Belief Revision and Relevance[1]

Peter Gärdenfors

University of Lund

1. Belief Revisions as Minimal Changes of Relevant Beliefs

The theory of belief revision deals with models of states of belief and *transitions* between states of belief. The goal of the theory is to describe what should happen when you update a state of belief with new information. In the most interesting case, the new information is *inconsistent* with what you believe. This means that some of the old beliefs have to be deleted if one wants to remain within a consistent state of belief. A guiding idea is that the change should be *minimal* so that as few of the old beliefs as possible are given up.

A central problem for the theory of belief revision is what is meant by a minimal change of a state of belief. The solution to this problem depends to a large extent on the *model* of a state of belief that is adopted. In the literature, two types of models dominate: One where a state of belief is described by a *set of sentences* from a given language, sometimes called a *belief set*, and another where states of belief are modelled by *probability functions* defined over the language. I shall briefly outline these models of belief revision in Section 3.

The criteria of minimality used in these models have been based on almost exclusively *logical considerations*. However, there are a number of non-logical factors that should be important when characterizing a process of belief revision. The focus of this article will be the notion of *relevance*. The key criterion to be developed here is the following:

(I) If a belief state K is revised by a sentence A, then all sentences in K that are *irrelevant* to the validity of A should be retained in the revised state of belief.

In my opinion, (I) has a solid intuitive support. However, a criterion of this kind cannot be given a technical formulation in a model based on belief sets built up from sentences in a simple propositional language because the notion of relevance is not available in such a language. In models based on probability functions, there is a standard definition of relevance that could be used to formulate the desired principle.

PSA 1990, Volume 2, pp. 349-365

However, as shall be shown below, the traditional definition suffers from some shortcomings that make it unsuitable to use in a more precise formulation of criterion (I).

So, before we can proceed to a version of criterion (I) that could be added to a theory of belief revision, based on either belief sets or probability functions, we must analyse the notion of relevance itself. This will be the purpose of Section 2. Only after this can we return to a theory of belief revision that incorporates the principle (I).

2. On the Logic of Relevance

In the traditional treatment of the notion of relevance, it is defined in terms of a probability function. However, since we want to develop an analysis of 'relevance' that can be used in various forms of models of belief states, a characterization that does not rely on probabilistic notions would be useful. This will be one of the aims of this section. Much of the material to be presented here is adopted from Gärdenfors (1978).

2.1 The Standard Definition

The traditional way of introducing the relevance relation is to define it with the aid of a given probability measure P in the following way:[2]

(D1) (a) A is *relevant* to C iff
$P(C/A) \neq P(C)$

(b) A is *irrelevant* to C iff
$P(C/A) = P(C)$

More general versions of this definition, but with the same basic idea have been studied by David (1979), Geiger (1990) and Pearl (1988). The implications for problems within philosophy of science are investigated by Salmon (1971, 1975) among others. Carnap (1950) points out that the theorems on irrelevance become simpler if the following definition of irrelevance is adopted instead:[3]

(b') A is *irrelevant* to C iff
$P(C/A) = P(C)$ or A is logically false.

If it is assumed that only logically false sentences have zero probability, then this definition has the consequence that any sentence A is either relevant or irrelevant to C. In the following sections I will adopt Carnap's suggestion, so when (D1) is mentioned, the conjunction of (a) and (b') is referred to.

There are two problems connected with (D1). One was already pointed out by Keynes (1921) who was among the first to discuss the concept of relevance. He observes that, intuitively, there is a stronger sense of 'relevance' which is not covered by (D1). In connection with his discussion of the 'weight' of arguments, he writes:[4]

> "If we are to be able to treat 'weight' and 'relevance' as correlative terms, we must regard evidence as relevant, part of which is favourable and part unfavourable, even if, taken as a whole, it leaves the probability unchanged. With this definition, to say that a new piece of evidence is 'relevant' is the same thing as to say that it increases the 'weight' of the argument."

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

Here Keynes is referring to the case when $P(C/A \& B) = P(C)$, even though $P(C/A) \neq P(C)$ and $P(C/B) \neq P(C)$, which, according to (D1), means that A & B is irrelevant to C, while both A and B, taken as separate pieces of evidence, are relevant to C.

In order to capture this stronger sense of 'relevance', Keynes proposes the following definition which, he believes, "is theoretically preferable":[5]

(D2) (a) A is *irrelevant* to C iff
there is no sentence B, which is derivable from A such that $P(C/B) \neq P(C)$.

(b) A is *relevant* to C iff A is not irrelevant to C.[6]

This definition has the consequence that if A is relevant to C, then, for any sentence B such that A & B is not logically contradictory, A & B is also relevant to C and thus it blocks the seemingly counterintuitive feature of (D1) mentioned above.

Carnap shows that the definition (D2) leads to the following trivialization result:[7] if neither C or $\neg C$ are logically valid, then A is irrelevant to C iff A is logically valid. This is certainly absurd. For most sentences C there are many other sentences that we judge as irrelevant to C. (D2) is therefore not the appropriate way to define the relevance relation in the stronger sense hinted at by Keynes. The question is now whether it is possible to give a definition of this relation that satisfies Keynes's requirement.

The second problematic feature of (D1) is that as soon as $P(A) = 1$, it follows that A is irrelevant to B, *for any sentence* B. This is highly counterintuitive because if P is taken to model the current state of belief, and A is a contingent sentence, then $P(A) = 1$ means that A is held to be a true fact, but this does not entail that there are no sentences that are relevant for the fact that A. We will return to this drawback of (D1) in Section 4.

I will now first show that Carnap's trivialization result is not dependent on the definition (D2) or any other definition in terms of probability measures. I will formulate some general criteria for the relevance relation and show that if Keynes's requirement is added, then the trivialization result will follow. When formulating the criteria, it will not be assumed that the relevance relation is to be explicated in terms of probability measures.

