Economics and Philosophy, **16** (2000), 287–313 Copyright © Cambridge University Press

ON WHY THE BEST SHOULD ALWAYS MEET[1]

Tor Sandqvist
Uppsala University

1. INTRODUCTION

It seems plausible, even truistic, that when an agent is faced with the choice of giving up one belief or another, the decision should be based on the relative strengths of these beliefs along some dimension of doxastic merit. This said, however, two non-trivial questions arise:

(1) *Which* dimension?
(2) *How* should the contraction outcome be affected by the distribution of beliefs along this dimension?

Although it might prima facie be expected that the answer to the second question should depend to some extent upon that to the first, most researchers in belief revision have concentrated almost exclusively on (2). The assumption seems to have been that, when dealing with the technicalities of various preference relations among beliefs, one may

[1] This is a considerably extended, and – I hope – improved version of a paper originally published in Hansson & Rabinowicz (eds.) 1995, pp. 125–35, as 'Why should the best always meet? On the intuitive basis of some contraction operations'. Readers who found that paper interesting will have reason to study the present one, but I do not believe that the converse statement holds.

Many people have contributed to sorting out my thoughts on these issues; in particular, I wish to thank John Cantwell, Sven Ove Hansson, Hans Rott, and an anonymous referee whose careful and no doubt laborious work alerted me to many serious shortcomings of an earlier version of this paper. Financial support during work on the paper has been provided by the research project *Kunskap, kunskapsförändring och icke-monoton inferens*, funded by the Swedish Council for Research in the Humanities and Social Sciences (HSFR).

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

safely abstract from the question of exactly what kinds of epistemic value these formal relations are supposed to reflect.

The present essay aims to cast some doubt upon this assumption. In Sections 2 through 4, I argue, of a certain family of contraction operations that, when described in purely formal terms, they lack intuitive plausibility. Then, in Sections 5 through 8, I try to remedy this situation for *one* of the members of the family – AGM's relational partial meet contraction – by giving a suitable informal reading of the preference relation figuring in it. Section 9 discusses some drawbacks of the suggested solution, thereby motivating the outlooks in the concluding sections, 10 and 11, to provide possible alternative approaches. The reader should be warned that the results will be mainly negative.

2. THE QUESTION

I shall consider, as a group, those contraction operations that follow a general strategy, which I shall call *meet of the best* (MOTB). This strategy, which is familiar to all students of belief revision, may be represented as involving three steps.

Step 1. Given an original set of beliefs S (upon which the requirement of logical closure is imposed in some models, in others that of disjunctive closure, and in still others no such conditions at all), and a sentence a which is to be retracted, obtain the set $S\perp a$, which has as its elements those subsets of S which satisfy some particular structural criterion. The most common criterion is that of being inclusion-maximal among those subsets of S which fail to imply a (under some consequence operator Cn), but other variants can be found in the literature as well. (Thus, the symbol '$\perp$' is here being used in a somewhat wider sense than is customary.)

Step 2. Among the elements of $S\perp a$, pick out those that are maximal with respect to a certain preference relation $\preceq$ over $\mathfrak{P}(S)$. The set of these elements, $\{A\colon A\in S\perp a \;\&\; (\forall B\in S\perp a\; B\preceq A)\}$, will be denoted '$max_{\preceq}(S\perp a)$'.

Step 3. Finally, take the intersection of the elements of $max_{\preceq}(S\perp a)$ as the agent's new set of beliefs, i. e. put $S\dot{-}a = \cap max_{\preceq}(S\perp a)$.

The AGM trio (Alchourrón, Gärdenfors and Makinson, 1985) having set the example with relational partial meet contraction, various authors followed the general scheme outlined above in constructing contraction operators. (Examples can be found, e.g., in Levi (1991, Chapter 4), Rott (1992, Section 1.5), Hansson (1994, Section 3), and Olsson (1997, Section 6).) By varying the original construction in different ways, for example, by placing different formal requirements on S and $\perp$, operators satisfying different rationality postulates were obtained. It is not my aim to go through the relative merits and defects of all these variants. Rather, what I wish to concentrate on is that which they have in common: the idea that

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

the outcome of a contraction operation should consist of whatever is common to a number of 'best' candidate sets of beliefs. The central question is: *what is the intuitive basis for this idea? Why should the best always meet?*

Before starting to discuss possible answers, I shall offer some considerations as to the point of posing this question in the first place. I imagine that some readers will be inclined to respond to it in something like the following manner:

> All that matters, given a certain operator of belief change, is how it behaves in practice; to what extent it *works*. This is to be assessed by representation theorems and by application to concrete examples. Any asking for further, 'intuitive' justification is pointless. The internal, technical workings of the operator are just a way of *representing* it and should not be ascribed any further, intrinsic significance.

There are two reasons why I do not agree with this point of view. First, it seems to me that a philosophically satisfying account of rational belief change should not only answer the question *what* constitutes such belief change, but also, in some sense, *why* it is rational. In the case of MOTB, one would not only want a proof that it leads to rational behaviour (as defined, e.g., by the Gärdenfors postulates), but also some feeling that this behaviour arises in a 'natural' way from the mechanics of the operator. If the construction seems in any way weird or arbitrary, this is, in my opinion, a matter of concern even if the internal oddities never reach the surface of cognitive behaviour. Rational behaviour, while indeed a *sine qua non* of models of belief change, is not the only *desideratum* thereof.

Secondly, there is an indirect value in having an operator the mechanics of which conform to our intuitions – a value which should appeal even to some of those who would dismiss the sentiments in the preceding paragraph as muddled and unscientific, and whose primary concern is with overt cognitive behaviour. For even if some operator has been found to satisfy certain rationality postulates, and to behave satisfactorily in application, one may still have the sneaking suspicion that there are counterexamples to be found, which have not yet been discovered. (By 'counterexamples' I do not, of course, mean instances falsifying postulates that have been proved to hold, but rather examples which strike one as irrational in spite of being in accordance with such postulates.) Presumably, there is less ground for such worries if not only the 'observed behaviour' but also the 'inner workings' of the operator accord with one's intuitions.

These considerations should not be understood as an argument that representation theorems and the axiomatic method are useless or

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

misdirected – my point is just that they do not constitute the *only* meaningful way of discussing belief dynamics.

Before we move on, a few words on terminology and notation are needed. '$A \prec B$' is always to be read as shorthand for '$A \preceq B$ and not $B \preceq A$', and similarly for the symbols '$<$' and '$\leq$', to be introduced later. When dealing specifically with the AGM version of MOTB, the expression '$K \perp a$', stands, as it did originally, for the set of inclusion-maximal subsets of K which fail to imply a, but when talking about MOTB constructions in general, the expression should be read as 'the set of subsets of K satisfying whatever structural criterion the construction in question makes use of'. When discussing models where the set of the agent's beliefs is required to be logically closed, we shall denote this set by 'K', but when no such assumption is being made, we shall use 'S'. In speaking of the *subsets* of a set of beliefs, we will mean different things according to context. When discussing models which do not require logical closure of belief sets, we will mean *all* subsets; when discussing models which do make this requirement, we will only be referring to the logically closed subsets. Symbolically, the set of these latter will be denoted '$\mathfrak{P}_{\mathrm{Cn}}(K)$'. By a *tautology* I shall mean any element of $\mathrm{Cn}(\emptyset)$, whether truth-functionally tautologous or not. When a contraction function $\dot{-}$ and a relation $\preceq$ are such that using $\preceq$ for contraction purposes according to the MOTB strategy gives rise precisely to $\dot{-}$, then we will say that $\preceq$ *generates* $\dot{-}$, or equivalently, that $\dot{-}$ is *based on* $\preceq$. Finally, when saying of a formal language that it is *finite*, we mean that it contains a finite number of atomic formulae.

3. FIRST RESPONSE

Now, then: why should the best always meet? I shall start by offering a first answer which springs quite naturally to mind. No claims to originality are made here; on the contrary, I try to formulate a position which, judging from the general gist of many discussions, I take to be fairly widespread among researchers in belief revision. The answer is as follows.

In a given contraction situation, a compromise must be attained between two conflicting goals. On the one hand, we want to sacrifice as little information as possible, as measured by some relation of informational value. On the other, we wish to minimize the risk of holding false beliefs. (For one discussion, in a somewhat wider context, of the relation between these two *desiderata* of inquiry, see Levi (1991, pp. 10 f.)) Steps 1–2 and 3 in the MOTB strategy correspond to each of the *desiderata*, respectively. In steps 1–2, we single out the most valuable candidate belief sets, hereby complying with the goal of minimal loss of informational value. Step 3 serves as a means of doxastic prudence, or avoidance

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

of arbitrariness: if we have no valid reason to prefer one particular belief/theory to another, we should suspend judgement of both of them, only keeping whatever they both agree on.

We can summarize the argument in the following little dialogue.

A: 'Why should the *best* always meet?'
B: 'In order to maximize informational value'.
A: 'And why should the best always *meet*?'
B: 'In order to minimize the holding of false beliefs'.

