A Bad Day Surfing is Better than a Good Day Working: How to Revise a Total Preorder

Richard Booth
Faculty of Informatics
Mahasarakham University
Mahasarakham 44150, Thailand
richard.b@msu.ac.th

Thomas Meyer and **Ka-Shu Wong**
National ICT Australia and
School of Computer Science and Engineering
UNSW, Sydney 2052, Australia
Thomas.Meyer@nicta.com.au, kswong@cse.unsw.edu.au

Abstract

Most approaches to *iterated belief revision* are accompanied by some motivation for the use of the proposed revision operator (or family of operators), and typically encode enough information for uniquely determining one-step revision. But in those approaches describing a family of operators, there is usually little indication of how to proceed uniquely after the first revision step. In this paper we take a step towards addressing that deficiency by providing a formal framework which goes beyond the first revision step. The framework is obtained by enriching the preference information starting from the following intuitive idea: we associate to each world x two abstract objects x^+ and x^-, with the intuition that x^+ represents x "on a good day", while x^- represents x "on a bad day", and we assume that, *in addition to* preferences over the set of worlds, we are given preferences over this set of objects as well. The latter can be considered as meta-information which enables us to go beyond the first revision step of the revision operator being applied.

Introduction

Total preorders (hereafter *tpos*) are used to represent preferences in many contexts. In particular they are a common tool in *belief revision* (Grove 1988; Katsuno & Mendelzon 1991; Rott 2001). In that setting they are taken to stand for plausibility orderings on the set of propositional worlds, which are used to encode the *dispositions* for change, or the *conditional beliefs* of an agent. The associated *belief set* is taken to be the set of those sentences true in all the minimal worlds. When new evidence α comes in, the plausibility ordering is used to calculate the new belief set, usually by setting it to be the set of those sentences true in all the minimal models of α. This ensures a unique new belief set, but does not provide enough information to obtain a new tpo, which may then serve as the target for the *next* revision input. Thus the question of modelling the dynamics of *tpos* is of critical importance to the problem of *iterated belief revision*.

The past ten years has seen a flurry of activity in this area, with (Darwiche & Pearl 1997) and (Nayak, Pagnucco, & Peppas 2003) being representative examples. Most approaches devote considerable effort to motivating the use of their proposed revision operator (or family of operators). But in those approaches describing a family of operators, there is usually little (or no) indication of how to choose among the available operators. In this paper we make a contribution towards overcoming that deficiency by providing a formal framework which obtains a unique *tpo* following one revision step, thereby going beyond just the belief set resulting from the revision input. The framework is obtained by enriching the preference information encoded in the tpo starting from the following intuitive idea: when we compare two different worlds x and y according to preference, often we are able to imagine different contingencies, according to whether all goes well in x and y or not. For example, given a choice between spending the day surfing at the beach and spending it in the office, we might think that even a bad day surfing is preferable to a good day working. Our idea is to associate to each world x two abstract objects x^+ and x^-, with the intuition that x^+ represents x "on a good day", while x^- represents x "on a bad day", and we assume that, *in addition to* the given tpo $\leq$ over the set of worlds, we are given a tpo $\preceq$ over this set of objects.

This meta-information allows us to uniquely determine the new tpo: when new evidence α comes in it casts a more favourable light on those worlds in which α holds. Thus the evidence signals a "good day" for all those worlds satisfying α, and a "bad day" for the $\neg\alpha$-worlds. The revised tpo $\leq^*_\alpha$ is obtained by setting $x \leq^*_\alpha y$ iff $x^\epsilon \preceq y^\delta$, where $\epsilon, \delta \in \{+, -\}$ depending on whether x, y satisfy α or not.

As we will see, one commonly assumed rule from belief revision which will *not* generally hold for our revision operators is that the input α is necessarily an element of the *belief set* associated to $\leq^*_\alpha$. Thus, at the belief set level, we are in the realm of so-called *non-prioritised* revision (Hansson 1999; Hansson *et al.* 2001).

The plan of the paper is as follows. We begin in the next section by describing our enriched preference state. Then we show how to use this enrichment to define a unique tpo-revision operator, and we axiomatically characterise the resulting family of operators. Initially we describe the properties of this family on a *semantic* level, i.e., in terms of how the ordering of individual worlds x, y undergo change. In the following section we give an alternative, *sentential* formulation in terms of *conditional beliefs*, and introduce the notion of what it means for one sentence to *overrule* another in the context of a tpo-revision operator. After this we study some

notions of strict preference which can be extracted from $\preceq$ and show how these are closely related to the 'overrules' relation. Next we examine two known special cases of our family and give an example which shows how rigid use of either of these can sometimes lead to counter-intuitive results. In the penultimate section we describe and axiomatise an interesting sub-class of our family which remains general enough to include the two special cases, before concluding.

Preliminaries: We work in a propositional language L generated by finitely many propositional variables. We use $\vdash$ and $\equiv$ to denote classical logical consequence and classical logical equivalence respectively. We sometimes also use Cn to denote the operation of closure under classical logical consequence. W is the set of propositional worlds. Given $\alpha \in L$, we denote the set of worlds which satisfy α by $[\alpha]$. Given any set $S \subseteq W$ of worlds, $Th(S)$ will denote the set of sentences true in all the worlds in S. A tpo is a binary relation $\leq$ which is both transitive and connected (for any x, y either $x \leq y$ or $y \leq x$). In what follows we assume a fixed but arbitrary initial tpo $\leq$ over W which we wish to revise. $<$ will denote the strict part of $\leq$, and $\sim$ the symmetric closure of $\leq$ (i.e. $x \sim y$ iff both $x \leq y$ and $y \leq x$). We are interested in functions $*$ which, for each $\alpha \in L$, return a new ordering $\leq^*_\alpha$, and we will refer to any such $*$ as *a revision operator for* $\leq$.