Because of the trivialization result, I conclude that Keynes's requirement has to be abandoned. However, this should not prevent us from seeking a definition of 'relevance' that is stronger than (D1) and that follows Keynes's (and our) intuitions as far as possible. I will present two criteria for the relevance relation which are weaker than Keynes's requirement but which are not satisfied by (D1). Their logical consequences will be investigated. Finally, I will propose a new definition of the relevance relation that satisfies one of these criteria and briefly investigate its properties.

2.2. Basic Criteria for the Relevance Relation

In this paper, relevance is taken to be a relation between sentences. I therefore assume that there be a given language $\mathfrak{L}$ where the sentences are taken from. This language is assumed to be closed under standard truth-functional operations. I will use A, B, C, etc. as symbols for sentences. If A is provable, I will write $\vdash A$. A sentence A is said to be *contingent*, if neither $\vdash A$, nor $\vdash \neg A$. The expression 'A is relevant to C' will be abbreviated $A \mathbb{R} C$, and similarly, 'A is irrelevant to C' will be written $A \perp\!\!\!\perp C$.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

I will now proceed to formulate some general criteria for the relevance relation. The criteria are not intended to be a complete characterization of the logic or 'relevance', but are rather meant to be as weak as possible.

(R0) If $\vdash A \leftrightarrow B$, then $A \mathbb{R} C$ iff $B \mathbb{R} C$.

This is a simple rule of replacement of logical equivalents.

(R1) $A \mathbb{R} C$ iff not $A \perp\!\!\!\perp C$.

Relevance and irrelevance are complementary and mutually exclusive relations. Carnap saw this criterion as an argument for changing (b) in (D1) to (b').

(R2) $A \mathbb{R} C$ iff $\neg A \mathbb{R} C$.

If one obtains some new information about the sentence C when learning that A, then one also learns something about C when $\neg A$ is added.

From (R1) and (R2) we can derive

(1) $A \perp\!\!\!\perp C$ iff $\neg A \perp\!\!\!\perp C$.

(R3) $(A \vee \neg A) \perp\!\!\!\perp C$.

Counting $A \vee \neg A$ as new evidence does of course not affect our judgement of the degree of truth of C. From (R3) and (R2) we can derive

(2) $(A \& \neg A) \perp\!\!\!\perp C$.

This is in accordance with Carnap's changing (b) in (D1) to (b'), and it enables us to formulate (R2) without restrictions.

A consequence of (R0), (R1), (R3) and (2) is

(3) If $A \mathbb{R} C$, then A is contingent.

The following condition is introduced in order to secure that relevance is a non-empty relation.

(R4) If C is contingent, then $C \mathbb{R} C$.

If it is assumed that only sentences which are logical consequences of the evidence have probability one, then (D1) fulfills the requirements (R0) – (R4). I take these criteria to be necessary for any explication of the relevance relation.

2.3. A Trivialization Result

We next turn to Keynes's requirement. In connection with his definition, which I call (D2), he gives the following argument:[8]

> "Any proposition which is irrelevant in the strict sense [i.e., according to (D2)] is, of course, also irrelevant in the simpler sense [i.e., according to (D1)] but if we were to adopt the simpler definition, it would sometimes occur that a part of evidence would be relevant, which taken as a whole was irrelevant."

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

This quotation motivates the following criterion:[9]

(R5) If A $\mathbb{R}$ C and not $\vdash \neg(A \,\&\, B)$, then (A & B) $\mathbb{R}$ C.

As we have already observed, (D1) does not satisfy (R5) for any non-trivial probability measure P. The following simple lemma will show the connection between (R5) and (D2) and throw some light on why Keynes chose this definition for his stronger concept of relevance.

LEMMA: If (R0) is assumed, then the following criterion is equivalent to (R5):

(4) If B $\mathbb{R}$ C, $A \rightarrow B$, and not $\vdash \neg A$, then A $\mathbb{R}$ C.

The proof of the lemma and the following three theorems can be found in Gärdenfors (1978).

I will now show that (R5) leads to strongly counterintuitive consequences, if combined with the criteria (R0) – (R4).

THEOREM 1: If the relations $\mathbb{R}$ and $\perp\!\!\!\perp$ satisfy (R0) – (R5), then every contingent sentence is relevant to every other contingent sentence.

This theorem presents us with a dilemma. On the one hand, there seems to be some truth in the observation that (D1) does not cover our intuitive conception of 'relevance', and, on the first impression, Keynes's requirement seems acceptable. On the other hand, the remaining criteria for the relevance relation, needed to derive the theorem, are seemingly innocent. However, the consequence that all non-trivial sentences are relevant to any contingent sentence is strongly counterintuitive.

In my opinion, the only reasonable way out of the dilemma is to reject the assumption that (R5) is valid. This does not mean, however, that (D1) has to be accepted as the correct definition of the relevance relation.

The unsatisfactory feature of (D1) is, roughly, that it makes *too few* sentences relevant. This view is supported by the quotations from Keynes (1921) given above. One way to find a more appropriate definition of the relevance relation is therefore to investigate further general criteria that may be added to the basic criteria (R0) – (R4) and that enlarge the set of relevant sentences.

2.4. Two Further Criteria

In this section I will investigate the logical consequences of the following criteria:

(R6) If A $\mathbb{R}$ C, B $\mathbb{R}$ C, and not $\vdash \neg(A \,\&\, B)$, then (A & B) $\mathbb{R}$ C.

(R7) If A $\perp\!\!\!\perp$ C and B $\perp\!\!\!\perp$ C, then (A & B) $\perp\!\!\!\perp$ C.

These criteria will be called 'the conjunction criterion for relevance' and 'the conjunction criterion for irrelevance' respectively. Neither of these criteria is fulfilled by (D1). (R6) is a special case of (R5) and thus trivially derivable from (R5). A consequence of Theorem 1 is that the sentences that are irrelevant to a sentence C are those that are logically valid or invalid. From this it is easy to see that (R7) too is derivable from (R0) – (R5). Thus (R6) and (R7) are consequences of (R5) in the presence of (R0) – (R4). In fact, the converse is also true.

