David C. Makinson

Conditional probability in the light of qualitative belief change

Article (Accepted version)
(Refereed)

Conditional Probability in the Light of Qualitative Belief Change

David Makinson

Abstract

We explore ways in which purely qualitative belief change in the AGM tradition throws light on options in the treatment of conditional probability. First, by helping see why it can be useful to go beyond the ratio rule defining conditional from one-place probability. Second, by clarifying what is at stake in different ways of doing that. Third, by suggesting novel forms of conditional probability corresponding to familiar variants of qualitative belief change, and conversely. Likewise, we explain how recent work on the qualitative part of probabilistic inference leads to a very broad class of 'proto-probability' functions.

Key words: conditional probability, belief revision, ratio rule, AGM, Hosiasson-Lindenbaum, Kolmogorov, Popper, Rényi, van Fraassen, cores, screened revision, hyper-revisionary probability, proto-probability, conditional plausibility measures.

1. Why Go Beyond the Ratio Rule?

Kolmogorov's axioms for one-place probability functions are simple and easy to work with, and the associated ratio definition of conditional probability is convenient to use (see appendix). They have become standard. So why go beyond them?

The reasons advanced in the literature are of two main kinds: a metaphysical complaint and a pragmatic appeal for greater expressiveness. We outline them in this section, and suggest that while the metaphysical grounds are less than compelling, there is indeed a need for greater expressive capacity. In the following section, we show how a comparison with the situation in qualitative belief revision makes that need all the more evident.

To keep the main text reader-friendly, most of the verifications and historical remarks are placed in an extended appendix, whose sections run parallel to the main text.

1.1. Doctrinal *vs* Pragmatic Considerations

It is commonly felt (see appendix) that all probability is 'really' conditional anyway, and we should bring this out by integrating it into our formal treatment. From a subjective perspective: a probability judgement is always made given a whole lot of background information, and so is in some sense conditional on that information. From a frequency standpoint: probability is some sort of limiting frequency of a type of item in a reference set, and if we enlarge or diminish the set, the frequency will in general change.

However, this perspective has its limitations. Considered as an argument, it may involve an infinite regress, as is most easily seen in the field-of-sets mode. Suppose we do take probability as a two-place function p: $F^2 \rightarrow [0,1]$ where F is a field of subsets of a set S. This still depends on the choice of the underlying set S. Turning p

into a three-place function p: $F^3 \rightarrow [0,1]$ will not help, as F^3 still depends on S, taking us one step further in an infinite regress. The only way to eliminate all such dependence is to fix the domain as the *universal* class. But practising probabilists never do this and, if done, it might as well be done from the beginning, with one-place functions.

Historically, the perspective is reminiscent of an early way of looking at classical first-order logic, according to which universal quantifications $\forall x \varphi(x)$ are at bottom always conditional, since their range depends on the choice of domain of discourse. On this view, the dependency should be made explicit from the outset by always quantifying over the entire universe, rewriting $\forall x \varphi(x)$ as $\forall x[Dx \rightarrow \varphi(x)]$ where D is the intended domain. Such a view had some philosophical currency for a while despite the difficulties of talking about a universal set (so the universe was thought of as a class rather than a set). But we have become accustomed to working with the simpler mode of representing universal quantification without running into difficulty, and the philosophical worries have simply withered away.

The historical precedent carries a methodological lesson. Even if all quantification or probability can be said to be in some sense conditional, this does not imply that the conditionality should always be brought into the formalism of the theory itself. It may sometimes be better treated as part of the business of applying the theory to specific problems.

Thus, it would seem that the doctrinal or metaphysical reasons for always taking conditional probability as primitive are less than compelling. Nevertheless, an important consideration remains. When conditional probability is defined by the ratio rule, it has limited expressive capacity. Sometimes we would like to allow propositions that have been accorded zero probability to serve as conditions for the probability of other propositions. This is impossible when $p(x|a)$ is understood as $p(a \wedge x)/p(a)$, for it is undefined when $p(a) = 0$.

The most famous example of this expressive gap is due to Borel. Suppose a point is selected at random from the surface of the earth. What is the probability that it lies in the western hemisphere, given that it lies on the equator? The condition of lying (exactly) on the equator has probability 0 under the random selection, but we would be inclined to regard the question as meaningful and even as having 1/2 for its answer. Examples have also arisen in the course of investigations in game theory in connection with strategic reasoning and weak dominance; for references see Halpern (to appear).

This complaint is more modest than the doctrinal claim, pointing to a gap rather than alleging a defect. It suggests that it could be helpful to have a more general conception of conditional probability that covers what we will call *the critical zone* – the case where the condition a is consistent but of zero probability – and that we should try to articulate it.