Enriching the preference state

We let $W^{\pm} = \{x^\epsilon \mid x \in W \text{ and } \epsilon \in \{+, -\}\}$. We assume $x^\epsilon = y^\delta$ only if both $x = y$ and $\epsilon = \delta$. We suppose, along with $\leq$, we are given some relation $\preceq$ over $W^{\pm}$. We expect some basic conditions on $\preceq$ and its interrelations with $\leq$:

($\preceq$1) $\preceq$ is a tpo over $W^{\pm}$

($\preceq$2) $x^+ \preceq y^+$ iff $x \leq y$

($\preceq$3) $x^- \preceq y^-$ iff $x \leq y$

($\preceq$4) $x^+ \prec x^-$

($\preceq$2) and ($\preceq$3) say that the choice between two worlds both on a good day, resp. both on a bad day, should be precisely the same as that dictated by $\leq$. ($\preceq$4) just says that given the choice between x on a good day and x on a bad day, we should choose x on a good day.

Definition 1 *Let $\preceq \subseteq W^{\pm} \times W^{\pm}$. If $\preceq$ satisfies ($\preceq$1)–($\preceq$4) we say $\preceq$ is a* $\leq$-faithful tpo (over $W^{\pm}$).

A $\leq$-faithful tpo $\preceq$ can be given a useful graphical representation. First recall that any tpo $\leq'$ can be equivalently represented as its linearly ordered set of *ranks*. The ranks of $\leq'$ are the equivalence classes $[\![x]\!]_{\sim'}$ modulo the symmetric closure $\sim'$ of $\leq'$, and they are ordered by the relation $[\![x]\!] \leq' [\![y]\!]$ iff $x \leq' y$. By ($\preceq$2), resp. ($\preceq$3), if x and y are two worlds in the same $\leq$-rank, then x^+ and y^+, resp. x^- and y^-, are in the same $\preceq$-rank. Thus if $R_1 < \cdots < R_m$ are the ranks of $\leq$ we can represent $\preceq$ as a $2 \times m$ table of numbers whose i^{th} column corresponds to rank R_i, and whose top and bottom rows correspond to $+$ and $-$ respectively. Then $x^\epsilon \preceq y^\delta$ iff the entry in $(\epsilon, [\![x]\!])$ is less than or equal to the entry in cell $(\delta, [\![y]\!])$. By ($\preceq$2) and ($\preceq$3) the numbers increase strictly monotonically from left to right, while ($\preceq$4) decrees they increase strictly monotonically from top to bottom. An example assuming just three $\leq$-ranks is shown below:

	R_1	R_2	R_3
$+$	1	2	3
$-$	3	4	5

Figure 1: A graphical representation of $\preceq$.

As this example shows, there is nothing to stop the same number appearing in *both* a cell in the "$+$" row and a cell in the "$-$" row. So in the above we see that if the rank of world x is R_1 and the rank of world y is R_3 then x^- and y^+ appear in the *same* $\preceq$-rank. In other words, x on a bad day is *equally preferred* to y on a good day.

Revision operators defined from $\preceq$

Now given a $\leq$-faithful tpo $\preceq$ over $W^{\pm}$ we want to use the information given by $\preceq$ to define a revision operator $* = *_\preceq$ for $\leq$. The idea is that the evidence α casts a favourable light on those worlds satisfying α. In other words, we consider worlds satisfying α to be having a "good day", with those worlds inconsistent with the evidence having a "bad day". We set, for any $\alpha \in L$ and $x \in W$:

$$r_\alpha(x) = \begin{cases} x^+ & \text{if } x \in [\alpha] \\ x^- & \text{if } x \in [\neg\alpha] \end{cases}$$

The revised tpo $\leq^*_\alpha$ is defined by setting, for each $x, y \in W$,

$$x \leq^*_\alpha y \text{ iff } r_\alpha(x) \preceq r_\alpha(y).$$

Definition 2 *For each $\leq$-faithful tpo $\preceq$ over $W^{\pm}$, we call $*_\preceq$ as defined above the revision operator* generated by $\preceq$.

Example 1 Consider the propositional language generated by the atoms p and q. We represent worlds as sequences of 0s and 1s, representing the valuations of p and q respectively (thus 01 represents a world where p is false and q is true). Let $\leq$ be the ordering on worlds depicted in the following:

10	11 00	01

Let $\preceq$ be the $\leq$-faithful tpo depicted in Figure 1. Revision by q can be represented pictorially as follows:

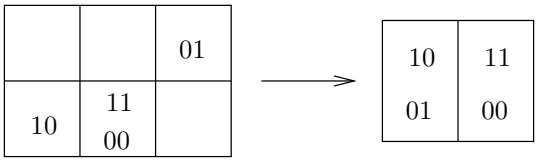
	11	01
10	00	

$\longrightarrow$

11	10 01	00

In the table on the left, worlds satisfying q are placed in the top row, with those not satisfying q placed in the bottom row.

The resulting ordering $\leq^*_q$, shown on the right, is obtained by reading the ranks from the corresponding cell in Figure 1. The resulting belief set, i.e., the set of sentences true in all the $\leq^*_q$-minimal worlds, is $Cn(p \wedge q)$. The revision of $\leq$ by $\neg p \wedge q$ can be similarly represented as follows:



This time the resulting belief set associated with $\leq^*_{\neg p \wedge q}$ is $Cn(p \leftrightarrow \neg q)$. Since $\neg p \wedge q \notin Cn(p \leftrightarrow \neg q)$, this example shows that new evidence is not always in the belief set associated to the new tpo.

What are the properties of $*_{\preceq}$? Consider the following list:

($*1$) $\leq^*_\alpha$ is a tpo over W
($*2$) $\alpha \equiv \gamma$ implies $\leq^*_\alpha = \leq^*_\gamma$
($*3$) If $x, y \in [\alpha]$ then $x \leq^*_\alpha y$ iff $x \leq y$
($*4$) If $x, y \in [\neg\alpha]$ then $x \leq^*_\alpha y$ iff $x \leq y$
($*5$) If $x \in [\alpha]$, $y \in [\neg\alpha]$ and $x \leq y$ then $x <^*_\alpha y$
($*6$) If $x \in [\alpha]$, $y \in [\neg\alpha]$ and $y \leq^*_\alpha x$ then $y \leq^*_\gamma x$
($*7$) If $x \in [\alpha]$, $y \in [\neg\alpha]$ and $y <^*_\alpha x$ then $y <^*_\gamma x$

($*1$) just says revising a tpo over W should result in another tpo over W, while ($*2$) is a syntax-irrelevance property. The next three rules are all familiar from the literature on iterated belief change. ($*3$) and ($*4$) appear respectively as (CR1) and (CR2) in Darwiche and Pearl's (1997) well-known list of four postulates. They say that after revising by α, the relative ordering between α-worlds, respectively $\neg\alpha$-worlds, remains unchanged. ($*5$) was proposed independently by Booth & Meyer (2006) and Jin & Thielscher (2005). It is easily seen to be stronger than the other two rules in the Darwiche-Pearl list (which can be obtained by replacing $\leq$ by $<$ (CR3) and $<^*_\alpha$ by $\leq^*_\alpha$ (CR4) respectively). It says if an α-world x was considered at least as preferred as a $\neg\alpha$-world y *before* receiving α, then after revision it should be considered *strictly* more preferred. These three rules were considered characteristic of a family of operators called *admissible* revision operators (Booth & Meyer 2006).