This piece of reasoning certainly has some intuitive appeal. However, if one tries to expound it in greater detail, problems arise. Consider B's first reply. It seems a reasonable assumption that what is intended is that 'informational value' is to be 'maximized' in the *final outcome* of the contraction. Surely, one would argue, the fact that the sets singled out in step 2 are, in some sense, the 'best', is of no merit *in itself*, but only insofar as it confers informational value on that which we finally end up with, namely $S\dot{-}a$. But if this is accepted, then B's argument is valid only if some version of the following principle is assumed to hold:

(P) In any given contraction situation, for any collection of candidate belief sets, the value of the intersection of these sets is somehow positively correlated with the values of the individual sets themselves.

(P) is necessary because, without it, the following question stands unanswered: given (*a*) that we want to maximize the informational value of $S\dot{-}a$, and (*b*) that $S\dot{-}a$ is to be the intersection of some collection of subsets of S, why select, as input to this operation of intersection, exactly those subsets of S that are, *in themselves*, maximal with respect to informational value? And this, of course, is just another way of asking what relevance B's first reply has to the question it is supposed to answer.

4. THE PARTY EXAMPLE

How, then, does (P) fare when subjected to critical examination? As might be expected, this depends on the interpretation of crucial expressions figuring in it, like 'value' and 'somehow positively correlated'. One formally straightforward reading, which renders (P) definitely *false*, is the following.

(P1) Let $\preceq$ be any preference relation over $\mathfrak{P}(S)$, let $a \in S$, and let $\mathfrak{A}$ and $\mathfrak{B}$ be subsets of $S \perp a$ such that for each $A \in \mathfrak{A}$ and each $B \in \mathfrak{B}$, $A \preceq B$. Then $\cap\mathfrak{A} \preceq \cap\mathfrak{B}$.

It is not difficult to invent simple formal counterexamples to (P1). However, since my ambitions in this paper are conceptual rather than formal, I shall instead present a somewhat complex example, intended

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

not only to serve as a formal refutation of (P1), but also to help the reader in acquiring an intuitive feeling for the problems at hand. Formally, the example will be carried out in the AGM framework. (Due to space limitations, a number of technical claims must be made in this section without proof. Readers demanding such proofs are invited to contact the author personally.)

The rationale behind the example is as follows: it seems reasonable to suppose that, whatever the origins of epistemic value, the value of a set of beliefs bears some correlation with the values of its constituent beliefs. But if this is so, then there may very well be two or more sets of beliefs that are each very valuable but such that their intersection is practically worthless – namely, if whatever makes each of them so valuable fails to be that which they all have in common.

Here is the story. A fancy party for wealthy and famous people has recently taken place, and, an avid consumer of gossip, I am eager to know who was there and who was not. Suppose that my investigations lead me to believe that (a_1) Amanda was there; (a_2) Arthur was there; (b) Bethena was there; (c_1) Charles was there; and (c_2) Cynthia was there. My belief set, then, is $K = \mathrm{Cn}(\{a_1, a_2, b, c_1, c_2\})$.

Some more notation before we continue: let $\mathfrak{L}$ be the set of the sentences a_1, a_2, b, c_1, c_2, and their truth-functional combinations; for any subset A of K, let $[A]$ be the set of maximal consistent subsets of $\mathfrak{L}$ containing A (i.e., informally put, let $[A]$ be the set of possible states of the world with respect to a_1, a_2, b, c_1, and c_2 that are consistent with A); and let $card(A)$ be the number of elements in A. Now suppose my preference relation $\preceq$ over $\mathfrak{P}_{\mathrm{Cn}}(K)$ to be such that for any $A,B \subseteq K$:

1. If $card(A \cap \{a_1, a_2\}) < card(B \cap \{a_1, a_2\})$, then $A \prec B$.
2. Unless rule 1 says otherwise, if $card(A \cap \{b\}) < card(B \cap \{b\})$, then $A \prec B$.
3. Unless rules 1–2 say otherwise, if $card(A \cap \{c_1, c_2\}) < card(B \cap \{c_1, c_2\})$, then $A \prec B$.
4. Unless rules 1–3 say otherwise, if $card([B]) < card([A])$, then $A \prec B$.
5. Unless rules 1–4 say otherwise, $A \prec B$.

In other words, $\prec$ is a lexicographic ordering concerned to maximize, in the first place, the containing of a_1 and a_2, in the second place that of b, in the third place that of c_1 and c_2, and in the fourth place logical strength in the sense of being consistent with the smallest number of possible worlds. Sets equal in all these respects are deemed equally valuable.

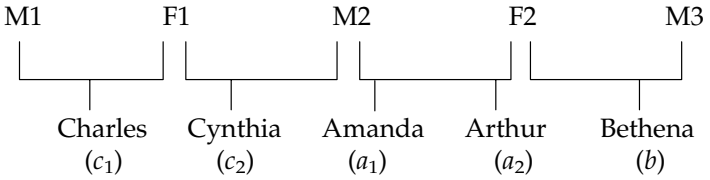
This $\prec$ has several pleasing properties. It follows directly from the characterization that it is connected. Moreover, it is both maximizing[2]

[2] The maximizing property was first introduced by Hansson (1992, p. 418). $\preceq$ is maximizing if, for every A, B: $A \prec B$ if $A \subset B$.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

and transitive. However, in spite of the apparently well-behaved nature of this relation, we shall find that it refuses to comply with (P1) in a concrete contraction situation.

Let us move on with our story. As everyone knows, rich and fashionable people are amazingly promiscuous; consequently, a rather complex set of family relations holds between the above-mentioned people. These relations are best described by means of a family tree. (Since the parents will play no further role in the example, their names are not given here.)



There are in the second generation exactly four pairs of individuals where the one has no parent in common with the other: Amanda and Charles; Arthur and Charles; Bethena and Charles; and Bethena and Cynthia. Regarding groups of three or more, there is no combination that does not involve at least one pair of (half-)siblings.

Now suppose that I learn from an extremely reliable source that at the party there were in fact no two persons who were even half-siblings. This means I have to retract, from my set S of beliefs, the somewhat cumbersome sentence $d = (c_2 \wedge (c_1 \vee a_1 \vee a_2)) \vee (b \wedge (a_1 \vee a_2)) \vee (a_1 \wedge a_2)$. (Strictly speaking, it means I have to *revise* my belief set with $\neg d$. But, assuming the Levi identity: $K^*a = (K \dot{-} \neg a) + a$, this in turn requires that I first contract with respect to d, and it is this part of the process which is currently of interest.) MOTB now demands that we identify the elements of $K \perp d$. None of these contains more than two atomic sentences, but there are four containing exactly two; each corresponding to one of the non-sibling pairs just listed. Let us dub them 'K_1' through 'K_4' as follows.

K_1: a_1 and c_1 K_2: a_2 and c_1
K_3: b and c_1 K_4: b and c_2

Of course, K_1 through K_4 contain many non-atomic sentences as well, but these do not concern us at present. What *is* a matter of concern is the following: according to our characterization of $\preceq$, either one of K_1 and K_2 is judged better than either one of K_3 and K_4; yet the intersection of the former is judged inferior to the intersection of the latter – since, of atomic sentences, $K_1 \cap K_2$ contains only c_1, while $K_3 \cap K_4$ possesses the more valuable-rated b. This contradicts (P1). Moreover, K_1 and K_2 happen to be precisely the $\preceq$-maximal elements of $K \perp d$, meaning that the MOTB strategy will lead me to conclude that Charles (c_1) was definitely at the

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

party, while all the other people are doubtful cases. Thus, MOTB delivers a belief state which, according to my preference ordering, is inferior to what I would have got if, instead of letting the best meet, I had gone for the intersection of, for example, the submaximal K_3 and K_4.

Summing up: not only is (P1) false in general; it is false in some cases where $\preceq$ bears not an altogether implausible structure, and false in such a way as completely to undermine B's first response in our little dialogue above – having the *best* meet does not always maximize informational value.

What to do, then? How is MOTB to be justified? This question applies, and is equally important, regardless of which particular MOTB contraction we are dealing with; whether it is defined for logically closed sets or for belief bases, and whether or not restrictions such as transitivity and connectivity are placed on the preference relations involved. Though in the Party Example the belief set happens to be logically closed, and the relation $\preceq$ transitive and connected, the example will of course constitute just as serious a problem for MOTB models where these properties are not required as it does for the AGM variety. This is not to say, however, that a solution which works satisfactorily with one brand of MOTB is guaranteed to be equally applicable to all the others. In what follows, I shall only concern myself with the relational Partial Meet Contraction (PMC) of AGM, focusing specially on what seems to be commonly regarded as its most interesting special case, namely transitively relational PMC.