As remarked by e.g. Rényi, in mathematical practice one can sometimes 'work around' the problem. The idea is that when a is in the critical zone, we could take $p(x|a)$ to be the limit of the values of $p(x|a')$ for a suitable infinite sequence of non-critical approximations a' to a. This is natural for some examples, such as Borel's

hemisphere/equator one. However, it is possible only for suitable domains – notably fields based directly or indirectly on the real numbers – satisfying appropriate conditions. Moreover, the outcome will depend on our choice of the approximating sequence. In the hemisphere/equator problem, we get the answer 0.5 only if each of the approximating 'equators' has constant width around the globe. If each is thicker in the west than in the east, then the figure will be higher. So the procedure provides neither a general solution nor, within its domain, a unique one.

This is not to dismiss Rényi's way around the problem out of hand. In practical situations, it may often be the best thing to do. Suppose that in an empirical investigation we have been working extensively with a particular one-place probability function, and we unexpectedly find ourselves needing to conditionalize on a proposition to which it accorded value zero. Should we go back and reconstruct everything in terms of an intrinsically two-place function? To do so poses two difficulties. In the first place, we need to specify, in a principled manner, the behaviour of the two-place function over the critical zone. We may find that there is more arbitrariness in the decisions required there than in choosing a particular approximating sequence – especially so if, as in the case of the equator example, there is a sequence that suggests itself quite naturally. Once the new two-place probability function has been specified there remains the job of rewriting, in terms of it, all the work so far done in the empirical investigation, and checking that it continues to run. In such circumstances, the simplest thing to do may often be to follow Rényi's work-around.

Let us return, however, to the theoretical level. How should a two-place probability function behave over the critical zone? There are, of course, quite trivial ways of regulating it. One, due to Carnap 1950, is to declare that the zone is empty: whenever $p(x) = 0$ then x is inconsistent. This is sometimes known as the regularity condition. It has the immediate effect that the *ratio definition* of $p(x|a)$ as $p(a \wedge x)/p(a)$ covers all instances of the right argument a except when a is inconsistent. For inconsistent a, one can then either leave $p(x|a)$ undefined, or take it to have value 1 for all values of the left argument x. However, as remarked e.g. by Spohn 1986, this is more like a way of avoiding than solving the problem. It abolishes by fiat the distinction between logical impossibility and total improbability.

Moreover, as noted by Harper 1975 (page 229), Carnap's restriction creates an internal inelegance: the set of functions is not closed under left projection of conditionalization, i.e. in Bayesian terminology, under update. To see this, let p be a proper one-place Kolmogorov function satisfying Carnap's regularity condition, and consider the two-place function $p(\cdot|\cdot)$ determined by the ratio definition. Now take a contingent proposition a with $1 \neq p(a) \neq 0$, and form the left projection $p_a(\cdot)$ alias $p(\cdot|a)$ of the two-place function. By the definition of left projections (see appendix) we have $p_a(x) = p(x|a)$ so substituting $\neg a$ for x, we have $p_a(\neg a) = p(\neg a|a) = p(\neg a \wedge a)/p(a) = 0/p(a) = 0$ since $p(a) > 0$. Thus $p_a(\neg a) = 0$ even though $\neg a$ is consistent, violating the regularity condition as applied to p_a. Even when p satisfies the regularity condition, the left projection of its conditionalization under the ratio definition (briefly, its update) need not do so.

Another trivial way of covering the critical zone is to put $p(x|a) = 1$ for every value of x when $p(a) = 0$. This might be called the *ratio/unit definition* of conditional

probability. But while this renders the function always-defined, and is very convenient in many contexts, it does not do much to increase expressive power since it makes $p(x|a) = p(y|a) = 1$ whenever the condition a is in the critical zone. Hopefully we should be able to get something more discriminating; the two-place function should in some sense be essentially conditional.

1.2. Some Notational Niceties

In the following sections, we compare various options for axiomatizing conditional probability in the light of qualitative belief revision. When doing so, we follow certain notational conventions for clarity. In particular, we distinguish $p(x|a)$ from $p(x,a)$, writing:

- $p(x|a)$ with a bar when it is understood as a two-place operation defined from a one-place one by the ratio rule, i.e. by putting $p(x|a) = p(a \wedge x)/p(a)$ when $p(a) > 0$, possibly with the extension that puts $p(x|a) = 1$ when $p(a) = 0$ (in which case we call it the ratio/unit rule).

- $p(x,a)$ with a comma when taking p as an undefined (arbitrary or primitive) two-place operation defined over all or part of L^2.

Care will always be taken to specify the arity (number of places) of a function under consideration, either by mentioning it explicitly, or by using place-markers as in $p(\cdot)$, $p(\cdot|\cdot)$, $p(\cdot,\cdot)$.

Throughout, Cn is the operation of classical consequence; we also write $\approx$ for the relation of classical equivalence.

2. Exploring the Critical Zone

In this section we weigh the significance of the critical zone. We begin by observing that an analogous zone already arises on the qualitative level for AGM belief change, and explaining how this helps bring out the conceptual options underlying different systems for two-place conditional probability. We then review those systems, presenting them in a modular way that makes manifest the intuitive rationales for apparently technical choices.