THEOREM 2: (R5) is derivable from (R6) and (R7) together with (R0) – (R4).

This theorem shows that (R6) and (R7) can not both be acceptable since Theorem 1 would then be derivable. In the sequel, it will be shown that neither (R6) nor (R7) is alone sufficient for (R5).

From (R6) and (R2) it is easy to derive the following condition:

(5) If A $\mathbb{R}$ C, B $\mathbb{R}$ C and not $\vdash$A v B, then (A v B) $\mathbb{R}$ C.

For a fixed sentence C, we see by (R2), (R6) and (5) that the set of sentences relevant to C is closed under truth-funcional operations, as long as these operations do not yield sentences that are logically valid or contradictory.

Using (R0) – (R4) one can show that (R7) is equivalent to

(6) If (A & B) $\mathbb{R}$ C, then A $\mathbb{R}$ C or B R C.

In words, this condition could be interpreted as saying that if a sentence is relevant, then some of its parts are also relevant. In a sense, this is the converse of (R5) which says that if a part of a sentence is relevant, then the sentence as a whole is relevant. As we have seen, (6) is derivable from (R0) – (R5).

Analogous to the case above is the possibility of deriving the following condition from (R7) and (1):

(7) If A $\perp\!\!\!\perp$ C and B $\perp\!\!\!\perp$ C, then (A v B) $\perp\!\!\!\perp$ C.

For a given sentence C, we conclude from (1), (R7) and (7) that the sentences relevant to C will be closed under truth-functional operations (with no restrictions). And, conversely, if the irrelevant sentences are closed under truth-functional operations, (1), (R7) and (7) will be fulfilled. This connection will be utilized in Section 5.

These results provide us with some ideas of the power of conditions (R6) and (R7). But, as we have seen, we cannot require both to be satisfied for a reasonable relevance relation. It is argued in Gärdenfors (1978) that (R7) is valid, but there are good counter-examples to (R6). Thus, the appropriate conditions for a relevance relation on this level seems to be (R0) - (R4) together with (R7). Next I would like to show that it is possible to improve the definition of irrelevance so that these conditions will be satisfied.

2.5 An Amended Definition of Irrelevance

In Gärdenfors (1978, p. 362) it is argued that in order to establish that A is irrelevant to C, it is not sufficient that P(C/A) =P(C), but we must also know that if we learned that A, then no sentences that are now irrelevant to C would become relevant to C on the new evidence A. This is in accordance with the earlier idea that (D1) makes too few sentences relevant. This argument motivates the following definition:

(D3) (a) A $\perp\!\!\!\perp$ C iff
P(A) = 0 or P(C/A) = P(C) and for all B such that P(C/B) = P(C) and P(A&B) $\neq$ 0, it also holds that P(C/A&B) = P(C).

(b) A $\mathbb{R}$ C iff not A $\perp\!\!\!\perp$ C.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

Note that the condition that P(C/A) = P(C) is a special case of "for all B such that P(C/B) = P(C) and P(A&B) ≠ 0, it also holds that P(C/A&B) = P(C)", namely, the case when B is a tautology. This means that (D3) can be simplified to:

(D3') (a) A ⫫ C iff
P(A) = 0 or for all B such that P(C/B) = P(C) and P(A&B) ≠ 0, it also holds that P(C/A&B) = P(C).

(b) A ℝ C iff not A ⫫ C.

THEOREM 3: (D3) satisfies (R0) - (R4), and (R7).

It is easy to show by a small finite example that (D3) can be satisfied nontrivially so that no trivialization result is possible for the set of conditions (R0) - (R4), and (R7).[10] But (D3) still suffers from the second drawback mentioned for (D1), i.e., the property that if P(A) = 1, then A ⫫ C for all C. In order to get around this problem, we need yet another amendment of the definition. This will be the topic of Section 4.

Furthermore, the following feature of (D3) is worth noticing. It is easy to verify that (D1) satisfies the following principle:

(8) A ℝ C iff C ℝ A

However, this symmetry principle is, in general, not satisfied by (D3). The following kind of example might be a counterexample to (8): Let A be the proposition that a mother is blond and C that her daugther is blond. Even though probability calculus tells us that P(A/C) ≠ P(A) if and only if P(C/A) ≠ P(C), our intuitions seem to be that A ℝ C but not C ℝ A since a mother's being blond can be a *cause* of her daughter's being blond, but not the other way around (cf. the results obtained by Tversky and Kahnemann 1982). I conclude that the fact that (D3) does not satisfy (8) need not be a drawback of the definition in relation to (D1). On the contrary, this feature may be in full accordance with our intuitions about relevance.

3. Belief Revision Models

The definitions (D1) - (D3) all have the drawback that if P(A) = 1, then A ⫫ C for all C. What one would like to have is that even if A is known, i.e., if P(A) = 1, C can be relevant to A for some sentences C. One way of capturing this idea is to say that *if A had not been known*, the information that C would have affected the probability of A. However, in order to formulate this idea more precisely, we need an account of belief revision and contraction processes.

In this section I will outline two models of belief revision. The first is based on belief sets as models of epistemic states, as developed in, for example, Alchourrón, Gärdenfors, and Makinson (1985) and Gärdenfors (1988). The second is based on probability functions as models of epistemic states. Models of this kind can be found in, for example, Harper (1975) and Gärdenfors (1986, 1988).

3.1 Belief Revision Models Based on Belief Sets

One way of modelling epistemic states is to describe them by *belief sets* which are sets of sentences from a given language. The interpretation is that if a sentence A belongs to a belief set K, this means that A is accepted as true in the state of belief modelled by K. Belief sets are assumed to be closed under logical consequences (classical

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

logic is generally presumed), which means that if K is a belief set and K logically entails B, then B is an element in K. A belief set can be seen as a partial description of the world — partial because in general there are sentences A such that neither A nor $\neg A$ are in K.