5. OUTLINE OF THE SOLUTION

We have seen that there are counterexamples to (P1): not every preference relation $\preceq$ has the property that sets of preferable candidate belief sets have preferable intersections. However, for our purposes we do not really need this strong thesis: it would suffice if we could argue that every *rational* $\preceq$ has the property in question. MOTB is a recipe for constructing a contraction operation out of a given preference relation. Given a good preference relation we want it to provide us with a good contraction operation. But if, given an *irrational* preference ordering, it produces an irrational contraction operation, this should not be a matter of concern any more than if a cooking recipe should fail to produce a delicious dish out of spoiled ingredients.

The key to my suggested solution, then, lies in arguing that all counterexamples to (P1) involve irrational preference relations. (This means having, perhaps surprisingly, to reject the prima-facie sensible preference relation of the Party Example as an irrational one – more on this later.) I will give an account of what it is that makes a belief set valuable, and formulate a number of conditions which must be fulfilled

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

by any preference relation reflecting this kind of value. The conditions will both (*a*) guarantee that sets of preferable candidate belief sets have preferable intersections, and (*b*) have some intuitive appeal irrespectively of this guarantee. Both of these features are essential: (*a*) is what we are after in the first place, but any conditions without the property (*b*) would be *ad hoc*.

For reasons that will become clear, I will call preference relations satisfying my criteria *exclusiveness relations*. We will find that the requirement that preference relations be exclusiveness relations does not impose any restrictions on contraction behaviour: for every relational PMC there is some exclusiveness relation which generates it. Moreover, for every transitively relational PMC there is some transitive exclusiveness relation which generates it.

6. GROVE SETS THE STAGE

The notion of exclusiveness is intuitively based on the concept of, and will be formally defined in terms of, possible worlds. Before we can move on, therefore, it will be useful to recapitulate some well-known facts about how the possible worlds framework is related to the AGM contractions we are trying to justify. In the following, by a 'possible world' we shall always mean a maximal consistent subset of the language; W will be the set of all possible worlds; and, for any set of beliefs A, $[A]$ is the set of all possible worlds that are compatible with, that is, supersets of, A. When a is a formula, '$[a]$' abbreviates '$[\{a\}]$'.

In his 1988 paper, 'Two Modellings for Theory Change', Adam Grove demonstrated that a close connection holds between the relational partial meet contraction of AGM on one hand, and Grove's own possible worlds model on the other. The connection is as follows.

In AGM's MOTB modelling, a belief state is represented by (1) a set K of beliefs, and (2) a preference relation $\preceq$ over the subsets of K. Grove's model also makes use of a preference relation (here labelled '$\leq$'), but over possible worlds. According to Grove, the intended informal interpretation of '$u \leq v$' is u's being at least as 'close' to $[K]$ (p. 159), or at least as 'compatible [...] with our current beliefs' (p. 160) as v is. For the purposes of this paper, I shall say that $\leq$ reflects the agent's judgements of the relative *implausibility* of worlds. The relation is subject to the condition that all and only the worlds in $[K]$ are $\leq$-minimal. That is to say, the most plausible worlds are precisely the ones that are compatible with all of the agent's beliefs. All the others are regarded by the agent as out of the question – only more or less strongly so.

Though Grove's own focus is on belief *revision*, the relation $\leq$ may also be put to use in contraction as follows. When forced to give up some belief a in K, the agent must start to regard at least some of the $\neg a$-worlds

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

as possibly true. Since the agent believes a, the $\neg a$-worlds will all initially be ruled out – but some less so than others. If we let '$min_{\leq}[\neg a]$' denote the set of $\leq$-minimal (most plausible, least strongly ruled out) $\neg a$-worlds, the agent's new contracted belief set is defined to consist of those sentences which hold throughout both $[K]$ and $min_{\leq}[\neg a]$; symbolically, $K\dot{-}a = \cap[K](\cup min_{\leq}[\neg a])$. Note that while the new belief set $K\dot{-}a$ is thus by definition the largest set compatible with *all* the worlds in $[K]\cup min_{\leq}[\neg a]$, it does not follow in general that $K\dot{-}a$ is compatible *only* with the worlds in $[K]\cup min_{\leq}[\neg a]$ – though, if $\mathfrak{L}$ is finite, the latter holds as well. (For a counterexample in the infinite case, see Appendix: Proof of Observation 3.)

Let us now return to the AGM framework, and consider a special subset of $\mathfrak{P}(K)$: namely $\cup\{K\perp p : p\in K \wedge p\notin \mathrm{Cn}(\varnothing)\}$, i.e. the set of those subsets of K which can ever figure in some $K\perp p$, where p is a non-tautologous member of K. This set will henceforth be referred to as '$\mathfrak{R}$'. What is important about $\mathfrak{R}$ is that (as Grove showed) it stands in a one-to-one correspondence with W-$[K]$: for each $A\in\mathfrak{R}$, there is a uniquely determined element of W-$[K]$ – we may call it 'w_A' – such that $[A] = [K]\cup\{w_A\}$. Inversely, for each $u\in W$-$[K]$, there is a uniquely determined element of $\mathfrak{R}$ – namely $K\cap u$, which we may abbreviate as 'R_u' – such that $[R_u] = [K]\cup\{u\}$. Thus, a subset of K is in $\mathfrak{R}$ iff it is compatible with exactly one world outside K. Moreover, it is an element of $K\perp a$ iff it is compatible with exactly one world in $[\neg a]$.

Now suppose we have a relation $\preceq$ over $\mathfrak{P}_{\mathrm{Cn}}(K)$ and a relation $\leq$ over W such that for all $A,B\in\mathfrak{R}$, $A\preceq B$ iff $w_B\leq w_A$ – i.e., B is more valuable than A iff B's characteristic world is more plausible than that of A. We then have, for any A and any a, that $A\in max_{\preceq}(K\perp a)$ iff $w_A\in min_{\leq}[\neg a]$. Since $A = K\cap w_A$, it follows that $\cap max_{\preceq}(K\perp a) = \cap([K]\cup min_{\leq}[\neg a])$, which is precisely to say that AGM contraction always delivers the same result as the sphere-based contraction described two paragraphs back. Put differently: provided that the agent ranks elements of $\mathfrak{R}$ precisely according to the plausibility of their corresponding worlds, AGM's MOTB contraction will deliver the largest subset of K which allows all minimally implausible $\neg a$-worlds.

What is crucial in the present context is that this conclusion holds regardless of how the agent orders, with respect to $\preceq$, the various subsets of K that are *not* elements of $\mathfrak{R}$ – i.e., subsets compatible with more than one world in addition to $[K]$. These are never taken into consideration when using $\preceq$ for purposes of contraction, wherefore these sets may be ordered in a completely arbitrary fashion without changing the behaviour of the contraction operation based on $\preceq$. What this *does* affect, however, is the standing, with respect to $\preceq$, of the contraction *outcome* compared with other subsets of K – since, in the typical case, $K\dot{-}a$ will be the intersection of several elements of $K\perp a$, $[K\dot{-}a]$ thus comprising more

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

than one world outside $[K]$. And, of course, the status of $K\dot{-}a$ with regard to $\preceq$ is precisely what decides the possibility of finding disturbing situations such as the one in Example 1. This fact will come in handy when, in the next section, we place restrictions on $\preceq$ so as to validate a sufficiently strong version of (P) without thereby altering the range of permissible contraction behaviours.

7. EXCLUSIVENESS

We are now in a position to introduce the notions of *exclusiveness* and *exclusiveness relation*. As it turns out, these notions will only help us in the special case where the language is finite, but in defining them we will nevertheless take a more general stance, keeping the infinite case in mind as well.

The idea is that a belief set is valuable insofar as it excludes, (rules out, is logically incompatible with) possible worlds. Two things are taken into account when determining the exclusiveness of a belief set: *which* worlds are excluded, and *how many* worlds are excluded. The first of these considerations requires some discrimination among possible worlds. So, in the spirit of Grove, our agent considers worlds more or less implausible and, therefore, more or less important to rule out. It is taken to be a prime concern of the agent to guarantee that the implausibility of the non-excluded worlds does not exceed a certain value, and the stronger such a guarantee a certain belief set gives, the greater its exclusiveness. Formally, taking '$A \preceq_e B$' to mean that the set of beliefs B is at least as exclusive as A, we have as a 'First Principle of Exclusiveness':

(E1) If there is a $u \in [A]$ such that for every $v \in [B]$, $v < u$, then $A \prec_e B$.

Our second principle concerns only elements of $\mathfrak{R}$. As we explained, any two such elements differ with respect to exactly one possible world; each will be compatible with all elements of $[K]$ plus one more world. This being the only difference between the $\mathfrak{R}$-elements, it seems a natural assumption that the relation between these respective worlds should determine the relation between the $\mathfrak{R}$-elements in question (just as in the case we discussed in Section 6). If the one world is more implausible than the other, the corresponding $\mathfrak{R}$-element will be less exclusive than the other, and if the worlds are equal (or perhaps incommensurable), then the $\mathfrak{R}$-elements will be equal (incommensurable). In short:

(E2) If $A, B \in \mathfrak{R}$, then $A \preceq_e B$ iff $w_B \leq w_A$.