So far each of our rules mention only *one* revision input sentence α (modulo logical equivalence). By analogy with the AGM postulates for *belief set* revision (Alchourrón, Gärdenfors, & Makinson 1985), we might consider them as the set of *basic* postulates for tpo-revision. One thing largely missing from the literature on iterated belief change is a serious study of *supplementary* rationality properties which bestow a certain amount of *coherence* on the results of revising $\leq$ by *different* sentences. The last couple of properties do this. First, suppose evidence α is received, and let $x \in [\alpha]$, $y \in [\neg\alpha]$, but suppose $y \leq^*_\alpha x$. We propose that if x is not more preferred than y, *even after* receiving evidence which clearly points more to x being the case than it does to y, then there can be *no* evidence which will lead to x being more preferred to y. This is expressed by ($*6$). Similarly ($*7$) says if x is deemed *strictly* less preferred than y after receiving α then x must be strictly less preferred after receiving *any* input.

It turns out that these properties provide an exact characterisation of the revision operators we consider.

Theorem 1 *Let $*$ be any revision operator for $\leq$. Then $*$ is generated from some $\leq$-faithful tpo $\preceq$ over $W^{\pm}$ iff $*$ satisfies ($*1$)–($*7$).*

To show the completeness part of Theorem 1, starting from any revision operator $*$ for $\leq$ we can define an ordering $\preceq_*$ over $W^{\pm}$ as follows. Let $x, y \in W$ and $\delta, \epsilon \in \{+, -\}$. If $\delta = \epsilon$ then we set

$$x^\delta \preceq_* y^\delta \text{ iff } x \leq y.$$

This obviously ensures $\preceq_*$ complies with ($\preceq 2$) and ($\preceq 3$). Now suppose $\delta \neq \epsilon$. If $x = y$ then we simply set $x^+ \prec_* x^-$, to ensure compliance with ($\preceq 4$). Otherwise we set

$$x^+ \preceq_* y^- \text{ iff } x \leq^*_x y, \qquad x^- \preceq_* y^+ \text{ iff } x \leq^*_y y.$$

Here, when we use a world x as a subscript in $\leq^*_x$, we are using it to denote any sentence α such that $[\alpha] = \{x\}$ (note that if $*$ satisfies ($*2$) the precise choice of α is irrelevant). Then if $*$ satisfies ($*1$)–($*7$) then $\preceq_*$ is a $\leq$-faithful tpo and the revision operator generated from $\preceq_*$ is precisely $*$.

Some social choice-like conditions

In this subsection we discuss some more properties satisfied by our revision operators. These properties are recognisable as versions of properties familiar from the theory of *social choice*, or *preference aggregation* (Arrow 1963). The problem of preference aggregation is the problem of finding some function f which, given any list of tpos (over some given set X of *alternatives*) $\leq_1, \ldots, \leq_n$, with the $\leq_i$s representing the preferences over X of the *individuals* in a group, will return a new single ordering $f(\leq_1, \ldots, \leq_n)$ over X which adequately represents the preferences of the *group* as a whole. Now, we can think of our problem of determining $\leq^*_\alpha$ as a highly specialised case of this problem. To do this we need to repackage the new evidence $\alpha \in L$ into tpo-form. The simplest way to do this is as follows.

Definition 3 *For any $\alpha \in L$, the tpo $\leq^\alpha$* generated by *α is the tpo over W given by $x \leq^\alpha y$ iff $x \in [\alpha]$ or $y \in [\neg\alpha]$.*

In other words $\leq^\alpha$ is the tpo over W consisting of (at most) two ranks: the lower one containing all the α-worlds and the upper one containing all the $\neg\alpha$-worlds. Then we can think of revision of $\leq$ by α as an aggregation of $\leq$ with $\leq^\alpha$. (This manoeuver is also carried out by Glaister, 1998 and Nayak, 1994. An alternative way of generating tpos from sentences, based on the Hamming distance between two propositional worlds, is mentioned by Benferhat *et al.*, 2000.)

Many properties of preference aggregation operators have been proposed. One well-known property, known as the *Pareto* condition, is that, given two alternatives x and y, if every individual prefers x *at least as much as* y, and if at least one individual *strictly* prefers x over y, then the group should *strictly* prefer x over y. In our case, this condition translates into the following property:

(Pareto) If $x \leq y$ and $x \leq^{\alpha} y$, and at least one of these two inequalities is strict, then $x <^{*}_{\alpha} y$

The case of the above rule where $\leq^{\alpha}$ is strict is nothing other than ($*5$), while the case where $x \sim^{\alpha} y$ and $x < y$ is easily seen to follow mainly from ($*3$) and ($*4$). Thus we have:

Proposition 1 *Every revision operator $*$ generated by some $\leq$-faithful tpo $\preceq$ over $W^{\pm}$ satisfies (Pareto).*

Another well-known property from preference aggregation is known as *Independence of Irrelevant Alternatives*, which states that for any two alternatives x and y, the group preference between x and y should depend only on how each individual ranks x and y. More precisely, if we were to replace individual i's tpo $\leq_i$ by any other tpo $\leq'_i$ which ranks x and y in exactly the same way as $\leq$, then x and y would be ranked in exactly the same way in $f(\leq_1, \ldots, \leq'_i, \ldots \leq_n)$ as in $f(\leq_1, \ldots, \leq_i, \ldots \leq_n)$. It turns out that our family of operators satisfy a restricted version of this rule, which we call *Independence of Irrelevant Alternatives in the Input*. Given $\alpha, \gamma \in L$, and $x, y \in W$, let's say α and γ *agree on x and y* iff either both $x <^{\alpha} y$ and $x <^{\gamma} y$, or both $x \sim^{\alpha} y$ and $x \sim^{\gamma} y$, or both $y <^{\alpha} x$ and $y <^{\gamma} x$. In other words α and γ both "say the same thing" regarding the relative plausibility of x and y.