Belief sets model the statics of epistemic states. I now turn to their *dynamics*. What we need are methods for updating belief sets. Three kinds of updates will be discussed here:

(i) Expansion: A new sentence together with its logical consequences is *added* to a belief set K. The belief set that results from expanding K by a sentence A will be denoted K^+_A.

(ii) Revision: A new sentence that is *inconsistent* with a belief set K is added, but in order for the resulting belief set to be consistent, some of the old sentences of K are deleted. The result of revising K by a sentence A will be denoted K^*_A.

(iii) Contraction: Some sentence in K is retracted without adding any new beliefs. In order for the resulting belief set to be closed under logical consequences, some other sentences from K must be given up. The result of contracting K with respect to A will be denoted K^-_A.

Expansions of belief sets can be handled comparatively easily. K^+_A can simply be defined as the logical consequences of K together with A :

(Def +) $K^+A = \{B: K \cup \{A\} \vdash B \}$

As is easily shown, K^+_A defined in this way is closed under logical consequences and will be consistent when A is consistent with K.

It is not possible to give a similar explicit definition of revisions and contractions in logical and set-theoretical notions only. To see the problem for revisions, consider a belief set K that contains the sentences A, B, A & B $\rightarrow$ C and their logical consequences (among which is C). Suppose that we want to revise K by adding $\neg C$. Of course, C must be deleted from K when forming $K^*_{\neg C}$, but at least one of the sentences A, B, or A & B $\rightarrow$ C must also be given up in order to maintain consistency. There is no purely *logical* reason for making one choice rather than the other, but we have to rely on additional information about these sentences. Thus, from a logical point of view, there are several ways of specifying the revision of a belief set. What is needed here is a method of determining the revision.

As should easily be seen, the contraction process faces parallel problems. In fact, the problems of revision and contraction are closely related, being two sides of the same coin. To establish this more explicitly, we note, firstly, that a revision can be seen as a composition of a contraction and an expansion. Formally, in order to construct the revision K^*_A, one first contracts K with respect to $\neg A$ and then expands $K^-_{\neg A}$ by A which amounts to the following definition:

(Def *) $K^*_A = (K^-_{\neg A})^+_A$

Conversely, contractions can be defined in terms of revisions. The idea is that a sentence B is accepted in the contraction K^-_A if and only if B is accepted in both K and K^*_A. Formally:

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

(Def -) $K^-_A = K \cap K^*_{\neg A}$

These definitions indicate that revisions and contractions are *interchangable* and a method for explicitly constructing one of the processes would automatically yield a construction of the other.

There are two methods of attacking the problem of specifying revision and contraction operations. One is to present *rationality postulates* for the processes. Such postulates are introduced in Gärdenfors (1984), Alchourrón, Gärdenfors and Makinson (1985) and discussed extensively in Gärdenfors (1988), and they will not be repeated here. A guiding idea for these postulates is that changes should be *minimal*, so that when changing beliefs in response to new evidence, one should continue to believe as many of the old beliefs as possible.

The second method of solving the problems of revision and contraction is to adopt a more *constructive* approach and build computationally oriented *models* of belief revision that can take a belief set (or some representation of such a set) together with a sentences to be added as input and which then gives a revised belief set as output. One idea in this area is that the sentences that are accepted in a given belief set K have different degrees of *epistemic entrenchment*. When determining which sentences to delete in the revision, the basic recipe is that one gives up those with the lowest degrees of epistemic entrenchment and retains those with the highest degree (Cf. Gärdenfors and Makinson (1988) for this approach).

In the theory of belief revision, the rationality postulates and the model approach are connected via *representation theorems* which say that all models in a certain class (for example using epistemic entrenchment to determine the revision function) satisfy a certain set of postulates (for example, the postulates (K*1) - (K*8) for revision as presented in Gärdenfors (1988)), and vice versa, any revision method satisfying these postulates can be identified with one of the models in the given class.

3.2 Belief revision models based on probability functions

Two central dogmas of Bayesianism are that states of belief can be represented by *probability functions* and that rational changes of belief can be represented by *conditionalization* whenever the information to be added is consistent with the given state of belief. However, there are other kinds of changes of belief that cannot easily be modelled by the conditionalization process. Sometimes we have to revise our beliefs in the light of some evidence that contradicts what we had earlier mistakenly accepted. And when $P(A) = 0$, where P represents the present state of belief and A is the new evidence to be accommodated, the conditionalization process is undefined.

And sometimes we give up some of our beliefs. This kind of change of belief is here, like above, called a *contraction*, and the goal of this subsection is to present a way of modelling this process for a probabilistic model of a state of belief.

In parallel with the situation for belief sets, one can distinguish three kinds of probabilistic belief changes:

(i) *Expansion*: where we start from $P(A) = \alpha$ for some sentence A and some α, $0 < \alpha < 1$, and where the expanded probability function P^+_A satisfies the criterion $P^+_A(A) = 1$.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

(ii) *Contraction*: where we start from $P(A) = 1$ for some sentence A, and where the contracted function P^-_A satisfies the criterion $P^-_A(A) = \alpha$, for some α, $0 < \alpha < 1$.

(iii) *Revision*: where we start from $P(A) = 0$ for some sentence A, and where the revised probability function P^*_A satisfies the criterion $P^*_A(A) = 1$.

Expansions are normally modelled by conditionalization, i.e., $P^+_A(B) = P(B/A)$, for all B, but there is no similar explicit definition of the contraction and revision processes. However, in the same way as for belief sets, it is possible to define revisions of probability functions in terms of their contractions:

(Def P*) $P^*_A(B) = (P^-_{\neg A}(B))^+_A$

Or, using that expansion is modelled by conditionalization: $P^*_A(B) = P^-_{\neg A}(B/A)$ which is always well defined since $P^-_{\neg A}(A) > 0$ according to the postulate (P-1) below. This means that if we can give a satisfactory definition of probabilistic contraction, we have thereby also solved the problem of defining a probabilistic revision. So let us focus on probabilistic contractions.