2.1. A Leaf from the AGM Book

It is instructive to compare the situation for probability change with that for qualitative belief change in the AGM tradition initiated in Alchourrón, Gärdenfors and Makinson 1985.

There, expansion is one thing, revision another. Let K be any belief set, i.e. a set of propositions closed under the operation Cn of classical consequence, i.e. $K = Cn(K)$. The *expansion* of K by a is defined simply by putting $K+a = Cn(K \cup \{a\})$. However *revision* is defined by putting $K*a = Cn((K \dot{-} \neg a) \cup \{a\})$, where $\dot{-}$ is a suitable contraction operation forming from K a subset that no longer implies the item contracted (when it is not itself logically true), and satisfying certain regularity conditions.

We thus have two different kinds of change side by side. Again, they differ in the *critical zone* which, in this qualitative context, is the case where we modify the belief set *K* by a proposition *a* that is *itself consistent but inconsistent with K*. In this critical zone, expansion creates blow-out to the set of all propositions of the language, while revision forces removal of items from the belief set. Outside the critical zone, the two operations coincide. This basic difference should not be obscured by talk of expansion being a special case of revision. That is just a sloppy way of saying that the values of the two operations are the same outside the critical zone; neither operation is a special case of the other.

This basic conceptual difference reflects itself in the different formal properties of expansion and revision. There are principles that hold for expansion but not for revision, and conversely. In particular:

- Expansion never loses anything from the initial belief set, i.e. $K \subseteq K+a$. This is sometimes known as the principle of *belief preservation*. In contrast, revision eliminates material from the belief set whenever the input *a* is in the critical zone.

- When *a* is inconsistent with *K*, expansion gives us blow-out: both $a, \neg a \in K+a = Cn(K \cup a) = L$ (the whole language). In contrast for revision, even when *a* is inconsistent with *K* then, as long as *a* is itself consistent, so is $K*a$. This property of revision is known as the principle of *(input) consistency preservation*.

The pattern is replicated in the probabilistic context. There too we are looking at two different kinds of operation, which coincide outside but differ inside the critical zone – which in this context, we recall, is the case where *a* is consistent but $p(a) = 0$. One is *expansionary*, the other is *revisionary*.

- The expansionary operation is given by the ratio/unit definition. It satisfies a probabilistic analogue of qualitative belief preservation: $p(x|a) = 1$ whenever $p(x) = p(x|\top) = 1$. Expressed with left projections, $p_a(x) = 1$ whenever $p_\top(x) = 1$. In other words, conditionalizing never reduces the corresponding belief set: writing $B(p)$ for $\{x: p(x) = 1\}$ we always have $B(p) \subseteq B(p_a) = \{x: p_a(x) = 1\} = \{x: p(x|a) = 1\}$; see the appendix for detailed verification. No juice is lost. In contrast, a revisionary operation would allow for loss of material from the associated belief set.

- When $p(a) = 0$, the expansionary operation blows-out to the unit function (irrespective of *a*'s own consistency): in that case $p_a(x) = p(x|a) = 1$ for all *x*, so that $B(p_a) = L$; see the appendix for a full verification. In contrast, a revisionary conditional probability function would never give us the unit function when the condition *a* is itself consistent.

These two kinds of conditionalization should not be thought of as competing for the position of 'the correct one'. Like expansion and revision in the qualitative context, they can work side by side, as different kinds of conditionalization. But how can the revisionary conception best be expressed?

There are two main approaches to the problem. One is to define a family of revision operations that take one-place probability functions to others. That is the path taken by Gärdenfors in a pioneering paper of 1986 (integrated into his book of 1988). The other approach is to define a family of two-place probability functions. That is the direction followed in varying manners by Hosiasson-Lindenbaum 1940, Rényi 1955, 1970, 1970a, Popper 1959 and others in their wake.

Although different in appearance, the two approaches are intimately related – indeed at bottom the same – as hinted by Gärdenfors 1988 and observed explicitly by Lindström and Rabinowicz 1989. Here, we consider only the approach using two-place probability functions. Our initial questions are: What are the essential conceptual differences between the differing axiom systems for two-place probability, and what are their advantages and disadvantages?

2.2. Bird's-Eye View of Available Systems

The usual presentations of axiom systems for two-place probability functions can be quite confusing. The systems are not always formulated in an intuitively evident manner. They can also be difficult to compare due to differing choices of right domain – sometimes the whole of L, sometimes the consistent propositions in L, sometimes an arbitrary subset of L lying between $\{x: p(x,\top) > 0\}$ and L itself. To facilitate comparison and focus on essentials, we formulate all systems as functions defined with unrestricted right domain and thus on *the whole* of L^2. We also present the systems in a *modular way*, that is, with a common basis and differing in what is added to it.