(E1) and (E2) are conditions on combinations of implausibility orderings $\leq$ and relations of relative exclusiveness $\preceq_e$. They allow us, given an

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

agent's $\leq$, to say whether a given $\preceq$ could possibly reflect the exclusiveness judgements of that agent. But what we really want is a set of conditions applicable to agents as described in the PMC framework. Since that framework makes no mention of any relation $\leq$ on the set of possible worlds, (E1) and (E2) are of no use as they stand. What we can and will do, though, is to import these requirements into an existentially quantified 'master condition': $\preceq$ is an exclusiveness relation only if there is some relation $\leq$ on the set of possible worlds such that $\preceq$ and $\leq$ jointly satisfy (E1) and (E2) (plus other conditions to be specified). Thus conceived, the property of being an exclusiveness relation becomes a purely structural one, having nothing to do with whether $\preceq$ in any way reflects some particular (actual or hypothetical) agent's cognitive state. So strictly speaking, rather than 'exclusiveness relation', we should write something like 'relation whose structural properties do not rule out its being some agent's exclusiveness relation' – but no confusion should result from using the former, more convenient term.

Besides the qualitatively oriented (E1) and (E2), we now add a quantitative requirement: since larger sets rule out more worlds, an exclusiveness relation should be maximizing.

(E3) If $A \subset B$, then $A \prec_e B$.

Furthermore, while requiring connectivity of $\preceq_e$ would perhaps mean an unnecessary restriction of generality (we may want to allow for sets of beliefs to be incommensurable in this respect) it is hard to see how relative exclusiveness could fail to be reflexive. This condition, then, should be added explicitly.

(E4) $A \preceq_e A$.

It remains to consider whether, in addition to the joint conditions on $\preceq$ and $\leq$ expressed in (E1) and (E2), and the conditions (E3) and (E4) on $\preceq$ *simpliciter*, we should place any rationality constraints on $\leq$ in itself. Let us look to Grove, our main source of inspiration, and decide how much to import into the present context of the conditions he places (p. 160) on $\leq$. There are four conditions to consider (we adapt Grove's formulations to our own notation):

'(S$\leq$1): $\leq$ is connected ($\forall u,v \in W$: $u \leq v$ or $v \leq u$)'. By the same reasoning as for $\preceq_e$ itself, this requirement is too strong and should be weakened to that of reflexivity:

(E5) $w \leq w$.

'(S$\leq$2): $\leq$ is transitive'. This second condition of Grove's is often considered intuitively reasonable, but belief revision modellings without

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

transitivity have been studied before,[3] and if the concept of exclusiveness can be brought to bear on such modellings as well, then so much the better. This is not to say that there can be no interest in studying specifically exclusiveness relations based on transitive $\leq$-relations; the contention is just that there is no reason to include transitivity in the concept of exclusiveness *per se*. This condition, then, should be dropped.

'(S$\leq$3): If $[a]\neq\emptyset$, then $\{u\in[a]: u\leq v\ \forall v\in[a]\} \neq \emptyset$'. This condition is commonly known as the 'limit assumption', and is often regarded as a necessary evil to ensure that a given contraction strategy can always be applied in a meaningful way. But such a guarantee is no concern of ours, our goal being rather to establish, *given* that a particular contraction operation can possibly be applied, that it can also be so *rationally*. So, again, while the condition may well be a desirable one, we should not include it in the notion of exclusiveness. Thus, we drop (S$\leq$3).

'(S$\leq$4): w in W is $\leq$-minimal ($w\leq v\ \forall v$) if and only if $w\in[K]$'. This is part of what we mean when we say, albeit vaguely, that $\leq$ reflects increasing 'implausibility', and must therefore be retained. Indeed, in the absence of transitivity, we must add to it a condition that would otherwise follow from it: that elements of K are always considered strictly less implausible than elements outside K.

(E6) If $w\in[K]$ then $w\leq v$, and
if $w\in[K]$ and $v\notin[K]$ then $w<v$.

We can now regiment our terminology as follows:

Definition: Let K be a consistent belief set in the language $\mathfrak{L}$, let $\mathfrak{P}_{\mathrm{Cn}}(K)$ be the set of logically closed subsets of K, W the set of maximal consistent subsets of $\mathfrak{L}$, and $\preceq$ a relation on $\mathfrak{P}_{\mathrm{Cn}}(K)$. Then $\preceq$ is an *exclusiveness relation for K in $\mathfrak{L}$* if there is some relation $\leq$ on W such that (E1)–(E6) hold for all $A,B\in\mathfrak{P}_{\mathrm{Cn}}(K)$ and all $v,w\in W$.

The following incarnation of (P) can now be seen to hold:

(PE) Let $\preceq_e$ be an exclusiveness relation for K in a finite language $\mathfrak{L}$, let $a\in K$, let $\mathfrak{A}$ and $\mathfrak{B}$ be subsets of $K\perp a$, and suppose that there is an $A\in\mathfrak{A}$ such that for every $B\in\mathfrak{B}$, $A\prec_e B$. Then $\cap\mathfrak{A}\prec_e\cap\mathfrak{B}$.

Proof: since all the elements of $\mathfrak{A}$ and $\mathfrak{B}$ are elements of $K\perp a$, and hence of $\mathfrak{K}$, it follows from (E2) that for any $B\in\mathfrak{B}$, $w_B<w_A$ – and, therefore, by (E6), $u<w_A$ for any $u\in[B]$. So for every $u\in\cup\{[B]:B\in\mathfrak{B}\}$, $u<w_A$. But, when $\mathfrak{L}$ is finite, $\cup\{[B]:B\in\mathfrak{B}\} = [\cap\mathfrak{B}]$, so we may also write: for every $u\in[\cap\mathfrak{B}]$,

[3] For instance in the original AGM paper, which investigates as separate questions (1) the effect of requiring contraction to be based on some relation over $\mathfrak{P}(K)$, and (2) the additional effect of requiring this relation to be transitive. (Alchourrón *et al.*, 1985, §4.)

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

$u<w_A$. Furthermore, since $w_A\in[A]$ and $A\in\mathfrak{U}$, we have $w_A\in[\cap\mathfrak{U}]$. By (E1), then, $\cap\mathfrak{U}\prec_e\cap\mathfrak{B}$, Q.E.D.

(PE) enables us, as desired, to answer the question, 'Why should the *best* always meet?' with: 'In order to maximize informational value'. Or, more fully spelled out: 'For a rational agent, the best elements of $K\perp a$ will be precisely the most exclusive ones. When the exclusiveness relation is connected and transitive in $\mathfrak{R}$, (PE) implies that the meet of a subset of $K\perp a$ containing non-maximally exclusive elements will always be less exclusive, that is, of less informational value, than the meet of the maximally exclusive ones. Thus, having the best meet will maximize informational value'.

Admittedly, this argument is limited to the case where the $\mathfrak{R}$-part of $\preceq_e$ is connected and transitive but, since these are commonly regarded as a reasonable rationality criteria, we do seem to have made some progress. However, a critic may object that, while the above argument explains why, *on the assumption* that the outcome of the contraction operation is to be the intersection of a subset of $K\perp a$, that subset should consist of $\preceq_e$-maximal elements, it does not justify that assumption in the first place.

By way of an answer, we can formulate an intuitively reasonable maxim which seems, in the case where $\mathfrak{L}$ is finite and $\leq$ is connected and transitive, to pick out relational PMC uniquely as the optimal contraction operation.

> **Maxim for belief contraction:** In contraction, always maximize exclusiveness as far as is possible without arbitrary discrimination between possible worlds.

The argument leading from this maxim to relational PMC goes as follows. In order to retract a from K, we must incorporate into $[K]$ at least one element of $[\neg a]$. In order to maximize exclusiveness, we should, according to (E1) and given connectivity and transitivity of $\leq$, only incorporate $\leq$-minimal $\neg a$-worlds. We may either incorporate all such worlds, or only some of them. While the latter alternative is preferable with regard to (E3), adopting it would violate our maxim in that any incorporation of some maximally plausible worlds at the expense of others would have to be arbitrary. We must therefore incorporate all minimally implausible $\neg a$-worlds. Principle (E3) is not redundant, however: without it, there would be nothing to prevent us, once we have accepted some $\neg a$-worlds, from incorporating any other world which is no more implausible than those already admitted, and we might end up with the set of all worlds, $\neg a$-containing or not, that are no worse than the best $\neg a$-worlds.[4] (E3) implies that a belief set allowing precisely [K]

[4] The resulting contraction function is known as 'severe withdrawal' (Rott & Pagnucco,

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

and all minimally implausible $\neg a$-worlds is preferable to any set allowing other worlds as well. Here, the non-arbitrariness requirement does not present a problem: sufficient ground for discrimination between the worlds outside $[\neg a]$ and those inside it is provided by the fact that only the latter are affected by our obligation to include at least one $[\neg a]$-world.