(IIA-Input) If α and γ agree on x and y then $x \leq^{*}_{\alpha} y$ iff $x \leq^{*}_{\gamma} y$

That this is a property of our family of tpo-revision operators can be straightforwardly shown by considering an arbitrary $\leq$-faithful tpo $\preceq$ over $W^{\pm}$. But in fact we can show the following:

Proposition 2 *Let $*$ be any revision operator for $\leq$ which satisfies ($*1$) and ($*3$)–($*5$). Then $*$ satisfies (IIA-Input) iff $*$ satisfies both ($*6$) and ($*7$).*

Thus, given the "basic" properties ($*1$)–($*5$) for tpo-revision, requiring $*$ to satisfy the two "supplementary" properties ($*6$) and ($*7$) amounts to enforcing (IIA-Input). Note this equivalence does not require the presence of the syntax-irrelevance property ($*2$). In fact, since sentences which are logically equivalent agree on *all* worlds x and y, we see that ($*2$) actually *follows from* (IIA-Input). Consequently, we have established that in the list ($*1$)–($*7$), property ($*2$) is *redundant*.

For more discussion on social choice-like conditions and their relevance to tpo-revision we refer the reader to the work of Glaister (1998).

On the sentential level

So far all our properties of tpo-revision operators have been expressed on the "semantic level", directly in terms of worlds. But there is also a *sentential* level on which we can recast our properties. For any tpo $\leq'$ over W and any $\beta \in L$ we let $\min(\beta, \leq')$ denote the set of $\leq'$-minimal elements of $[\beta]$, i.e., $\min(\beta, \leq') = \{x \in [\beta] \mid \nexists y \in [\beta] \text{ s.t. } y <' x\}$. Then we define:

$$\leq' \circ \beta = Th(\min(\beta, \leq')).$$

$\leq' \circ \beta$ represents what is believed in $\leq'$ on the *supposition* that β is the case. If $\lambda \in \leq' \circ \beta$ then we might also say $\beta \rightarrowtail \lambda$ is a *conditional belief* in $\leq'$. Note that we do not necessarily assume this is the same thing as saying λ would be believed after receiving β explicitly *as evidence*. This is because we want to support non-prioritised revision, so in particular β itself might not necessarily be believed after receiving it as evidence (it might be simply too far-fetched). Nevertheless, new evidence will have some impact on the set of conditional beliefs. Note that this notation enables us to denote the belief set associated to $\leq'$ by $\leq' \circ \top$.

We can give all the properties ($*2$)–($*7$) an equivalent formulation in terms of $\circ$, thus giving a set of sound and complete properties for our family of revision operators which has a different flavour:

($\circ 2$) If $\alpha \equiv \gamma$ then $\leq^{*}_{\alpha} \circ \beta = \leq^{*}_{\gamma} \circ \beta$

($\circ 3$) If $\beta \vdash \alpha$ then $\leq^{*}_{\alpha} \circ \beta = \leq \circ \beta$

($\circ 4$) If $\beta \vdash \neg\alpha$ then $\leq^{*}_{\alpha} \circ \beta = \leq \circ \beta$

($\circ 5$) If $\neg\alpha \notin \leq \circ \beta$ then $\alpha \in \leq^{*}_{\alpha} \circ \beta$

($\circ 6$) If $\alpha \notin \leq^{*}_{\alpha} \circ \beta$ then $\alpha \notin \leq^{*}_{\gamma} \circ \beta$

($\circ 7$) If $\neg\alpha \in \leq^{*}_{\alpha} \circ \beta$ then $\neg\alpha \in \leq^{*}_{\gamma} \circ \beta$

($\circ 2$) just says revising by logically equivalent sentences yields the same set of conditional beliefs. ($\circ 3$) and ($\circ 4$) are essentially the well-known (C1) and (C2) of Darwiche & Pearl (1997), while ($\circ 5$) corresponds to rule (P) of Booth & Meyer (2006), also referred to as *Independence* by Jin & Thielscher (2005). The correspondences between these last three rules and their counterparts in the previous section were proved in those papers. (Although these papers all assume the "prioritised" setting for belief revision in which revision inputs are always believed after revision.) The last two rules are neatly explained with the help of the following terminology:

Definition 4 *Given any revision operator $*$ for $\leq$ and given $\alpha, \beta \in L$, we shall say β* overrules *α (relative to $*$) iff either β is inconsistent or $\alpha \notin \leq^{*}_{\alpha} \circ \beta$. We shall say β* strictly overrules *α (relative to $*$) iff $\neg\alpha \in \leq^{*}_{\alpha} \circ \beta$.*

The inclusion of the clause "β is inconsistent" in the definition of "overrules" allows for a smoother exposition. This way we get the intuitively expected chain of implications $\beta \vdash \neg\alpha$ implies β strictly overrules α, which implies β overrules α. If $*$ satisfies ($\circ 5$) then this in turn implies $\neg\alpha \in \leq \circ \beta$. Now suppose that evidence γ is received and we then make a further supposition β. ($\circ 6$) says if β overrules α and β is consistent then α will not be believed, while ($\circ 7$) says if β *strictly* overrules α then α will be *rejected*.

Proposition 3 *Let $*$ be a revision operator for $\leq$ which satisfies ($*1$). Then for each $i = 2, \ldots, 7$, $*$ satisfies ($*i$) iff $*$ satisfies ($\circ i$).*

Corollary 1 *Let $*$ be a revision operator for $\leq$. Then $*$ is generated from some $\leq$-faithful tpo $\preceq$ over $W^{\pm}$ iff $*$ satisfies ($*1$) and ($\circ 2$)–($\circ 7$).*

This sentential reformulation is useful, since there are some interesting properties which can be formulated in sentential terms, but for which obvious semantic counterparts do not exist. For example:

(Disj1) $(\leq^*_\alpha \circ\beta) \cap (\leq^*_\gamma \circ\beta) \subseteq (\leq^*_{\alpha\vee\gamma} \circ\beta)$

(Disj2) $(\leq^*_{\alpha\vee\gamma} \circ\beta) \subseteq (\leq^*_\alpha \circ\beta) \cup (\leq^*_\gamma \circ\beta)$

These two properties were essentially first proposed by Lehmann, Magidor, & Schlechta (2001), and seem to be natural properties to have. The first one says if a conditional belief is held both after receiving evidence α and after receiving evidence γ, then it is also held after receiving their disjunction as evidence. The second one says a conditional belief is not held after receiving a disjunction as evidence, *without* being held after receiving just one of the disjuncts in isolation.