The leading idea is to exploit Rényi's insight that for 'most' values of the right argument of the two-place function, the left projections should be proper one-place Kolmogorov functions, adding that in the remaining cases they should be the unit function. We obtain modularity by making a different specification of what counts as 'most' for each system.

We begin with the basic *van Fraassen system*, which was formulated in the field-of-sets mode by van Fraassen 1976 and 1995. Expressed in the propositional mode for two-place functions p: $L^2 \rightarrow [0,1]$, its axioms are the following three of *right extensionality*, *left projection*, and *product*:

(vF1) $p(x,a) = p(x,a')$ whenever $a \approx a'$

(vF2) p_a is a one-place Kolmogorov probability function with $p_a(a) = 1$

(vF3) $p(x \wedge y, a) = p(x,a) \cdot p(y, a \wedge x)$ for all formulae a, x, y.

In (vF1), recall that we are using $\approx$ for classical equivalence. Note that (vF2), as formulated here, says that p_a is a one-place Kolmogorov function, but it *does not say whether it is proper or improper* (the unit function). Indeed, the axioms are consistent with p_a being the unit function for every $a \in L$.

Despite their modesty, the van Fraassen axioms have surprisingly many useful consequences. The following were already noticed by van Fraassen 1976, 1995, Arló

Costa 2001, Arló Costa and Parikh 2005. For the convenience of the reader, we recall brief verifications in the appendix.

- Left extensionality: $p(x,a) = p(x',a)$ whenever $x \approx x'$.

- When $y \in Cn(x)$ then $p(x,a) \leq p(y,a)$.

- When $p(\cdot)$ is defined as $p(\cdot,\mathsf{T})$, then we have the ratio rule (though not its unit extension to the critical zone, i.e. the ratio/unit rule).

- When a is a contradiction, then p_a is the unit function.

- The set Δ of all $a \in L$ such that p_a is the unit function is an ideal. That is, it is closed downwards (whenever $a \in Cn(b)$ and $a \in \Delta$ then $b \in \Delta$) and also closed under disjunction (whenever $a,b \in \Delta$ then $a \vee b \in \Delta$).

- p_a is the unit function iff $p(a,b) = 0$ for all b such that p_b is a proper Kolmogorov function.

Van Fraassen 1976, 1995 called the $a \in L$ such that p_a is a proper Kolmogorov function *normal*, and the remaining $a \in L$ *abnormal* – of course, modulo the function $p(\cdot,\cdot)$. In that terminology, the set of all abnormal formulae form a non-empty ideal containing the contradictions, and a formula a is abnormal iff $p(a,b) = 0$ for all normal b. Apart from that, the van Fraassen axioms do not tell us much about which formulae are normal, which abnormal.

Popper's system goes some way to filling the gap. It may be obtained by adding a single axiom, stating that p_a is normal whenever $p(a,\mathsf{T}) > 0$.

(Positive): when $p(a,\mathsf{T}) > 0$ then p_a is a proper Kolmogorov function.

This still leaves unspecified the status of p_a when a is in the critical zone, i.e. consistent but with $p(a,\mathsf{T}) = 0$. The other systems fill this gap in three different ways. *Carnap's system* does so trivially, by declaring that the zone is empty:

(Carnap) When a is consistent then $p(a,\mathsf{T}) > 0$.

This is equivalent to what we would get by staying with one-place functions as primitive, using the ratio/unit definition to generate two-place functions, but declaring that only contradictions can get the value 0.

The *Unit system* fills the gap almost as trivially, by adding instead an axiom saying that any left projection from a point in the critical zone has constant value 1:

(Unit) When a is consistent but $p(a,\mathsf{T}) = 0$, then p_a is the unit function.

This is equivalent to what we would get by keeping one-place functions as primitive and using the ratio/unit definition to generate two-place ones, without requiring that only contradictions can get the value 0.

Hosiasson-Lindenbaum's system (briefly HL) regulates the critical zone by treating its elements just like consistent propositions outside the zone. It adds to the Popper axioms:

(HL) When a is consistent but $p(a,\top) = 0$, then p_a is a proper Kolmogorov probability function.

Thus, in terms of Rényi's leading idea mentioned above, 'most values of the right argument' means, for an arbitrary $p(\cdot,\cdot)$:

- In the Hosiasson-Lindenbaum system: all propositions above or in the critical zone,

- In the Unit system: all propositions above the critical zone but none of those in it,

- In the Popper system: all propositions above the critical zone plus those in an unspecified subset (possibly empty) of it,

- In Carnap's system: any of the first three, since the critical zone is declared empty.

For the van Fraassen system, the content of 'most values of the right argument' is a little more complex and we return to it in a moment.