What our maxim prescribes then, is, if possible, to go to a contracted set of beliefs allowing, in addition to $[K]$, all and only minimally implausible $\neg a$-worlds. And when $\mathfrak{L}$ is finite, this *is* possible: as explained in Section 6, the sought-for set is precisely the one delivered by relational PMC.

8. EXCLUSIVENESS AND THE PARTY EXAMPLE

How does the requirement that $\preceq$ be an exclusiveness relation take care of our Party Example – the one that started all the trouble? Simple: the relation $\preceq$ characterized by rules 1–5 cannot be one of exclusiveness. Why? Suppose for *reductio* that it is, and take some $B \in \mathfrak{R}$ such that $a_1 \notin B$, $a_2 \notin B$, and $b \in B$. Now, for any $u \in [a_1]$–$[K]$, there is an element A of $\mathfrak{R}$ such that $a_1 \in A$ and $u = w_A$. By rule 1, $B \prec A$, so, by (E2), $u < w_B$. By analogous reasoning, $u < w_B$ for any $u \in [a_2]$–$[K]$; and by (E6), $u < w_B$ for any $u \in [K]$. It follows that w_B is strictly more implausible than any world in $[a_1]$ or in $[a_2]$. Now regard $C = \mathrm{Cn}(\{a_1 \vee a_2\})$. Since every world in C is either in $[a_1]$ or in $[a_2]$, (E1) gives us $B \prec C$. But since $b \in B$ while C is devoid of atomic sentences, rule 1 would have it that, on the contrary, $C \prec B$.

Informally and more generally put: if an exclusiveness relation places supreme weight on a set's containing certain special sentences (as, in our example, it does on a_1 and a_2), then it must not, as in the Party Example, be completely indifferent to a set's containing disjunctions of such sentences. This is a consequence of the basic intuition of exclusiveness: for if the containing of either one of two sentences guarantees the exclusion of certain particularly unpalatable worlds, then so does the containing of the disjunction of those sentences. Thus, we have arrived at a kind of solution to our problem. If maximizing exclusiveness is taken to be the main concern of cognitive agents, relational PMC can be justified – and transitively relational PMC more convincingly so – in such cases where the marking off relation satisfies our conditions for exclusiveness. Other cases, such as the one in Example 1, can be rejected as involving irrational preference orderings.

It must be realized that, while this is indeed in one sense a restriction on the original AGM model, there is also a sense in which AGM escapes

1999, p. 509) or 'Rott contraction' (Cantwell, 1999, p. 56). I do not wish to suggest that Rott contraction is necessarily a bad thing – but what we are discussing and are trying to justify is AGM contraction.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

entirely unscathed. For although the proposal rules out certain *preference relations*, it does not affect the set of *contraction operations* construable as relational PMC's. This is because, given an arbitrary contraction-defining relation $\preceq$ on the subsets of K, it is always possible to find an exclusiveness relation which generates the same contraction function as does $\preceq$. This follows from the fact that it is only the $\Re$-part of $\preceq$ that is relevant to contraction behaviour – and given this fraction, the remainder of the relation can always be adjusted to obtain an exclusiveness relation.[5] Thus, we have the following:

> **Observation 1:** If $\dot{-}$ is an operation of relational partial meet contraction, then there is an exclusiveness relation $\preceq_e$ such that $\dot{-}$ is the contraction operation based on $\preceq_e$.
> (See Appendix for proof sketch.)

This result concerns relational PMC in general. What about transitively relational PMC? Of course, the above general result implies as a special case that every transitively relational PMC operation is generated by *some* exclusiveness relation. But if our goal is to justify transitively relational PMC, as it were, *qua* transitive, we would like the exclusiveness relation in question to be transitive, too. Fortunately, it turns out that the exclusiveness relation generated for Observation 1 can always, in cases of transitively relational PMC, be straightforwardly extended to a transitive relation without losing the characteristics of an exclusiveness relation.

> **Observation 2:** If $\dot{-}$ is an operation of transitively relational partial meet contraction, then there is a transitive exclusiveness relation $\preceq_e$ such that $\dot{-}$ is the contraction operation based on $\preceq_e$.
> (See Appendix for proof sketch.)

What is wrong with the Party Example, then, is not the agent's exhibited doxastic *behaviour* – it is quite all right to believe initially that all five persons were at the party, while, after contraction by d, only being sure about Charles. Indeed, it may be quite sensible to have a preference relation $\preceq$ such that its $\Re$-part is correctly described by rules 1–5: namely, if worlds containing more a-sentences are considered more plausible than worlds containing fewer a-sentences; if, among worlds equal in this respect, b-worlds are more plausible than non-b-worlds; etc. For such an agent, K_1 and K_2 will indeed be the $\preceq$-maximal elements of

[5] This is a bit of an over-simplification. Since we have no guarantee that the $\Re$-part of $\preceq$ will be reflexive, whereas this is required of the exclusiveness relation, we may have to make some adjustments to the $\Re$-part of $\preceq$ as well. However, as is shown in the proof, this can be done without altering the behaviour of the contraction operation.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

$K \perp d$. But – and this is the crucial difference – such an agent will also value $K \dot{-} d = K_1 \cap K_2$ highly, since this set contains $a_1 \vee a_2$, hereby excluding all those implausible worlds where neither a_1 nor a_2 is true.

9. DRAWBACKS

In the preceding two sections, I have focused on the strengths of the notion of exclusiveness; it is time to turn to its weaknesses. I shall present a number of serious objections to it, along with outlines of possible responses to these objections. I do not, personally, have a definite opinion about the extent to which my rejoinders actually manage to fend off the objections. The reader must decide for him or herself.

First objection. Not all readers will happily reject the preferences described in the Party Example as irrational. Why, they will ask, should we not be allowed, if we like, to value a_1 and a_2 highly while caring next to nothing for their disjunction? Of course, it has just been demonstrated that this judgement is incompatible with the principles of exclusiveness, but does that not show that those principles are too restrictive?

One possible response to this objection is to try and shift the burden of proof. After all, rules 1–5 are merely descriptive of $\preceq$: they say nothing of its intuitive rationale. Why do I regard a_1 and a_2 as supremely valuable? Because they are especially likely to be true? But in that case, surely, their disjunction is also likely to be true. Or is it just that I would really *like* to believe in a_1 and/or a_2, whereas their disjunction holds no special attraction for me? In that case, arguably, basing my deliberation on $\preceq$ amounts to mere wishful thinking, and has no legitimate claim to rationality. Until rules 1–5 are backed by some convincing intuition, the Party Example is under-described and therefore useless.

Another, more diplomatic answer to the objection is the following. Granted, some agents may legitimately base their cognitive deliberations on criteria other than exclusiveness. For such agents, the MOTB strategy is not an option. However, for those agents who do go for exclusiveness, MOTB is the rational strategy. Perhaps the best need not *always* meet, but we have at least shown that there are *some* contexts where they should. Relational PMC is not completely senseless.

Second objection. The exclusiveness-based defence of relational PMC only works given the identification of epistemic value with exclusiveness. But not a word has been said in defence of that identification.

Granted. As a defence of relational PMC the argument of the preceding section is only as strong as the concept of exclusiveness is intuitively appealing. Again, though, a pluralistic attitude may be fruitful. Even if one cannot accept exclusiveness as the *only* property of belief sets worthy of attention, it is at least *one* such property. An agent

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

concentrating solely on exclusiveness may then be regarded as one limiting case of rationality, and as such may merit some interest.

Third objection.[6] The restriction (E1) on exclusiveness relations is strongly lexicographic, in that a belief set B is preferable to a belief set A as soon as some world u not excluded by A is more implausible than any world that is not excluded by B. But what if B allows a great number of distant worlds, though not quite as implausible as u, while A only allows one such distant possibility, namely, u? Should not, intuitively speaking, the numerosity of the worlds allowed by B overrule the fact that the single world u allowed by A is slightly more distant, that is, should we not have $B \prec A$, contrary to (E1)?

This objection will be seen as all the more serious when we observe that without (E1), (PE) falls and with it our defence of MOTB. To see this, suppose (1) that A rules out all worlds except u and the worlds in $[K]$; (2) that all worlds allowed by B, except those in $[K]$, are equally distant, namely slightly less so than u; and finally (3) that, contrary to (E1), we humour the objector's intuitions, letting $B \prec A$. Now let $\mathfrak{A}=\{A\}$ and $\mathfrak{B} = \{R_w : w \in [B]\text{–}[K]\}$ (i.e., $\mathfrak{B}$ is the set of those belief sets compatible with all worlds in $[K]$ plus exactly one w in $[B]$–$[K]$). We have $\mathfrak{A} \subseteq \mathfrak{R}$, $\mathfrak{B} \subseteq \mathfrak{R}$, and $(\forall A' \in \mathfrak{A})\ (\forall B' \in \mathfrak{B})\ A' \prec B'$. Now, if there should happen to be some a such that $\mathfrak{A}$ and $\mathfrak{B}$ are both subsets of $K \perp a$, the elements of $\mathfrak{B}$ being, moreover, the $\preceq$-maximal ones in $K \perp a$ – then we will find ourselves in a predicament analogous with that in the Party Example: why should the best meet, giving us $K \dot{+} a = \cap\mathfrak{B} = B$, when taking instead the (trivial) intersection $\cap\mathfrak{A} = A$ would deliver a better result, as measured by $\preceq$? (E1), or something like it, is needed to rule out this situation, so if its lexicographic nature is counterintuitive, this is a grave matter, as our objective is precisely to make MOTB look intuitively reasonable.