Proposition 4 *Every revision operator $*$ generated from some $\leq$-faithful tpo $\preceq$ over $W^\pm$ satisfies (Disj1) and (Disj2).*

We prove this result by considering an arbitrary $\leq$-faithful tpo $\preceq$, rather than trying to derive these rules syntactically from ($*1$) and ($\circ 2$)–($\circ 7$). A key property used in the proof is that, for any $\alpha, \gamma \in L$ and $x \in W$, $r_{\alpha\vee\gamma}(x) = \min\{r_\alpha(x), r_\gamma(x)\}$.

The next result shows that $\leq^*_\alpha \circ\beta$ falls neatly into one of three categories. Note we don't need ($\circ 6$) and ($\circ 7$) for this.

Proposition 5 *Let $*$ be any revision operator for $\leq$ satisfying ($*1$) and ($\circ 2$)–($\circ 5$), and let the overrules relations be given relative to $*$. Then for all $\alpha, \beta \in L$,*

$$\leq^*_\alpha \circ\beta = \begin{cases} \leq \circ(\alpha \wedge \beta) & \text{if } \beta \text{ doesn't overrule } \alpha \\ (\leq \circ(\alpha \wedge \beta)) \cap (\leq \circ\beta) & \\ & \text{if } \beta \text{ overrules } \alpha, \text{ but not strictly} \\ \leq \circ\beta & \text{if } \beta \text{ strictly overrules } \alpha \end{cases}$$

Thus if β doesn't overrule α then making the supposition β after receiving α as evidence is the same as supposing α and β *together* in the initial tpo $\leq$. If β *strictly* overrules α then evidence α is just ignored when making the further supposition β. In the intermediate case where β overrules α, but not strictly, supposing β following evidence α results in a mixture of these two.

In particular note what happens when $\beta \equiv \top$. We see that $\leq^*_\alpha \circ\top$ equals either *(i)* $\leq \circ\alpha$, or *(ii)* $(\leq \circ\alpha) \cap (\leq \circ\top)$, or *(iii)* $\leq \circ\top$. Thus either the evidence is fully incorporated into the belief set using the AGM revision operator corresponding to $\leq$ (Katsuno & Mendelzon 1991) (case *(i)*), or the belief set remains unchanged (case *(iii)*), *or* there is an intermediate possibility (*(ii)*), which amounts to removing $\neg\alpha$ from the initial belief set using the AGM contraction operator corresponding to $\leq$. That is, we don't commit to believing the evidence, but we leave open the possibility that it *might* hold. We will have more to say on these notions of overruling in the next section.

Notions of strict preference

In this section we shall assume a fixed $\leq$-faithful tpo $\preceq$ over $W^\pm$. Given $\preceq$ we can define two more interesting preference orderings over W:

$$x \ll y \text{ iff } x^- \preceq y^+, \qquad x \lll y \text{ iff } x^- \prec y^+$$

In other words, $x \ll y$, resp. $x \lll y$, is saying that x, even on a bad day, is at least as preferred as, resp. strictly preferred to, y. The next proposition collects some properties of these two orderings.

Proposition 6
(i) $\lll \subseteq \ll \subseteq <$ *(where recall $<$ is the strict part of the initial tpo $\leq$).*
(ii) $\lll$ *and* $\ll$ *are both strict partial orders (i.e., irreflexive and transitive).*
(iii) $\lll$ *and* $\ll$ *both satisfy the* filtered *condition (Freund 1991), i.e., for all $x, y \in W$ and $\beta \in L$, if $x, y \in [\beta] \setminus \min(\beta, <')$ then there exists $z \in [\beta]$ such that $z <' x$ and $z <' y$.*

(Recall for a strict partial order $<'$, $\min(\beta, <') = \{x \in [\beta] \mid \nexists\, y \in [\beta] \text{ s.t. } y <' x\}$.) By *(i)* we see $<$, $\ll$ and $\lll$ form progressively more stringent notions of strict preference. If we let $* = *_\preceq$ then we see $x \lll y$ implies $r_\gamma(x) \prec r_\gamma(y)$ for all $\gamma \in L$, and so $x <^*_\gamma y$ for any γ. Thus $\lll$ can also be viewed as a set of *core*, or *protected* strict preferences in $<$ which are always preserved in any revision. Meanwhile we have $x \ll y$ implies $x \leq^*_\gamma y$ for any γ. Thus $\ll$ may be viewed as a set of *weakly protected* strict preferences, in the sense that if $x \ll y$ then no evidence will ever cause this preference to be *reversed*.

It turns out that these relations $\ll$ and $\lll$ are closely related to the notions of overruling and strict overruling from Definition 4.

Proposition 7 *Let the overrules relations be given relative to $*_\preceq$. Then* (i) *β overrules α iff* $\min(\beta, \ll) \subseteq [\neg\alpha]$. (ii) *$\beta$ strictly overrules α iff* $\min(\beta, \lll) \subseteq [\neg\alpha]$.