When contracting a state of belief with respect to a belief A, it will be necessary to change the probability values of other beliefs as well in order to comply with the axioms of probability calculus. However, there are, in general several, ways of fulfilling these axioms. An important problem concerning contractions is how one determines which among the accepted beliefs, i.e., those A's where $P(A) = 1$, are to be retained and which are to be removed. One requirement for contractions of probability functions is that the *loss of information* should be kept as small as possible.

In Gärdenfors (1986) and (1988), I have formulated a number of postulates for contractions of probability functions. These postulates are based on the idea that the contraction P^-_A of a probability function P with respect to A should be *as small as possible* in order to minimize the number of beliefs that are retracted. In a sense that will be made more precise later, contractions can be viewed as 'backwards' conditionalizations. The postulates for probabilistic contractions can also be regarded as generalizations of a set of postulates for contractions in the case of belief sets.

I will here give a brief presentation of the postulates for probabilistic contraction. Formally, this process can be represented as a function from $\mathcal{P} \times \mathfrak{L}$ to $\mathcal{P}$, where $\mathcal{P}$ is the set of all probability functions and $\mathfrak{L}$ is the language that describes the space of events over which these functions are defined.[11] The value of such a contraction function, when applied to arguments P and A, will be called the contraction of P with respect to A, and it will be denoted P^-_A. Let us say that an event A is *accepted* in the state of belief represented by P iff $P(A) = 1$.

The first postulate is a requirement of 'success' simply requiring that A not be accepted in P^-_A, unless A is logically valid, in which case it can never be retracted:

(P-1) $P^-_A(A) < 1$ iff A is not logically valid.

It should be noted that this postulate does not say anything about the magnitude of $P^-_A(A)$. This leaves open a range of possibilities for an explicit construction of a contraction function. None of these possibilities will be ruled out by the remaining postulates. The value of $P^-_A(A)$ can be seen as a parameter in the construction of P^-_A.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

The second postulate requires that the contraction P^-_A is only dependent on the content of A, not on its linguistic formulation:

(P-2) If A and B are logically equivalent, i.e., describe the same event, then $P^-_A = P^-_B$.

The following postulate is only needed to cover the trivial case when A is not already accepted in P:

(P-3) If $P(A) < 1$, then $P^-_A = P$.

So far, the postulates have only stated some mild regularity conditions. The next one is more interesting:

(P-4) If $P(A) = 1$, then $P^-_A(B/A) = P(B)$, for all B in $\mathfrak{A}$.

This means that if A is first retracted from P and then added again (via conditionalization), then one is back in the original state of belief. This postulate, which will be called the *recovery* postulate, is one way of formulating the idea that the contraction of P with respect to A should be minimal in the sense that an unnecessary loss of information should be avoided. It also makes precise the sense in which contraction is 'backwards' conditionalization.

The final postulate is more complicated and concerns the connection between P^-_A and $P^-_{A\&B}$:

(P-5) If $P^-_{A\&B}(\neg A) > 0$, then $P^-_A(C/\neg A) = P^-_{A\&B}(C/\neg A)$, for all C.

In order to understand this postulate, we first present one of the arguments that has been proposed as a justification for conditionalization. Unlike all other changes of P to make A certain, conditionalization does not distort the probability ratios, equalities, and inequalities among sentences that imply A. In other words, the probability proportions among sentences that imply A are the same before and after conditionalization.

Now, if contraction may be regarded as 'backwards' conditionalization, then a similar argument should be applicable to this process as well. More precisely, when contracting P with respect to A, some sentences that imply $\neg A$ will receive non-zero probabilities, and when contracting P with respect to A&B some sentences that imply $\neg A$ or some sentences that imply $\neg B$ (or both) will receive non-zero probabilities. If, in the latter case, some sentences that imply $\neg A$ receive non-zero probabilities, i.e., if $P^-_{A\&B}(\neg A) > 0$, then the two contractions should give the same proportions of probabilities to the sentences implying $\neg A$, i.e., $P^-_A(C/\neg A)$ should be equal to $P^-_{A\&B}(C/\neg A)$, for all C. But this is exactly the content of (P-5).

This completes the set of postulates for probabilistic contraction functions. It should be noted that the postulates do not determine a unique contraction, but that they only introduce rationality constraints on such functions. Among other things, the value of $P^-_A(A)$ can be any number greater than 0 and smaller than 1. It is argued in Gärdenfors (1988) that rationality constraints are not enough to determine a unique contraction function (just as the probability axioms do not determine a unique rational probability function), but pragmatic factors must be added in order to single out the actual contraction. In the book, I introduce an ordering of 'epistemic entrenchment' among the beliefs to be used when determining which beliefs are to be given up when

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

forming a particular contraction. The heuristic rule is that when we have to give up some of our beliefs, we try to retain those with the greatest epistemic entrenchment.

4. A Final Definition of Relevance

It is now time to return to the desideratum (I) formulated in the introductory section:

(I) If a belief state K is revised by a sentence A, then all sentences in K that are *irrelevant* to the validity of A should be retained in the revised state of belief.

We can now use the terminology introduced above to formulate this criterion more precisely:

(I*) If $A \perp\!\!\!\perp C$, then $P^*_A(C) = P(C)$

When states of belief are modelled by belief sets, this criterion has as a special case:

(8) If $A \perp\!\!\!\perp C$, then $C \in K^*_A$ iff $C \in K$

Parallel criteria should also be valid for contraction:

(I-) If $A \perp\!\!\!\perp C$, then $P^-_A(C) = P(C)$

(9) If $A \perp\!\!\!\perp C$, then $C \in K^-_A$ iff $C \in K$

By using (Def P*), it is easy to show that (I*) will follow from (I-). So what we want is a definition of irrelevance that will satisfy (I-) in addition to (R0) - (R4) and (R7).