It is easy to check that these axiom systems are equivalent to their usual presentations (see appendix), giving us the sets **Carnap**, **Unit**, **HL**, **Popper, van Fraassen** of functions. The modular arrangement makes it clear at a glance, from their very formulation, what the relations between the systems are. Specifically, we have **Carnap = Unit$\cap$HL $\subset$ Unit, HL $\subset$ Unit$\cup$HL $\subset$ Popper $\subset$ Popper$\cup\{\mathbf{1}(\cdot,\cdot)\}$ = van Fraassen**, where $1(\cdot,\cdot)$ is the unit two-place function putting $p(x,a) = 1$ for all a,x, and $\subset$ is proper inclusion.

The first four relations were established by Leblanc and Roeper (1989 theorems 4 and 15, table 5, figure 15; also 1999 chapter 3 section 2), with however rather laborious verifications from the usual formulations of the systems, and without mentioning the historical role of Hosiasson-Lindenbaum as a key contributor. With the present modular formulation, the inter-relations become trivial, except for the inclusion **van Fraassen $\subseteq$ Popper$\cup\{\mathbf{1}(\cdot,\cdot)\}$** and the proper part of the inclusion **Unit$\cup$HL $\subset$ Popper**. We comment on these in turn.

The inclusion **van Fraassen $\subseteq$ Popper$\cup\{\mathbf{1}(\cdot,\cdot)\}$** amounts to observing that Popper's system may be obtained from that of van Frassen by adding an axiom saying that $p(\cdot,\cdot)$ is not the unit two-place function, i.e. that $p(x,b) \neq 1$ for some x,b. This is known from the work of van Fraassen 1976, 1995, but for convenience we give a brief verification in the appendix.

Since **van Fraassen = Popper$\cup\{\mathbf{1}(\cdot,\cdot)\}$**, the two classes differ by only a single function – the two-place unit function. The relation between the systems of Popper and van Fraassen is thus analogous to that between the original system of Kolmogorov for proper one-place probability functions, and the extension obtained by adding the improper (unit) one-place function. Further, we can locate the van Fraassen

system in the framework of Rényi's intuitive idea of 'most values of the right argument' as follows:

- In the van Fraassen system, 'most' means: *either* none at all (in the case of the two-place unit function), *or* (for all other functions $p(\cdot,\cdot)$) all propositions above the critical zone plus those in an unspecified subset of it (like Popper).

For the proper part of the inclusion **Unit$\cup$HL $\subset$ Popper**, we need a 'mixed' function, failing axioms (Unit) and (HL) but satisfying the Popper axioms. Such a function was already supplied by Leblanc and Roeper 1989 in the form of a rather enigmatic 64-element table; in the appendix we equip the same example with an intuitive rule-based formulation. The relations between the classes are pictured in the diagram of Figure 1 below.

The reader may be surprised that we have not mentioned the axiomatic system of Rényi 1955, also in his later books 1970, 1970a. This is not neglect: Rényi's work is indeed capital, providing the leading idea on which most subsequent presentations (including the present one) are based. Rather, his system takes a form rather different from those above. He presents a *scheme for a range of axiomatizations*, with the right domain of the function serving as a parameter. For a suitable choice of this parameter (and a little massage) we may obtain the axiomatization of Popper, and likewise of Hosiasson-Lindenbaum. Thus, strictly speaking (and taking into account the chronology), Popper's axioms could be called the Rényi/Popper postulates. These historical matters are reviewed more fully in the appendix.

Figure 1. Hasse Diagram for Classes of Two-Place Probability Functions
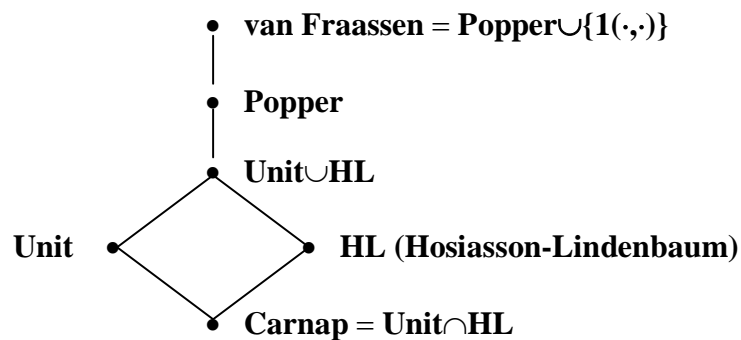
3. Comparative Attractions

Are there any reasons for preferring one of these systems to another? From our discussion so far, there are three serious contenders going beyond the ratio/unit account, namely the systems of Hosiasson-Lindenbaum, Popper, and van Fraassen. In this section we discuss possible grounds for preferring one to the other, coming to the conclusion that the choice is not a matter of correctness but of policy, particularly regarding two questions. The first is conceptual: *how revisionary* do we want our conditional probability to be? This separates Hosiasson-Lindenbaum on the one hand from Popper and van Fraassen on the other. The second question is more technical: do we want the class of functions to be *closed under Bayesian update*? Its answer groups

Hosiasson-Lindenbaum and Popper in contrast with van Fraassen. Despite these differences, there is an overall conceptual unity: we show that the two broader classes of functions may be transformed into the narrowest one by passing from the classical to suitably chosen supraclasical background consequence relations.