For each of the two overrules relations we may consider an interdefinable *inference* relation. We define:

$$\beta \Rightarrow \alpha \text{ iff } \beta \text{ overrules } \neg\alpha$$

$$\beta \Rrightarrow \alpha \text{ iff } \beta \text{ strictly overrules } \neg\alpha.$$

Using fundamental results by Freund (1991) and Kraus, Lehmann, & Magidor (1990), classifying various families of nonmonotonic inference relations, Proposition 7 together with the properties of $\ll$ and $\lll$ now allows us to deduce many properties of $\Rightarrow$ and $\Rrightarrow$, and thereby of the overrules relations:

Corollary 2 *The binary relations $\Rightarrow$ and $\Rrightarrow$ are both (consistency-preserving)* preferential *inference relations, in the sense of Kraus, Lehmann, & Magidor (1990). Furthermore they both satisfy the rule of* Disjunctive Rationality, *i.e., if $\beta \vee \gamma \Rightarrow \alpha$ then either $\beta \Rightarrow \alpha$ or $\gamma \Rightarrow \alpha$.*

The first part is a consequence of the fact that $\ll$ and $\lll$ are strict partial orders (Kraus, Lehmann, & Magidor 1990). In particular it implies $\Rightarrow$ and $\Rrightarrow$ **both** satisfy the following rules (among others):

$$\frac{\beta \Rightarrow \alpha,\ \alpha \vdash \gamma}{\beta \Rightarrow \gamma} \quad \text{(Right Weakening)}$$

$$\frac{\beta \Rightarrow \alpha,\ \beta \Rightarrow \gamma}{\beta \Rightarrow \alpha \wedge \gamma} \quad \text{(And)}$$

$$\frac{\beta \Rightarrow \alpha,\ \beta \Rightarrow \gamma}{\beta \wedge \gamma \Rightarrow \alpha} \quad \text{(Cautious Monotony)}$$

Switching things around in terms of the corresponding overrules relations, Right Weakening implies if β (strictly)

overrules α then β (strictly) overrules every sentence logically *stronger* than α. The And-rule tells us that if β (strictly) overrules both α and γ separately, then it (strictly) overrules their *disjunction*. While Cautious Monotony translates into the rule that if β (strictly) overrules α, then so does $\beta \wedge \neg\gamma$, *provided* β (strictly) overrules γ.

The second part of Corollary 2 follows from results by Freund (1991) and Proposition 6*(iii)*. It implies a disjunction $\beta \vee \gamma$ cannot (strictly) overrule α without at least one of its disjuncts doing so. However it's possible for neither $\Rightarrow$ nor $\Rrightarrow$ to satisfy the well-known rule Rational Monotony (Kraus, Lehmann, & Magidor 1990) (and thus also Monotony). I.e., if $\beta \Rightarrow \alpha$ and $\beta \not\Rightarrow \neg\gamma$ then $\beta \wedge \gamma \Rightarrow \alpha$. This is because it can be shown that the relations $\ll$ and $\lll$ are not in general *modular*, i.e., they do not verify the property $x <' y$ implies there exists z such that either $x <' z$ or $z <' y$. (A counterexample for $\ll$ can be found by taking the initial tpo from Example 1 with the $\preceq$ defined earlier in Figure 1, and then by taking $x = 10$ and $y = 01$.)

Limiting cases

In this section we investigate some special limiting cases of our family of revision operators. Firstly, suppose we insist on the following strengthening of property ($\preceq$4):

($\preceq$L) $\quad x^+ \prec y^-$.

In other words, given a choice between any world on a good day and any world on a bad day, we choose the world on a good day every time. This is equivalent to the limiting case where $\ll \;=\; \emptyset$ (thus also $\lll \;=\; \emptyset$). Hence this condition can be thought of as expressing "minimal confidence" behind the initial tpo $\leq$. Note that adding this rule to ($\preceq$2) and ($\preceq$3) is enough to specify a unique tpo over $W^{\pm}$, thus causing ($\preceq$1) to become redundant. Indeed we are left with the tpo defined by, for all $x, y \in W$ and $\delta, \epsilon \in \{+, -\}$, $x^\delta \preceq y^\epsilon$ iff either ($\delta = +$ and $\epsilon = -$) or ($\delta = \epsilon$ and $x \leq y$). In terms of the graphical representation of $\preceq$, this corresponds to the case where every number in the "+" row is *strictly less than* every number in the "−" row:

	R_1	R_2	$\cdots$	R_n
+	1	2	$\cdots$	n
−	$n+1$	$n+2$	$\cdots$	$2n$

The revision operator $*_{\mathrm{L}}$ defined by this $\preceq$ then reduces to:

$$x \leq^*_\alpha y \text{ iff either } x <^\alpha y \text{ or } (x \sim^\alpha y \text{ and } x \leq y)$$

This is the well-known *lexicographic* revision operator studied and axiomatised in the context of iterated belief revision (Glaister 1998; Spohn 1988; Nayak, Pagnucco, & Peppas 2003). It amounts to $\leq^\alpha$ being refined by $\leq$. We can characterise $*_{\mathrm{L}}$ within our family in the following way:

Proposition 8 *If $*$ is generated from some $\leq$-faithful tpo over $W^{\pm}$ satisfying ($\preceq$L) then $*$ satisfies:*

($*$L) $\quad$ *If $x \in [\alpha]$ and $y \in [\neg\alpha]$ then $x <^*_\alpha y$.*

Furthermore if $*$ is any *revision operator for $\leq$ which satisfies ($*$L) then the $\leq$-faithful tpo $\preceq_*$ defined right after Theorem 1 satisfies ($\preceq$L).*

From this result we see that $*_{\mathrm{L}}$ is axiomatically characterised by ($*$1)–($*$7) plus ($*$L). However it is easy to see that ($*$L) implies ($*$5)–($*$7). ($*$1) also becomes redundant, since ($*$3), ($*$4) and ($*$L) are enough to force the unique tpo $\leq^*_\alpha$, and we already established after Proposition 2 that ($*$2) can be removed. Hence ($*$3), ($*$4) and ($*$L) form a sound and complete axiomatisation for $*_{\mathrm{L}}$. The sentential counterpart of ($*$L) is the rule *Recalcitrance* of Nayak, Pagnucco, & Peppas (2003), i.e.,

($\circ$L) $\quad$ If $\beta \not\vdash \neg\alpha$ then $\alpha \in \leq^*_\alpha \circ\beta$.

Note also that new evidence is always believed after lexicographic revision. A characterisation of $*_{\mathrm{L}}$ in terms of social choice-like conditions was given by Glaister (1998), who referred to it as "J-revision".

At the other extreme, suppose instead we insist on

($\preceq$P) $\quad x < y$ implies $x^- \prec y^+$.