We need, then, an argument to make implausibility look much more undesirable than sheer numerosity in worlds allowed by a belief set. Here is an attempt: under our interpretation of the distance relation $\leq$, the fact that a world is very distant *means* precisely that the agent is heavily disinclined to regard it as a serious possibility – or equivalently, is very much inclined to adopt a belief set which rules it out. On the other hand, great quantities in worlds allowed just amounts to great uncertainty, which – while arguably an undesirable state for a cognitive agent to be in – is not in itself a good argument not to adopt a belief state. Just saying, 'No, I don't like this belief set, because it's too weak and I want to know a lot of things' will not do.

Note, though, that we must not go too far in discrediting quantity-based preferences if another of our principles – (E3) – is to pass our scrutiny. Ideally, we want a delicately balanced way of looking at things

[6] Wlodek Rabinowicz, private correspondence.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

which makes quantity of worlds not totally unimportant, yet absolutely less weighty than plausibility.

Fourth objection. The notion of exclusiveness is, for intuitive purposes, dependent on the notion of 'implausibility of possible worlds', in terms of which it has been defined. But the nature of the latter relation is by no means intuitively clear. If we want to propound PMC as preferable to Grovean, world-based contraction on the grounds of its greater intuitive clarity, then we have to make it work without recourse to the idea of implausibility. On the other hand, if we do allow ourselves the luxury of an implausibility relation, then, arguably, we have no need, intuitively or formally, for PMCs in the first place: we can simply make do with the more straightforward world-based Grove contractions. Thus, the notion of exclusiveness is either illegitimate (if we are not allowed to talk of implausibility) or superfluous (if we are).

Perhaps this is right. But even if PMC cannot be made to look, all things considered, superior to world-based contraction, it is good to see that it can at least be regarded as sensible.

Fifth objection. (PE) is limited to cases where the language is finite. When we drop this restriction, there are examples bearing the same problematic structure as that in the Party Example. More precisely, taking Cn to be classical sentential logic, we have:

Observation 3: There are $\mathfrak{L}$, K, a, $\preceq_e$, $\mathfrak{A}$, and $\mathfrak{B}$ such that
$\mathfrak{L}$ is a language with a denumerable set of atomic formulae;
K is a logically closed subset of $\mathfrak{L}$;
$\preceq_e$ is an exclusiveness relation for K in $\mathfrak{L}$ that is connected in $\mathfrak{R}$ and transitive;
$a \in K$;
$\mathfrak{A}, \mathfrak{B} \in K \perp a$;
$\forall A \in \mathfrak{A}\ \forall B \in \mathfrak{B}\ A \prec_e B$; yet
$\cap \mathfrak{B} \prec_e \cap \mathfrak{A}$.
(See Appendix for an example.)

This observation is the more significant as, in forming the notion of exclusiveness, we took care to look at the infinite case as well as the finite one – as in the discussion of Grove's condition (S$\leq$3), which is automatically fulfilled in the finite case, and which we would therefore not even have bothered to consider if we had only been concerned with finite models.

The only apology I can think of is that having something that works in the finite case is better than having nothing at all; especially since, if we are trying to model real-world thinkers, finite models are more realistic.

Sixth objection. The notion of exclusiveness is tailor-made to work with AGM's version of MOTB. But there are numerous other contraction models of MOTB form. What about them?

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

Answer (or, rather, non-answer): the present paper has raised, but does not attempt to solve, the problem of justifying the MOTB strategy in general. Even such (hypothetical) readers as find the notion of exclusiveness entirely satisfactory in the AGM case should be aware that this in itself will not handle the general issue. To what extent the present solution can be extended to other models is not a question that I shall address here.

10. RELATED IDEAS

The strategy, employed here, of defining, or at least motivating, a preference relation over sets of beliefs in terms of some other relation, possibly over some other entities, has of course been suggested by previous authors in various contexts. In view of the limitations of the notion of exclusiveness, it would be interesting to know whether any of these earlier suggestions might serve us better. There is no room for a comprehensive overview here, but I shall at least look at a number of constructions, pointing out why, with a possible exception, they are not suited to the task we have set ourselves in this essay. (In some cases, acquaintance with the cited papers may be necessary on the reader's part in order to find my discussion intelligible.)

Levi (1991, pp. 127 f.) introduces a notion of what he calls 'damped informational value', which is a quality possessed by potential belief sets in addition to, and determined by, their primary form of epistemic value (which Levi calls 'probability-based informational value'). Elements of $K\perp a$ always have both kinds of value in equal degree, but for intersections of subsets of $K\perp a$, they may differ.[7] For any such subset $\mathfrak{A}$, the damped informational value of $\cap\mathfrak{A}$ is postulated to be equal to the smallest damped informational value of any element of $\mathfrak{A}$, whereas the probability-based value of the intersection will typically be less than that of any $\mathfrak{A}$-element. Levi's explicit aim in introducing this notion is to legitimize taking, as contraction outcome, an intersection of elements of $K\perp a$, rather than a single element of $K\perp a$. But it will be seen that his construction in a sense could serve our purposes as well: sets of better elements of $K\perp a$ will, by postulation, have better intersections – as measured by damped informational value.

Unfortunately, Levi does not even attempt to give an intuitive account of what, in the 'real world' of cognitive deliberation, the technical notion damped informational value is supposed to represent, nor any reason why a rational agent should strive to maximize it in the

[7] Note that we are here diverging both from Levi's notation and from that of AGM, using the symbol '$\perp$' in the wide sense of this paper's Section 2. Levi's criterion for inclusion in $K\perp a$ (as we put it) is that of being *a-saturatable*. A subset X of K is *a*-saturatable iff the result of expanding X with respect to $\neg a$ is a maximal consistent subset of the language.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

first place. This means that if applied as-is for our purposes, the notion will be open to fatal charges of *ad hoc*-ness, and therefore useless.

Rott (1991, p. 158) defines a function which, given an epistemic entrenchment ordering[8] of the agent's beliefs, delivers a preference relation over arbitrary subsets of the language. Formally, there is nothing to prevent us from basing a relational PMC function on this relation. The question we face is: would this make intuitive sense, and if so, is Rott's relation perhaps a better alternative to our own exclusiveness relation?

Perhaps, but there is a problem. Rott proves (p. 161) that his relation – which we may label '$\preceq$' – has the following property:

(P4) $M \preceq M \cap N$ or $N \preceq M \cap N$.

It is not difficult to see that a relation satisfying (P4) cannot in general be maximizing (the condition we have labelled '(E3)'). This means that there may well be, among non-a-implying subsets of K, sets that are $\preceq$-maximal without being inclusion-maximal. It then ceases to be obvious why, in contraction, we are to restrict our initial attention to elements of $K \perp a$. Why should just the best elements of $K \perp a$ meet, if there may be just as good subsets of K sitting around outside $K \perp a$? Also, as we saw in Section 7, (E3) is essential for the argument from the Maxim for belief contraction. These difficulties must be addressed if Rott's construction is to be a credible alternative to the notion of exclusiveness.

Grove (1988, p. 164) proposes a method for constructing a relation over sentences out of his distance relation $\leq$ on possible worlds. (Formally, the method parallels that of Lewis (1973, p. 53), but since, of the two, only Grove is directly concerned with belief dynamics, we shall restrict our attention to his paper.) If we denote the defined relation by '$\leq_G$', Grove's construction, in effect, is this: for any two sentences a and b, $a \leq_G b$ iff there is a $u \in [a]$ such that for every $v \in [b]$, $u \leq v$. This definition could of course be extended to arbitrary sets of sentences, but a relation thus defined would be useless for PMC purposes. Since, for every $K' \in K \perp a$, $[K']$ contains $\leq$-minimal worlds, all elements of $K \perp a$ will be '$\leq_G$-best' elements, and we will find ourselves in full meet contraction.

There may be natural modifications of Grove's idea which dodge this particular objection. We might, for instance, focus our attention on the worlds excluded by elements of $K \perp a$ rather than worlds allowed by them. But we would still face the problem that arises from Rott's construction: the defined relation will not in general be maximizing.