3.1. Hosiasson-Lindenbaum *vs* Popper *vs* van Fraassen

The Hosiasson-Lindenbaum system is not just revisionary − it is radically so, satisfying without reserve the probabilistic counterpart of consistency preservation. That is, for *every* proposition *a*, if it is consistent then p_a is a proper Kolmogorov function. The only values of the right argument that project to the unit function are the inconsistent ones.

On the other hand Popper's system is more compromising. Its spirit was expressed by Leblanc 1989, who asked: "Can't there be *some* statement of *L* that is 'utterly unbelievable', so unbelievable indeed that − should you believe it − you'd believe anything, and yet is not truth-functionally false?". It is 'variably revisionary', in that it leaves unspecified the extent to which a function satisfying the axioms is expansionary, and how far it is revisionary. As one extremal case it covers functions $p(\cdot,\cdot)$ that are purely expansionary, i.e. p_a blows out to the unit function for every *a* in the critical zone as well as for inconsistent *a*. These are the functions satisfying the Unit axiom above. At the other extreme it covers the Hosiasson-Lindenbaum functions, where p_a never blows out in the critical zone. In between, it covers many 'mixed' functions, where for certain *a*,*b* in the critical zone p_b is the unit function while p_a is a proper Kolmogorov function. Van Fraassen's system is also variably revisionary, but covers just one more function than does Popper's: the two-place unit function $p(\cdot,\cdot) = 1(\cdot,\cdot)$, for which p_a is the one-place unit function for any choice of *a* whatsoever in the right domain.

Thus, if we a looking for a notion of conditional probability that is as revisionary as possible, we will naturally turn to the Hosiasson-Lindenbaum functions; if we wish to allow variation in the extent to which it is revisionary, we will favour the Popper or van Fraassen functions.

On a more technical level, of the three classes of functions, that of van Fraassen is the only one that is closed under Bayesian update. The point is very similar to that made by Harper 1975 regarding Carnap's regularity condition (see section 1.1 above), and may be expressed as follows.

As well as passing from an unconditional to a conditional function, we often need to strengthen the condition of an already conditional one. It is useful to express this as an operation taking a two-place function $p(\cdot,\cdot)$ to another two-place function $p_{\wedge b}(\cdot,\cdot)$ by the rule $p_{\wedge b}(x,a) = p(x,a\wedge b)$. This is a familiar move in the Bayesian tradition, where it is called *update*. But the operation breaks the boundaries of the class of Hosiasson-Lindenbaum functions. It may happen that while *a* is consistent, *a*∧*b* is not, in which case for any function $p(\cdot,\cdot)$ satisfying the Hosiasson-Lindenbaum axioms, $(p_{\wedge b})_a$ is the unit function despite the consistency of *a*, so that $p_{\wedge b}$ does not satisfy axiom (HL).

Indeed, the update operation also breaks the boundaries of the class of Popper functions. To see this, consider any Popper function $p(\cdot,\cdot)$, and let *a* be an inconsistent

proposition. Then $p_{\wedge a}(a,\top) = p(a,a) = 1 > 0$, while for all values of x we have $(p_{\wedge a})_a(x) = p_{\wedge a}(x,a) = p(x,a) = 1$ since a is inconsistent, so that $(p_{\wedge a})_a$ is the unit function. These two facts together contradict the distinctive Popper axiom (Positive).

The only way to keep our class of functions closed under Bayesian update is to generalize to the class of all van Fraassen functions. Thus, if we regard closure under update as important or convenient, we will thus naturally gravitate towards the van Fraassen system. On the other hand, it may be suggested that from the point of view of a revisionary concept of conditional probability, Bayesian update as defined above is quite inappropriate, suitable only for an expansionary notion with conditional probability given by the ratio or ratio/unit definition. This is discussed further in the appendix.

On the other hand, the gap between the three classes of function may be less significant than appears at first sight, for every van Fraassen function may be transformed into a Hosiasson-Lindenbaum one by suitably expanding the underlying consequence relation.