This rule is equivalent to saying $\lll = <$. (Thus also $\ll = <$.) This property expresses maximal confidence behind the initial tpo $\leq$, or skepticism towards new evidence. Adding this rule to ($\preceq$2)–($\preceq$4) is again enough to specify $\preceq$ completely. It is not difficult to show this time we are left with $x^\delta \preceq y^\epsilon$ iff either $x < y$ or $[x \sim y$ and $(\delta = +$ or $\epsilon = -)]$:

	R_1	R_2	$\cdots$	R_n
+	1	3	$\cdots$	$2n-1$
−	2	4	$\cdots$	$2n$

The associated revision operator $*_{\mathrm{P}}$ is then given by

$$x \leq^*_\alpha y \text{ iff either } x < y \text{ or } (x \sim y \text{ and } x \leq^\alpha y).$$

This is a "reverse" lexicographic method, studied in the context of iterated belief revision (Papini 2001). This time it corresponds to $\leq$ being refined by $\leq^\alpha$. In this case new evidence is not always believed.

Proposition 9 *If $*$ is generated from some $\leq$-faithful tpo over $W^{\pm}$ satisfying ($\preceq$P) then $*$ satisfies*

($*$P) $\quad$ *If $x \in [\neg\alpha]$, $y \in [\alpha]$ and $x < y$ then $x <^*_\alpha y$.*

Furthermore if $*$ is any *revision operator for $\leq$ which satisfies ($*$P) then the $\leq$-faithful tpo $\preceq_*$ defined right after Theorem 1 satisfies ($\preceq$P).*

This result implies $*_{\mathrm{P}}$ may be characterised axiomatically by ($*$1)–($*$7) plus ($*$P). However we may significantly simplify this list by observing that, in the presence of the other rules, ($*$6), ($*$7) and ($*$P) are together equivalent to the single short-and-sweet rule

($*$p) $\quad < \subseteq <^*_\alpha$.

Again ($*$1) becomes redundant, and so we arrive at the following characterisation of $*_{\mathrm{P}}$.

Proposition 10 *$*_{\mathrm{P}}$ is the unique revision operator for $\leq$ which satisfies (*3)–(*5) plus (*p).*

It is easy to see the sentential counterpart of (*p) is the following rule:

(○p) $\quad \leq \circ\beta \ \subseteq \leq^*_\alpha \circ\beta$.

(○p) states that *all* conditional beliefs in $\leq$ are preserved after revision.

As the following example shows (partly based on one by Darwiche & Pearl, 1997), rigid use of either of these limiting cases $*_{\mathrm{L}}$ and $*_{\mathrm{P}}$ can lead to counter-intuitive results.

Example 2 Suppose we have a murder trial with two main suspects, John and Mary. Let p represent "John is the murderer" and q represent "Mary is the murderer". Furthermore let r represent "The victim is an alien from outer-space".

Initially we believe the murder was committed by one person, either John or Mary. However we *wouldn't be surprised* to discover that either both or neither were involved in the crime. What *would* be surprising – indeed highly shocking – would be if we found out the victim was an alien. However we are still capable of imagining a hypothetical situation in which this turns out to be the case, and we think this would not alter our belief that either John or Mary acted alone. If we were to represent all this using a tpo $\leq$, it seems the following is the best candidate:

100	110	101	111
010	000	011	001

Now during the trial we receive testimony that John is the murderer, leading us to revise $\leq$ by p. Supposing we then receive testimony that Mary is the murderer, the most reasonable conclusion would be that both John and Mary were involved in the murder. But using the operator $*_{\mathrm{P}}$ gives

$$\leq_p^{*_{\mathrm{P}}} \circ q = Cn(\neg p \wedge q \wedge \neg r)$$

We are forced to drop our belief that John is the murderer.

Now consider the situation where we receive testimony that John is the murderer, followed by the supposition that if John is the murderer, then the victim is an alien. In this case it seems the reasonable thing to do is drop the acquired belief that John is the murderer. However, using the operator $*_{\mathrm{L}}$ gives

$$\leq_p^{*_{\mathrm{L}}} \circ(p \rightarrow r) = Cn(p \wedge \neg q \wedge r)$$

That is, we end up believing John murdered an alien!

The move to our more general family of tpo-revision operators enables a correct treatment of both these scenarios simultaneously. Consider the $\leq$-faithful tpo $\preceq$ represented by:

	R_1	R_2	R_3	R_4
+	1	2	5	6
−	3	4	7	8

In the first case where we receive evidence pointing towards John's guilt followed by the supposition Mary did it, we have

$$\leq_p^* \circ q = Cn(p \wedge q \wedge \neg r)$$

which is the intuitive result. In the case where we receive evidence for John being the murderer, followed by supposing that if John is the murderer then the victim is an alien, we have

$$\leq_p^* \circ(p \rightarrow r) = Cn(\neg p \wedge q \wedge \neg r)$$

which is what we would expect.

A further sub-class

Close inspection reveals that both the limiting cases mentioned above share something in common – in both cases we have $\lll = \ll$. Writing out this condition in full, the unique $\preceq$ defined in each case satisfies:

($\preceq$5) $\quad x^- \preceq y^+$ iff $x^- \prec y^+$.

This condition states that no x^- appears in the same $\preceq$-rank as a y^+. In this section we take a look at the subclass of our family of revision operators defined by enforcing this condition.

Firstly, in terms of the graphical representation of $\preceq$ the effect of ($\preceq$5) is simple: it just means that no number is allowed to appear *twice* in the array of numbers. Another thing to notice is that if $\ll = \lll$ then the distinction between the overrules relation and the *strictly* overrules relation relative to $*_{\preceq}$ disappears – they collapse into the same binary relation. As for an axiomatic characterisation of this subfamily, the next result points the way:

Proposition 11 *If $*$ is generated from some $\leq$-faithful tpo over $W^{\pm}$ satisfying ($\preceq$5) then $*$ satisfies*

*(*8) For $x \in [\alpha]$ and $y \in [\neg\alpha]$, either $x <^*_\alpha y$ or $y <^*_\alpha x$.*

Furthermore if $*$ is any *revision operator for $\leq$ which satisfies (*8) then the $\leq$-faithful tpo $\preceq_*$ defined right after Theorem 1 satisfies ($\preceq$5).*

Condition (*8) means that after revising by α, there is a separation between α-worlds and $\neg\alpha$-worlds, in the sense that each $\leq^*_\alpha$-rank contains *either* only α-worlds or only $\neg\alpha$-worlds. This property is called (UR) by Booth & Meyer (2006), where it is shown that its sentential counterpart is:

(○8) $\quad$ If $\neg\alpha \notin \leq^*_\alpha \circ\beta$ then $\alpha \in \leq^*_\alpha \circ\beta$.