To arrive at such a definition, first note that it follows from the postulate (P-4) that if $P(A) = 1$, then $P(C) = P^-_A(C/A)$, so the requirement that $P^-_A(C) = P(C)$ may as well be written $P^-_A(C) = P^-_A(C/A)$ in this case. And according to (P-3), we have in the case when $P(A) < 1$, i.e., the case when A is not known in the state of belief represented by P, that $P(C) = P^-_A(C)$, for all C. This means that the equality $P^-_A(C) = P^-_A(C/A)$ is a generalized version of the classical equality $P(C) = P(C/A)$ used in (D1) to define irrelevance. The more general equality also covers the case when $P(A) = 1$.

Using this equality in combination with the construction in (D3), we can now formulate the final definition of irrelevance:

(D4) (a) $A \perp\!\!\!\perp C$ iff
$P(A) = 0$, or for all B such that $P^-_B(C) = P^-_B(C/B)$ and $P^-_A(A\&B) \neq 0$, it also holds that $P^-_A(C/A\&B) = P^-_A(C)$.

(b) $A \mathbb{R} C$ iff not $A \perp\!\!\!\perp C$.

In the same way as for (D3') we can conclude, by letting B be a tautology, that $A \perp\!\!\!\perp C$ entails as a special case $P^-_A(C/A) = P^-_A(C)$.

Before we establish the properties of (D4), we need to reconsider one of the general postulates for revision. (D3) had the drawback that if $P(A) = 1$, then $A \perp\!\!\!\perp C$ for all C. This feature made it possible to satisfy (R2) without any exceptions in the limiting cases. However, one of the purposes of formulating (D4) is to eliminate this drawback, but this means that when $P(A) = 0$ so that $A \perp\!\!\!\perp C$, it is still possible to have $\neg A \mathbb{R} C$.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

This violation of (R2) shows that it should be replaced by the following, slightly weaker version:

(R2') If $P(A) \neq 0$ and $P(A) \neq 1$, then $A \mathbb{R} C$ iff $\neg A \mathbb{R} C$.

The corresponding weakening of (1), which is equivalent to (R2') given (R1), is:

(1') If $P(A) \neq 0$ and $P(A) \neq 1$, then $A \perp\!\!\!\perp C$ iff $\neg A \perp\!\!\!\perp C$.

THEOREM 4: If the contraction function satisfies (P-1) - (P-5), then (D4) satisfies (R0), (R1), (R2'), (R3), (R4), (R7) and (I-).

Proof: (R0) and (R1) follow immediately from (D4). (R3) follows from (P-1) and (P-4) which entail that $P^-_{A \vee \neg A} = P$. To show (R2'), we proceed by verifying (1'). So assume that $P(A) \neq 0$, $P(A) \neq 1$, and $A \perp\!\!\!\perp C$: We want to show that $\neg A \perp\!\!\!\perp C$. By (P-3) we have $P^-_A = P = P^-_{\neg A}$. Suppose that for some B, $P^-_B(C) = P^-_B(C/B)$ and $P^-_{\neg A}(\neg A \& B) \neq 0$. We need to show that $P^-_{\neg A}(C/\neg A \& B) = P^-_{\neg A}(C)$, that is, $P(C/\neg A \& B) = P(C)$. If $P(A\&B) = 0$, then $P(C/\neg A \& B) = P(C/\neg A) = P(C)$, since $P(C/A) = P(C)$, and we are done. So suppose that $P(A\&B) = P^-_A(A\&B) \neq 0$. Since $A \perp\!\!\!\perp C$ we then know that $P(C/A\&B) = P(C)$. We need to distinguish two cases: (i) $P(B) < 1$. By (P-3) again, we then have $P^-_B = P$. Consider the following identities: $P(C) = P(C/B) = P(A/B) \cdot P(C/A\&B) + P(\neg A/B) \cdot P(C/\neg A \& B)$. Since $P(C/A\&B) = P(C)$, it follows from the fact that $P(A/B) + P(\neg A/B) = 1$ that $P(C/\neg A \& B) = P(C)$ which proves this case. (ii) $P(B) = 1$. Then $P(C/\neg A \& B) = P(C/\neg A)$. But, as established above $P(C/\neg A) = P(C)$ and we are done.

To prove (R4), it is sufficient to note that if C is contingent, then $P^-_C(C) < 1$ by (P-1), so $P^-_C(C) < P^-_C(C/C) = 1$ and hence $C \mathbb{R} C$.

The most difficult condition to verify is (R7). Assume that $A \perp\!\!\!\perp C$ and $B \perp\!\!\!\perp C$. The goal is to show $A\&B \perp\!\!\!\perp C$. If $P(A\&B) = 0$, this follows immediately from (D4). So suppose that $P(A\&B) \neq 0$. It follows that $P(A) \neq 0$ and $P(B) \neq 0$. From the facts that $A \perp\!\!\!\perp C$ and $B \perp\!\!\!\perp C$, we know that $P^-_A(C) = P^-_A(C/A)$ and $P^-_B(C) = P^-_B(C/B)$, and consequently that $P^-_A(C/A\&B) = P^-_A(C)$ and $P^-_B(C/A\&B) = P^-_B(C)$. As a preliminary we show that $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$.

Case 1: $P(A\&B) < 1$. It follows that either $P(A) < 1$ or $P(B) < 1$ and, from (P-3), that $P^-_{A\&B} = P$. If $P(A) < 1$, then also $P^-_A = P$, and since $P^-_A(C/A\&B) = P^-_A(C)$ it follows that $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$. Similarly, if $P(B) < 1$, then $P^-_B = P$ and since $P^-_B(C/A\&B) = P^-_B(C)$, we know also in this case that $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$. Hence, $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$.