In Nebel (1992, p. 59), a method is given for constructing a relation $\preceq$ over $\mathfrak{P}(K)$ out of a (not necessarily deductively closed) set C of believed

[8] Readers unacquainted with the notion of epistemic entrenchment may consult Gärdenfors & Makinson (1988), Gärdenfors (1992), or Hansson (1999) for useful introductions.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

sentences, plus a transitive and strongly connected relation $\leq_N$ over C. In the general case, Nebel's construction offers no guarantee that $\preceq$ will be suitable for MOTB contraction; it is not difficult to see that, for example, the problematic $\preceq$ of the Party Example can be thus construed (put $C = \{a_1, a_2, b, c_1, c_2\}$ and $c_1,c_2 <_N b <_N a_1,a_2$). However, Nebel also (in Section 5) discusses the case where, in effect, $C=K$ and $\leq_N$ satisfies the postulates of epistemic entrenchment. If this construction can be provided with a convincing intuitive basis, and shown to satisfy some sufficiently strong version of our principle (P), then it might make a viable alternative to the notion of exclusiveness. A thorough investigation of this suggestion must, however, fall outside the scope of the present essay.

11. CONCLUSION

I think I have pointed to a serious problem with the MOTB approach to belief revision. In formulating the notion of exclusiveness, I have made one salvage attempt, but, in view of its limitations, it can hardly be considered anything like a final solution. I offer it only as a starting point, open to refinements and improvements.

Quite conceivably, though, the best way of justifying the MOTB strategy is along lines rather different from those followed here. My notion of exclusiveness is based on judgements of plausibility, or, in a wide sense of the word, likelihood. But there are other possibilities: for instance, philosophers of a pragmatist bent may be able to find an interpretation of $\preceq$ in terms of pragmatic utility which will validate some suitable version of (P).

A completely different defence strategy might be to try and do without (P) altogether. To be sure, the argument outlined in Section 3 requires some version of (P) – but there may be other arguments. Perhaps it can somehow be maintained that having the best meet is a rational thing to do even if the meet in question is not particularly good in the same sense as the selected candidate belief sets.

It is thus too early to give up on the idea of MOTB contraction. There are many avenues to be explored in the quest for its justification – and perhaps doing so should be given a higher priority by researchers in belief revision. The numerous investigations into the formal properties of different MOTB contractions that have been performed to date would, to this author at least, hold considerably greater interest if someone could explain why any rational agent would want to engage in such contractions in the first place.

APPENDIX: PROOFS

The proofs offered here are a bit compressed; more painstaking versions may be requested from the author.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

Proof of Observation 1.

Let $\dot{-}$ be an operator of relational PMC. We need to find a relation $\leq$ on W and a relation $\preceq'$ on $\mathfrak{P}_{\mathrm{Cn}}(K)$ satisfying (E1) through (E6) and such that $\preceq'$ generates $\dot{-}$.

By definition, there is some relation $\preceq$ on $\mathfrak{P}_{\mathrm{Cn}}(K)$ which generates $\dot{-}$. We define $\leq$ from $\preceq$ and $\preceq'$ from $\leq$:

(Def$\leq$) $u \leq w$ iff
$u=v$ or $u \in [K]$ or ($u \notin [K]$, $v \notin [K]$, $R_v \preceq R_u$, and $R_u \preceq R_u$).

(Def$\preceq'$) $A \preceq' B$ iff
$(\exists u \in [A])(\forall v \in [B])v<u$ or $A \subseteq B$ or ($A,B \in \mathfrak{R}$ and $w_B \leq w_A$).

The definitions immediately give us (E4), (E5), and (E6). We also have the following property:

($\preceq'$1) $A \prec' B$ iff either $(\exists u \in [A])(\forall v \in [B])v<u$ or $A \subset B$.

We omit the straightforward, if slightly laborious, proof.

$\leq$ and $\preceq'$ satisfy (E1) and (E3). Immediate from ($\preceq'$1).

$\leq$ and $\preceq'$ satisfy (E2). Suppose $A,B \in \mathfrak{R}$. We want to show that $A \preceq' B$ iff $w_B \leq w_A$. *Right-to-left* is immediate from (Def$\preceq'$). For *left-to-right*, suppose $A \preceq' B$. We have three cases according to (Def$\preceq'$). In the *first* case, $(\exists u \in [A])(\forall v \in [B])v<u$. By (E6), the u in question cannot be in $[K]$, and therefore can only be w_A. Hence $w_B<w_A$, whence $w_B \leq w_A$. In the *second* case, $A \subseteq B$, which, since $A,B \in \mathfrak{R}$, implies $A=B$, wherefore, by (Def$\leq$), $w_B \leq w_A$. And in the *third* case, we have $w_B \leq w_A$ immediately.

Furthermore,

($\preceq'$2) For any $A,B \in \mathfrak{R}$, $B \preceq' A$ iff
$A=B$ or ($B \preceq A$ and $A \preceq A$).

Proof: By (E2) and (Def$\leq$). Finally,

$\preceq'$ generates $\dot{-}$.
Proof: since, *ex hypothesi*, $\preceq$ generates $\dot{-}$, we will be done if we can demonstrate that $\preceq$ and $\preceq'$ generate the same contraction operation. For this purpose, by the definition of relational partial meet contraction, it suffices to show that

(*) for any $a \in \mathfrak{Q}$, and any $A \in K \perp a$,
A is $\preceq'$-maximal in $K \perp a$ iff A is $\preceq$-maximal in $K \perp a$.

If $K \perp a$ is empty, (*) holds vacuously. Now suppose $K \perp a$ is non-empty.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

For *left-to-right*, suppose for *reductio* that A is $\preceq'$-maximal but not $\preceq$-maximal in $K\perp a$. Then, for some $B_0\in K\perp a$, $B_0\preceq' A$ but $\neg(B_0\preceq A)$. By ($\preceq'2$), this entails $A=B_0$. Hence, $\neg(A\preceq A)$. Now regard an arbitrary $B\in K\perp a$. Since A is $\preceq'$-maximal, $B\preceq' A$. By ($\preceq'2$) and $\neg(A\preceq A)$, this entails that $B=A$. Hence, A is the only element of $K\perp a$, wherefore, since $\neg(A\preceq A)$, $K\perp a$ contains no $\preceq$-maximal element. But, *ex hypothesi*, $\dot{-}$ is based on a selection function γ which selects from every non-empty $K\perp a$ (and we have assumed $K\perp a$ to be non-empty) a non-empty set of $\preceq$-maximal elements. Contradiction.

Right-to-left follows from ($\preceq'2$).

Proof of Observation 2.

Let $\dot{-}$ be an operator of transitively relational PMC. By Observation 5.1 of Achourrón *et al*. (1985, p. 523), there is some connected and transitive relation $\preceq$ on $\mathfrak{P}_{\mathrm{Cn}}(K)$ which generates $\dot{-}$. Define $\leq$ and $\preceq'$ in terms of $\preceq$ as in the preceding proof; any properties of $\leq$ and $\preceq'$ demonstrated there still hold. Now define $\preceq''$, our desired transitive exclusiveness relation, as follows.

(Def$\preceq''$) $A\preceq'' B$ iff there are $C_0,\ldots,C_n$ (where n is a natural number) such that $A=C_0$, $B=C_n$, and for every i ($1\leq i\leq n$), $C_{i-1}\preceq' C_i$.

A few lemmas are needed in order to show that $\preceq''$ has the desired properties:

($\preceq'3$) For any $A,B,C\in\mathfrak{P}_{\mathrm{Cn}}(K)$, if $A\preceq' B$ and $B\prec' C$, then $A\prec' C$.

We have six cases in all: three for $A\preceq' B$ (according to (Def$\preceq'$)) times two for $B\prec' C$ (according to ($\preceq'1$)). We only deal here with the three least straightforward ones, leaving the remainder to the reader. We will trade on the fact, following from (E6) and the transitivity of $\preceq$, that $\leq$ is also transitive.

First case: $(\exists u\in[A])(\forall v\in[B])v<u$ and $(\exists w\in[B])(\forall x\in[C])x<w$. It follows that $(\exists u\in[A])(\forall x\in[C])x<u$, whence $A\prec' C$ by ($\preceq'1$).

Second case: $A,B\in\mathfrak{R}$ and $w_B\leq w_A$ and $(\exists w\in[B])(\forall x\in[C])x<w$. Regard an arbitrary $x\in[C]$. For some $w\in[B]$, $x<w$. By (Def$\leq$), we cannot have $w\in[K]$. The only element in $[B]$-$[K]$ is w_B, so $w=w_B$. So, for our arbitrary $x\in[C]$, we have $x<w_B$ and $w_B\leq w_A$, wherefore $x<w_A$. Therefore, by ($\preceq'1$), $A\prec' C$.