The postulate (○8) does have a certain amount of intuitive appeal. It says that after receiving α as evidence and then making the supposition β, α should be believed as long as it is consistent to do so.

(*8), alias (○8), is quite a strong rule, and adding it to the list (*1)–(*7) causes some redundancies. Since (*8) implies the equivalence of $x \leq^*_\alpha y$ with $x <^*_\alpha y$ for $x \not\sim^\alpha y$, we see (*6) now follows from (*7). Meanwhile (*5) becomes equivalent to "if $x <^\alpha y$ and $x \leq y$ then $x \leq^*_\alpha y$" (i.e., (CR4) proposed by Darwiche & Pearl, 1997). But using the fact that $\leq = \leq_\top$ (which follows from (*3)), this is seen as just the instance of (*7) in which $\gamma = \top$. Hence (*5) also

disappears. Thus the class of tpo-revision operators generated by those $\leq$-faithful tpos over $W^{\pm}$ satisfying ($\preceq 5$) may be characterised as follows:

Theorem 2 *Let $*$ be a revision operator for $\leq$. Then $*$ is generated from some $\leq$-faithful tpo over $W^{\pm}$ satisfying ($\preceq 5$) iff $*$ satisfies ($*1$), ($*3$), ($*4$), ($*7$) and ($*8$).*

Of course we can if we wish replace the last four rules above with their sentential equivalents.

Conclusion and future work

We have introduced a new family of operators for revising total preorders by sentences based on the simple intuitive idea that when we compare possibilities, we are often able to imagine these possibilities with regard to "best case" and "worst case" scenarios. We have placed this family firmly in the context of the problem of iterated belief revision, and have shown that our results significantly extend current work on this topic.

On the level of belief sets, our operators fall within the realm of non-prioritised revision, in that revision inputs are not necessarily elements of the belief set associated to the revised preorder. This is in contrast to most works on iterated belief change, which are usually given in the "prioritised" setting (with the work of Booth, 2005 being an exception). We envisage prioritised revision by α as a two-stage process, with the first stage being carried out by one of the operators in this paper, and then the second stage consisting of an application of Boutilier's *natural revision* (1996) of the resulting tpo by α, i.e., the most preferred α-worlds are simply brought if necessary to the front of the new tpo. For the special case of the operator $*_{\mathrm{P}}$, this was already done by Booth & Meyer (2006), leading to the *restrained revision* operator (see Section 5 of that paper). For future work we plan to apply this to the more general family.

Another direction for future research is the investigation of larger families of revision operators, such as those obtained by weakening one, or both, of ($\preceq 2$) and ($\preceq 3$). Observe that this is equivalent to weakening ($*3$) and ($*4$), or ($\circ 3$) and ($\circ 4$). The weakening of ($\circ 4$) will be of particular interest, since it is essentially equivalent to the much-criticised postulate (C2) proposed by Darwiche & Pearl (1997).

Conversely, it would be interesting to consider special subclasses of our general family. We considered one in the last section. Another example could be the family obtained by taking $\ll$ or $\lll$ to be modular orderings. Finally note that our operators do not conform to the principle of *categorical matching* – from an initial tpo $\leq$ together with a $\leq$-faithful tpo $\preceq$ over $W^{\pm}$ they return a new tpo $\leq^*_\alpha$, but give no help on defining a new $\leq^*_\alpha$-faithful tpo over $W^{\pm}$ which can then be used to further revise $\leq^*_\alpha$. One way of rectifying this might be to preserve as much of $\lll$ and $\ll$ as possible.

Acknowledgements

Most of the paper was written during a visit by Richard Booth to the KRR group at NICTA in Sydney. Some valuable feedback was received from audience members during a presentation of this material to the KR group at Leipzig University. Thanks are also due to the reviewers for some helpful suggestions. National ICT Australia is funded by the Australia Government's Department of Communications, Information and Technology and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Centre of Excellence program. It is supported by its members the Australian National University, University of NSW, ACT Government, NSW Government and affiliate partner University of Sydney.

References

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

Arrow, K. 1963. *Social Choice and Individual Values*. John Wiley.

Benferhat, S.; Konieczny, S.; Papini, O.; and Pino Pérez, R. 2000. Iterated revision by epistemic states: axioms, semantics and syntax. In *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000)*, 13–17.

Booth, R., and Meyer, T. 2006. Admissible and restrained revision. *Journal of Artificial Intelligence Research*, accepted for publication.

Booth, R. 2005. On the logic of iterated non-prioritised revision. In *Conditionals, Information and Inference*, volume 3301 of *LNAI*. Springer. 86–107.

Boutilier, C. 1996. Iterated revision and minimal changes of conditional beliefs. *Journal of Philosophical Logic* 25(3):263–305.

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

Freund, M. 1991. Injective models and disjunctive relations. *Journal of Logic and Computation* 52:191–203.

Glaister, S. M. 1998. Symmetry and belief revision. *Erkenntnis* 49:21–56.

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

Hansson, S. O.; Fermé, E.; Cantwell, J.; and Falappa, M. 2001. Credibility-limited revision. *Journal of Symbolic Logic* 66(4):1581–1596.

Hansson, S. O. 1999. A survey of non-prioritized belief revision. *Erkenntnis* 50:413–427.

Jin, Y., and Thielscher, M. 2005. Iterated belief revision, revised. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05)*, 478–483.

Katsuno, H., and Mendelzon, A. 1991. Propositional knowledge base revision and minimal change. *Artificial Intelligence* 52:263–294.

Kraus, S.; Lehmann, D.; and Magidor, M. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167–207.

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

Nayak, A.; Pagnucco, M.; and Peppas, P. 2003. Dynamic belief change operators. *Artificial Intelligence* 146:193–228.

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

Papini, O. 2001. Iterated revision operations stemming from the history of an agent's observations. In *Frontiers of Belief Revision*. Kluwer, Dordrecht. 281–303.

Rott, H. 2001. *Change, Choice and Inference*. Oxford University Press.

Spohn, W. 1988. Ordinal conditional functions: A dynamic theory of epistemic states. In *Causation in Decision: Belief, Change and Statistics*, 105–134. Kluwer Academic Publishers.