Case 2: $P(A\&B) = 1$. It follows that $P(A) = 1$ and $P(B) = 1$. By (P-4), this entails that $P^-_{A\&B}(C/A\&B) = P(C)$, so it suffices to show that $P^-_{A\&B}(C) = P(C)$. Consider the following identity: $P^-_{A\&B}(C) = P^-_{A\&B}(A) \cdot P^-_{A\&B}(C/A) + P^-_{A\&B}(\neg A) \cdot P^-_{A\&B}(C/\neg A)$. It follows from (P-4) that $P^-_{A\&B}(C/A) = P^-_B(C)$ and thus from the assumption that $P^-_B(C) = P^-_B(C/B)$ and (P-4) again that $P^-_{A\&B}(C/A) = P(C)$. If $P^-_{A\&B}(\neg A) = 0$, the identity thus reduces to $P^-_{A\&B}(C) = P(C)$ which is what we wanted to show. On the other hand, if $P^-_{A\&B}(\neg A) \neq 0$, we can apply (P-5) to conclude that $P^-_{A\&B}(C/\neg A) = P^-_A(C/\neg A)$. But from the assumption that $P^-_A(C) = P^-_A(C/A)$ it follows that $P^-_A(C/\neg A) = P^-_A(C/A)$, which by (P-4) reduces to $P^-_A(C/\neg A) = P(C)$. So also in this case, the identity reduces to $P^-_{A\&B}(C) = P(C)$ and we have shown that $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$.

After this intermediate result, assume now that D is a sentence such that $P^-_{A\&B}(A\&B\&D) \neq 0$ and that $P^-_D(C) = P^-_D(C/D)$. Since $P^-_{A\&B}(C/A\&B) = P^-_{A\&B}(C)$

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

we can, by the same argument as above, conclude that $P^-_{A\&B\&D}(C/A\&B\&D) = P^-_{A\&B\&D}(C)$. We need to show that $P^-_{A\&B}(C/A\&B\&D) = P^-_{A\&B}(C)$ for all C. If $P(A\&B) = 1$, then this follows immediately from (P-4) and what was shown above. If $P(A\&B) < 1$, then $P(A\&B\&D) < 1$ and hence $P^-_{A\&B\&D} = P^-_{A\&B} = P$ and so $P^-_{A\&B}(C/A\&B\&D) = P^-_{A\&B\&D}(C/A\&B\&D) = P^-_{A\&B\&D}(C) = P^-_{A\&B}(C)$. This shows that (R7) is satisfied.

For (I-) finally, it is sufficient to note that if $P(A) = 1$, then $P^-_A(C) = P(C)$ follows from $P^-_A(C) = P^-_A(C/A)$ and (P-4), and if $P(A) < 1$, then $P^-_A(C) = P(C)$ is immediate from (P-3). This completes the proof of the theorem.

Apart from knowing that (D4) has the desired properties, we must also make sure that it does not lead to any triviality results. However, to prove this, one can use essentially the same example that was used in Gärdenfors (1978) to establish the non-triviality of (D3). The details are easy to verify but tedious so I will omit them.

5. Using Irrelevance in Constructions of Belief Revisions

The theorem proved in the preceeding section shows that if one starts from a probability contraction function, it is *possible* to define relations of relevance and irrelevance that satisfy the desired postulates. However, this procedure is like putting the cart in front of the horse, since it is more natural to take the irrelevance relation as *primitive* and then use this relation when *constructing* a belief revision function. In this section I will show how the notion of irrelevance can be exploited in such a construction. For simplicity, I shall only consider belief revision and contraction functions based on belief sets, but a similar approach can also be used for probabilistic revisions and contractions.

The key idea in the construction to follow is to use the irrelevance relation as a tool for *partitioning* the sentences in a belief set K:

(D≈) B *is relevance-equivalent to* C *in relation to* A, in symbols $B \approx_A C$, if and only if there is an E such that $E \perp\!\!\!\perp A$ and $E \rightarrow (B \leftrightarrow C)$.

The intuition behind this definition is that if the difference in the contents of B and C is irrelevant to A, then B and C should be treated equally when revising K with respect to A. We show that (D≈) indeed produces a partitioning of K:

THEOREM 5: If $\perp\!\!\!\perp$ satisfies (R0) - (R4) and (R7), then $\approx_A$ is an equivalence relation. In particular, if $B \perp\!\!\!\perp A$, then $B \approx_A T$. Furthermore, the mapping $B \mapsto |B|_A$, where $|B|_A$ is the equivalence class of B under $\approx_A$, is a Boolean homomorphism.

Proof: It is trivial that $\approx_A$ is reflexive and symmetric. We need to show that the relation is transitive as well. So assume that $B \approx_A C$ and $C \approx_A D$, that is, there are sentences E and F such that $E \perp\!\!\!\perp A$, $E \rightarrow (B \leftrightarrow C)$ and $F \perp\!\!\!\perp A$ and $\vdash F \rightarrow (C \leftrightarrow D)$. By (R7) it follows that $E\&F \perp\!\!\!\perp A$ and by propositional calculus it is easy to derive $\vdash E\&F \rightarrow (B \leftrightarrow D)$. Thus $B \approx_A D$, and we have shown that $\approx_A$ is an equivalence relation. Since $\vdash B \rightarrow (B \leftrightarrow T)$, it follows that if $B \perp\!\!\!\perp A$, then $B \approx_A T$.

To show that the mapping $B \mapsto |B|_A$ is a Boolean homomorphism, we need to show that equivalence classes are closed under negations and conjunctions.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

If $B \approx_A C$, then there is an E such that $E \perp\!\!\!\perp A$, $\vdash E \rightarrow (B \leftrightarrow C)$, and consequently, by propositional logic, $\vdash E \rightarrow (\neg B \leftrightarrow \neg C)$, and so $\neg B \approx_A \neg C$. If $B \approx_A C$ and $B' \approx_A C'$, there are sentences E and F such that $E \perp\!\!\!\perp A$, $\vdash E \rightarrow (B \leftrightarrow C)$, and $\vdash F \perp\!\!\!\perp A$ and $\vdash F \rightarrow (B' \leftrightarrow C')$. By (R7) it follows that $E\&F \perp\!\!\!\perp A$ and by standard propositional calculus it is easy to derive $\vdash E\&F \rightarrow (B\&B' \leftrightarrow C\&C')$ and hence $B\&B' \approx_A C\&C'$. This completes the proof.