To see this, consider any two-place function $p(\cdot,\cdot)$ satisfying the van Fraassen axioms. We have already noted (section 2.2) that the set Δ of all $a \in L$ such that p_a is the unit function is a non-empty ideal. That is, it contains all contradictions, whenever $a \in Cn(b)$ and $a \in \Delta$ then $b \in \Delta$, and whenever $a,b \in \Delta$ then $a \vee b \in \Delta$. Hence the set $\nabla = \{a: \neg a \in \Delta\}$ is a filter, i.e. whenever $b \in Cn(a)$ and $a \in \nabla$ then $b \in \nabla$, and whenever $a,b \in \nabla$ then $a \wedge b \in \nabla$). From this in turn it follows that if we define a supraclassical consequence operation Cn' by putting $Cn'(A) = Cn(A \cup \nabla)$ we have: $\bot \in Cn'(a)$ iff $\bot \in Cn(\{a\} \cup \nabla)$ iff $\neg a \in Cn(\nabla) = \nabla$ iff $a \in \Delta$ iff p_a is the unit function. That is, p_a is the unit function iff a is inconsistent modulo Cn'. Using this, it is not difficult to show that if $p(\cdot,\cdot)$ is a van Fraassen function modulo Cn then it is a Hosiasson-Lindenbaum function modulo Cn'.

In brief: any van Fraassen function (modulo classical Cn) is a Hosiasson-Lindenbaum function modulo a suitably defined supraclassical consequence operation Cn', with the abnormal elements becoming Cn'-inconsistent. From this point of view, the differences between the three classes of functions may be seen as a matter of the background logic rather than one of probability. Of course, it should be remembered that, unlike classical consequence, such supraclassical consequence relations are not closed under substitution of arbitrary formulae for elementary letters (Makinson 2005), so we are using a rather different kind of logic.

3.2. Does it Ever Make a Difference?

Can the choice of kind of conditional probability ever make a substantive difference to an application? One example where it does is the theory of 'cores', set out by Arló Costa 2001 and Arló Costa & Parikh 2005 building on ideas of van Fraassen 1995.

Cores were introduced to give a probabilistic account of an intuitive distinction between a broader class of 'plain' beliefs and a narrower one of 'full' beliefs, with the formal desiderata that the classes are distinct, non-trival, and both closed under classical consequence (and hence under conjunction).

Translating from the field-of-sets mode used by the authors mentioned, a *core* for a Popper function $p: L^2 \to [0,1]$ is defined to be a formula c such that (1) c is normal, that is, the left projection p_c of p from the right value c is a proper Kolmogorov function, and (2) for all formulae b inconsistent with c we have $p(b, c^+ \lor b) = 0$ for every consistent c^+ logically implying c.

Plain beliefs modulo p are then identified with those formulae logically implied by at least one core, while *full* beliefs are those implied by every core. The authors show that in the finite case, for any Popper function $p: L^2 \to [0,1]$ there is a unique strongest core c_0 and a unique weakest one c_1; so that in that case plain beliefs are those formulae logically implied by c_0, while full beliefs are those implied by c_1. Indeed, in the field-of-sets mode we have the same whenever the underlying set is countable and we assume countable additivity.

However, for plain beliefs so defined, there is a difficulty. In the finite case they turn out to be just the formulae x with $p(x, \top) = 1$. In the field-of-sets mode, and assuming countable additivity, this also holds whenever the underlying set is countable. This is given as the 'coincidence lemma' of Arló Costa 2001 page 578, and is also an immediate consequence of Lemma 3.1 of Arló Costa and Parikh 2005. Thus in these contexts, the definition of *plain* belief in terms of cores gives us nothing new, no matter how we choose our Popper function. Nevertheless, as Parikh has urged (personal communication), when we are working in the uncountable case, or in the countable one but without countable additivity, we may not have the same collapse.

It does not seem to have been noticed in the literature that for full beliefs as defined via cores, the outcome depends critically on whether or not we are working with a Hosiasson-Lindenbaum function. If we are, it turns out that the full beliefs become just the tautologies – which is hardly what was wanted. To show this, we need only verify that $\top$ is itself a core. Using the definition above, it suffices to check that $p(\bot, \top) = 0$ (which is immediate) and that whenever b is inconsistent while a is consistent then $p(b, a \lor b) = 0$. But by the inconsistency of b we have $p(b, a \lor b) = p(b, a)$; and since p is a Hosiasson-Lindenbaum function, its left projection p_a from consistent a is a proper Kolmogorov function, so by the inconsistency of b again, $0 = p_a(b) = p(b, a)$. Note that this argument does not depend on any cardinality assumptions.

Thus the use of cores for defining a formal notion of full belief is not robust between the two notions of two-place probability, Hosiasson-Lindenbaum and Popper. The construction gives a non-trivial account of full belief only when there is at least one consistent a such that p_a is the unit function. Some might take this as a reason for preferring Popper to Hosiasson-Lindenbaum functions; others might take it as casting doubt on the value of the above definition of a core.

4. Back and Forth between Belief Revision and Conditional Probability

4.1. Correspondence between AGM and HL

We have been using AGM belief revision to explain why we should take seriously a revisionary reading of two-place probability functions and to help throw light on the options available, notably those of Hosiasson-Lindenbaum, Popper, and van Fraassen.