Third case: $A,B\in\mathfrak{R}$ and $w_B\leq w_A$ and $B\subset C$. Since $B\in\mathfrak{R}$, there is an $a\in\mathfrak{L}$ such that B is an inclusion-maximal non-a-implying subset of K. This entails $a\in C$. We are now going to show that $K\subseteq C$. Regard an arbitrary $b\in K$. $(a\rightarrow b)\rightarrow a$ is tautologically equivalent to a, so since B fails to imply a, $(a\rightarrow b)\rightarrow a\notin \mathrm{Cn}(B)$. By the deduction property of Cn, $a\notin\mathrm{Cn}(B\cup\{a\rightarrow b\})$.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

Since $b\in K$, we have $a\rightarrow b\in K$, and therefore $B\subseteq \mathrm{Cn}(B\cup\{a\rightarrow b\})\subseteq K$. Since a is entailed by every subset of K which is a strict superset of B, we must have $B=\mathrm{Cn}(B\cup\{a\rightarrow b\})$, and hence $a\rightarrow b\in B$. This, in conjunction with $B\subset C$ and $a\in C$, gives us $b\in C$. Since this result was derived from the assumption that $b\in K$, $K\subseteq C$. Since $A\in\mathfrak{R}$, $A\subset K$, and we reach $A\prec' C$ via $A\subset C$.

($\preceq''$1) If $A\prec' B$ then $A\prec'' B$.

Suppose that $A\prec' B$. Then $A\preceq' B$ and hence, by (Def$\preceq''$), $A\preceq'' B$. Now suppose for *reductio* that $A\prec'' B$ does not hold; this requires that $B\preceq'' A$. By (Def$\preceq''$), then, there are $C_0,\ldots,C_n$ such that $B=C_0$, $A=C_n$, and for every i $(1\leq i\leq n)$, $C_{i-1}\preceq' C_i$. Specifically, $C_{n-1}\preceq' C_n=A$. This, taken together with $A\prec' B$ and ($\preceq'$3) gives us $C_{n-1}\prec' B$. Now, again by ($\preceq'$3), $C_{n-2}\preceq' C_{n-1}$ gives us $C_{n-2}\prec' B$. Continuing in this manner, we eventually reach $C_0\prec' B$, which contradicts $B=C_0$.

($\preceq''$2) If $A,B\in\mathfrak{R}$, then $A\preceq' B$ iff $A\preceq'' B$.

Left-to-right: immediate from (Def$\preceq''$).

Right-to-left: connectivity of $\preceq$ plus ($\preceq'$2) gives connectivity of $\preceq'$ in $\mathfrak{R}$. Then apply ($\preceq''$1).

We can now conclude that $\preceq''$ has all the desired properties.

$\leq$ and $\preceq''$ satisfy (E1). Directly from ($\preceq''$1) and the fact that $\leq$ and $\preceq'$ satisfy (E1).

$\leq$ and $\preceq''$ satisfy (E2). Directly from ($\preceq''$2) and the fact that $\leq$ and $\preceq'$ satisfy (E2).

$\preceq''$ satisfies (E3). Directly from ($\preceq''$1) and the fact that $\preceq'$ satisfies (E3).

$\preceq''$ satisfies (E4). Directly from (Def$\preceq''$) and the fact that $\preceq'$ satisfies (E4).

$\leq$ satisfies (E5) and (E6). Established in previous proof.

$\preceq''$ is transitive. Immediate from (Def$\preceq''$).

$\preceq''$ generates $\dot{-}$. Directly from ($\preceq''$2) and the fact that $\preceq'$ generates $\dot{-}$.

Proof of Observation 3.

Notation: in this proof, the variables 'n' and 'm' denote natural numbers. Now let

$\mathfrak{L}$ $=_{df}$ the language consisting of the denumerably infinite set of sentential letters $\{G_0, G_1, G_2,\ldots\}$ and their truth-functional combinations;
Cn $=_{df}$ classical sentential logic;
for every n, $w_n =_{df} \mathrm{Cn}(\{G_0, G_1,\ldots, G_{n-1}, \neg G_n, \neg G_{n+1},\ldots\})$;
$w_\omega =_{df} \mathrm{Cn}(\{G_0, G_1,\ldots,G_n,\ldots\})$; and
$W =_{df}$ the set of maximal consistent subsets of $\mathfrak{L}$.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

Clearly, all w_n as well as w_ω are elements of W.

Furthermore, let

$K =_{df} \mathrm{Cn}(\{\neg G_5\} \cup \{G_n \rightarrow G_m\colon n>m\})$;
$\leq =_{df}$ the relation over W such that $u \leq v$ iff u is located at least as far up as v in the following table:
w_0 through w_5
$w_{21}, w_{22}, w_{23}, \ldots$
w_6 through w_{20}
w_ω
all others

(Suggested informal reading of G_n: 'The number of errors in my paper is greater than n'. $\leq$ may be intuitively motivated by supposing that the agent has heard from a very convincing referee that the number of errors is at most 5, from a moderately convincing one that it is greater than 20; and that (s)he considers w_ω quite unreasonable (the paper may be too long, but it is, after all, finite!) and the worlds violating $\{G_n \rightarrow G_m\colon n>m\}$ weird beyond intelligibility.)

Define $\preceq_e$ from $\leq$ as $\preceq''$ is defined from $\leq$ in the proof of Observation 2, and let

$a =_{df} \neg G_{10}$;
$\mathfrak{B} =_{df} \{K \cap w_n\colon n>20\}$; and
$\mathfrak{A} =_{df} \{K \cap w_n\colon 11 \leq n \leq 20\}$.

Then $\mathfrak{A}, \mathfrak{B} \subseteq K \perp a$ and $\mathfrak{B}$ is the set of $\preceq_e$-maximal elements of $K \perp a$.

We have $\forall A \in \mathfrak{A}\ \forall B \in \mathfrak{B}\ A \prec_e B$. It remains to show that $\cap\mathfrak{B} \prec_e \cap\mathfrak{A}$. Since w_ω is more implausible than any world in $[\cap\mathfrak{A}] = \{w_0, \ldots, w_5, w_{11}, \ldots, w_{20}\}$, by (E1) it suffices to show that $w_\omega \in [\cap\mathfrak{B}]$. Suppose to the contrary that $w_\omega \notin [\cap\mathfrak{B}]$, or, equivalently, $\cap\mathfrak{B} \not\subseteq w_\omega$. Then for some $p \in \mathfrak{L}$, $(\forall B \in \mathfrak{B}\ p \in B) \wedge p \notin w_\omega$; hence $\forall n\ (n>20 \rightarrow p \in K \cap w_n) \wedge p \notin w_\omega$; hence $\forall n\ (n>20 \rightarrow p \in w_n) \wedge p \notin w_\omega$. But by induction over the length of p, we also have $\exists n \forall m (m \geq n \rightarrow (w_m \in [p] \leftrightarrow w_\omega \in [p]))$: contradiction.

The verification of the remaining claims listed in Observation 3 is left to the reader.

REFERENCES

Alchourrón, Carlos E., Peter Gärdenfors and David Makinson. 1985. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–30

Cantwell, John. 1999. Some logics of iterated belief change. *Studia Logica* 63:79–89

Gärdenfors, Peter. 1992. Belief revision: an introduction. In *Belief Revision*, pp. 1–28. Peter Gärdenfors (ed.). Cambridge University Press

Gärdenfors, Peter and David Makinson. 1988. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pp. 83–95. Moshe Y. Vardi (ed.). Morgan Kaufmann Publ.

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press

Grove, Adam. 1988. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–70

Hansson, Sven Ove. 1992. Similarity semantics and minimal changes of beliefs. *Erkenntnis*, 37:401–29

Hansson, Sven Ove. 1994. Taking belief bases seriously. In *Logic and Philosophy of Science in Uppsala*, pp. 13–28. Dag Prawitz and Dag Westerståhl (eds.). Kluwer Academic Publishers

Hansson, Sven Ove. 1999. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Kluwer Academic Publishers

Hansson, Sven Ove and Wlodek Rabinowicz. (eds.). 1995. *Logic for a Change: Essays Dedicated to Sten Lindström on the Occasion of His Fiftieth Birthday*. Uppsala Prints and Preprints in Philosophy

Levi, Isaac. 1991. *The Fixation of Belief and Its Undoing*. Cambridge University Press

Lewis, David. 1973. *Counterfactuals*. Harvard University Press

Nebel, Bernhard. 1992. Syntax based approaches to belief revision. In *Belief Revision*, pp. 52–88. Peter Gärdenfors (ed.). Cambridge University Press

Olsson, Erik J. 1997. A coherence interpretation of semi-revision. In *Coherence: Studies in Epistemology and Belief Revision*, pp 5:1–5, 29. Doctoral dissertation, Acta Universitatis Upsaliensis

Rott, Hans. 1991. Two methods of constructing contractions and revisions of knowledge systems. *Journal of Philosophical Logic*, 20:149–73

Rott, Hans. 1992. Modellings for belief change: prioritization and entrenchment. *Theoria*, 58:21–57

Rott, Hans and Maurice Pagnucco. 1999. Severe withdrawal (and recovery). *Journal of Philosophical Logic*, 28:501–47

https://doi.org/10.1017/S0266267100000262 Published online by Cambridge University Press