If we define the Boolean operations on the equivalence classes in the obvious way, i.e., $\neg|B|_A = |\neg B|_A$ and $|B|_A \,\&\, |C|_A = |B\&C|_A$ etc., it follows from the theorem above that the set $|K|$ of equivalence classes will form a belief set, where all sentences that are irrelevant to A belong to the same equivalence class, i.e., $|T|_A$. This class will function as the tautology in $|K|$. If A is contingent, there are at least two distinct equivalence classes in $|K|$ because we then have $|T|_A \neq |A|_A$.

If we assume that we have a belief contraction (or revision) function defined on $|K|$, then we can define a contraction (or revision) function on K in the following way

(D-) $\quad B \in K^-_A$ iff $|B| \in |K|^-_A$.

The value of this construction is shown by the fact that if the contraction function on $|K|$ satisfies (K-1) - (K-8) of Gärdenfors (1988), then the contraction function defined on K also satisfies (K-1) - (K-8) *as well as* (I-). The upshot is that if we have an irrelevance relation that satisfies the desired conditions defined on a belief set K, then we can, by the method presented here, construct a well-behaved contraction function for K that will satisfy (I-).

One important application area of belief revision theory is *updating logical databases*. From a computational point of view, the ultimate goal is to develop algorithms for computing appropriate revision and contraction functions for an arbitrary logical database (which models a state of belief). The proposed method will simplify the computations of belief revision functions since the number of equivalence classes, in all interesting cases, will be considerably smaller than the number of elements in the original belief set. Thus the amount of calculations required to compute the required belief revision will be reduced.

6. Conclusion

A general criterion for the theory of belief revision is that when we revise a state of belief by a sentence A, as much of the old information as possible should be retained in the revised state of belief. The motivating idea in this paper has been that if a belief B is irrelevant to A, then the general criterion entails that B should still be believed in the revised state. The problem was that the traditional definition of statistical relevance suffers from some serious shortcomings and cannot be used as a tool for defining belief revision processes. This led me to develop an amended notion of relevance that has the desired properties. In particular, the postulate (R7), is violated by the traditional definition. On the basis of the new definition, I have outlined how it can be used to simplify a construction of a belief revision method.

https://doi.org/10.1086/psaprocbienmeetp.1990.2.193079 Published online by Cambridge University Press

Notes

[1]Research for this article has been supported by the Swedish Council for Research in the Humanities and Social Sciences. I wish to thank Didier Dubois, David Miller, Henri Prade, Teddy Seidenfeld, and Wolfgang Spohn for helpful comments.

[2]Throughout this paper I will use probability measures defined on sentences. It is easy to translate the analysis presented here to probability measures defined on classes (or properties). 'Relevance' defined in terms of classes is a central concept in Salmon's and Greeno's theories of statistical explanation (cf. Salmon, 1971).

[3]Carnap (1950), p. 348.

[4]Keynes (1921), p.72.

[5]Keynes (1921), p. 55. I have changed Keynes's notation. Keynes gives no explicit definition of 'relevant' although (b) is obviously in line with his intentions.

[6]Carnap remarks in (1950) , p. 420, that this definition is essentially the same as his definition of 'complete irrelevance' with the condition added that A not be logically false.

[7]Carnap(1950), p. 420.

[8]Keynes(1921), p. 55.

[9]In order to avoid contradicting (2), it must be assumed that A & B is consistent.

[10]Such an example can be found on pp. 362-63 in Gärdenfors (1978).

[11]I will still assume that $\mathfrak{L}$ includes the standard propositional operators and is ruled by a logic including classical propositional logic.

References

Alchourrón, C.E., Gärdenfors P., and Makinson D. (1985), "On the logic of theory change: Partial meet contraction and revision functions", *The Journal of Symbolic Logic* 50: 510-30.

Carnap, R. (1950), *Logical Foundations of Probability*, Chicago: University of Chicago Press.

David, A.P. (1979), "Conditional independence in statistical theory", *Journal of the Royal Statistical Society, B*, 41, 1-31.

Geiger, D. (1990), "Graphoids: A qualitative framework for probabilistic inference", *Technical Report R-142*, Cognitive Systems Laboratory, UCLA.

Gärdenfors, P. (1978), "On the logic of relevance", *Synthese* 37: 351-67.

_ _ _ _ _ _ _ .. (1984), "Epistemic importance and minimal changes of belief", *Australasian Journal of Philosophy* 62, 136-57.

_ _ _ _ _ _ _ .. (1988), "The dynamics of belief: Contractions and revisions of probability functions", *Topoi* 5: 29-37.

_ _ _ _ _ _ _ . (1988), *Knowledge in Flux*, Cambridge, MA: MIT Press.

Gärdenfors, P. and D. Makinson (1988), "Revisions of knowledge systems using epistemic entrenchment", in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, M. Vardi (ed.). Los Altos, CA: Morgan Kaufmann.

Harper, W. L. (1975), "Rational belief change, Popper functions and counterfactuals", *Synthese* 30: 221-62.

Keynes, J. M. (1921), *A Treatise on Probability*, London: Macmillan.

Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.

Salmon, W. C. *et al.* (1971), *Statistical Explanation and Statistical Relevance*, Pittsburgh: Pittsburgh University Press.

Salmon, W. C. (1975), "Confirmation and relevance", in *Induction, Probability, and Confirmation*, Minnesota Studies in the Philosophy of Science, G. Maxwell and R.M. Anderson Jr. (eds.), Minneapolis: University of Minnesota Press, pp. 3-36.

Tversky, A. and Kahneman, D. (1982), "Causal schemas in judgments under uncertainty", in *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky (eds.). New York: Cambridge University Press.