, that $[(\Psi * A) * B] = [\Psi * A \wedge B]$. By AGM2 and closure of belief sets, we then recover $A \in [(\Psi * A) * B]$, as required. Assume instead that $\neg B \in \mathrm{Cn}(A)$. Since, by AGM2, we have $B \in [(\Psi * A) * B]$, it then follows by closure of belief sets that $\neg A \in [(\Psi * A) * B]$, as required. □

Observation 7. *RED3, IR and the negation of RED2 are jointly inconsistent.*

*Proof of Observation **??**.* Assume the negation of RED2, i.e. that there exist sentences A and B and doxastic states Ψ and Ψ', such that $[\Psi * C] = [\Psi' * C]$, for any truth-functional combination C of A and B, but $[(\Psi * A) * B] \neq [(\Psi' * A) * B]$. By IR, there then exists a Ψ'' such that $[\Psi'' * C] = [\Psi' * C]$, for any C, and $[(\Psi'' * A) * B] = [(\Psi * A) * B]$. Since, by assumption, $[(\Psi * A) * B] \neq [(\Psi' * A) * B]$, we therefore have $[(\Psi'' * A) * B] \neq [(\Psi' * A) * B]$. But this contradicts RED3, which would require, since $[\Psi'' * C] = [\Psi' * C]$, for any C, that $[(\Psi'' * A) * B] = [(\Psi' * A) * B]$. □

REFERENCES

Alchourrón, C., P. Gärdenfors & D. Makinson (1985). On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic* L, 2: 118–139.

Booth, R., S. Chopra, and T. Meyer (2005). Restrained Revision. *Proceedings of the Sixth Workshop on Nonmonotonic Reasoning, Action and Change (NRAC).*

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

Boutilier, C. (1996). Iterated Revision and Minimal Change of Conditional Beliefs. *Journal of Philosophical Logic*, XXV, 3: 263–305.

Darwiche, A. & Pearl, J. (1997). On the Logic of Iterated Belief Revision. *Artificial Intelligence*, LXXXIX, 1–2:1–29.

Grove, A. (1988). Two Modellings for Theory Change, *Journal of Philosophical Logic*, XVII, 2: 157-170.

Hansson, S.O. (1992). In Defense of the Ramsey Test. *Journal of Philosophy*, LXXXIX, 10: 522–540.

Katsuno, H. and A. O. Mendelzon (1991). Propositional Knowledge Base Revision and Minimal Change. *Artificial Intelligence*, LII, 3:263–294.

Nayak, A. C. (1994). Iterated Belief Change Based on Epistemic Entrenchment. *Erkenntnis*, XLI, 3:353–390.

Nayak, A. C., M. Pagnucco, P. Peppas (2003). Dynamic Belief Revision Operators *Artificial Intelligence*, 146: 193–228.

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*, XXVIII. Dordrecht: Springer.

Rott, H. (2011). Reapproaching Ramsey: Conditionals and Iterated Belief Change in the Spirit of AGM. *Journal of Philosophical Logic*, 40(2), pp. 155–91.

Stalnaker, R. (2009). Iterated Belief Revision. *Erkenntnis*, LXX, 2: 189–209.