Which of these three accounts of conditional probability corresponds formally to AGM belief revision? From the discussion so far, one would guess that it is the HL system, and indeed that turns out to be the case. There is a natural map (due essentially to Lindström and Rabinowicz 1989, building on Gärdenfors 1988 chapter 5) from the family of all HL conditional probability functions into the family of all AGM revision operations on consistent belief sets. The definition is straightforward. Given an HL function $p(\cdot,\cdot)$ we define a belief set $K = B(p)$ to be the 'top' of p, i.e. $B(p) = \{x: p(x,\mathsf{T}) = 1\}$ and the revision operation $*_p$: $\{K\} \mathsf{X} L \to 2^L$ or more briefly $*_p$: $L \to 2^L$ by putting $*_p(a) = \{x: p(x,a) = 1\}$. It is not difficult to show that this map has the stated properties and that, moreover, it is surjective in the finite case − though of course far from injective). Details and verifications are given in the appendix.

The qualitative AGM axioms may thus be seen as reflections of the quantitative ones of Hosiasson-Lindenbaum. To this extent, the 1985 AGM postulates may be said to go back to 1940!

The existence of such a map prompts a number of further questions. To what kinds of qualitative belief revision do the systems of Popper and of van Fraassen correspond? Conversely, what kinds of conditional probability correspond to variant procedures for belief revision, such as the 'screened revision' of Makinson 1997? Are there any further interesting notions of conditional probability with a revisionary spirit that might have qualitative counterparts?

To get a qualitative analogue of Popper (while keeping classical consequence as our background consequence relation) we need to abandon or weaken the AGM postulate (K∗5): $K*a$ is consistent whenever a is consistent. For van Fraassen, we also need to add the revision function that makes $K*a$ inconsistent for every a, and, as a result, qualify postulate (K∗3): $K*a \subseteq Cn(K \cup \{a\})$. For further discussion of these matters, see Arló-Costa 2001.

4.2. From Screened Revision to Screened Conditional Probability

Screened revision is a variant form of AGM belief revision. Its basic idea is to see the operation as made up of two steps: a pre-processing step possibly followed by application of an AGM revision. The pre-processor decides the question of *whether* to revise, and this is done by checking whether the proposed input is consistent with a central part of the belief set under consideration, regarded as a protected subset. If they are mutually inconsistent, the belief set remains unchanged; otherwise we apply an AGM revision in a manner that protects the privileged material. Clearly, such a composite process will not satisfy all the postulates of AGM revision: for example, the postulate of success, $a \in K*a$, will fail in the first case. For more details, see Makinson 1997.

What would a probabilistic analogue of this look like? Roughly speaking, using the language of Leblanc cited in section 3.1, when a is too unbelievable to take seriously as a condition, we put the probability of x on condition a to be just the *unconditioned* probability of x rather than 1. In other words, for such a we require that $p(\cdot,a) = p(\cdot,\mathsf{T})$ rather than $p(\cdot,a) = 1(\cdot,a)$. At the same time, we protect the negation of a, by requiring that $p(\neg a,b) = 1 = p(\neg a,\mathsf{T})$ for all b. Thus, on the semantic level the functions are like those of Popper except that $p_\mathsf{T}(\cdot)$ takes the place of $1(\cdot)$ as the left projection p_a of

abnormal a, and negations of 'unbelievable' elements continue to get the value 1 under all conditions.

This forces modification of the axioms. In particular, the axiom (vF2) of left projection must be weakened: we no longer always have $p_a(a) = 1$ since when a is unbelievable $p_a(a) = p_\mathsf{T}(a) = p(a,\mathsf{T}) = 0$. In another respect, however, (vF2) can be strengthened: we can require that the left projection from any point is always a proper Kolmogorov function, as we no longer have any use for the unit function. The product axiom (vF3) must also be weakened. To show this, consider any inconsistent a. Unrestricted use of the product axiom would give us that for all x: $p_\mathsf{T}(x) = p_a(x) = p(x,a) = p(x \wedge x,a) = p(x,a) \cdot p(x,a \wedge x) = p(x,a) \cdot p(x,a) = p_a(x) \cdot p_a(x) = p_\mathsf{T}(x) \cdot p_\mathsf{T}(x)$; so that for any x, $p_\mathsf{T}(x)$ is either 0 or 1 – which is quite undesirable behaviour. The problem of axiomatizing the class of such 'screened two-place probability functions' appears to be open.

4.3. From Hyper-revisionary Conditionalization to Hyper-revisionary Revision

As is well known, for any van Fraassen function $p(\cdot,\cdot)$ and $a \in L$, if $p(a,\mathsf{T}) > 0$ then $p(x,a)$ is determined by a natural relativization of the ratio rule: $p(x,a) = p(a \wedge x,\mathsf{T})/p(a,\mathsf{T})$. Indeed, this equality is almost immediate: the product axiom gives us $p(a \wedge x,\mathsf{T}) = p(a,\mathsf{T}) \cdot p(x,a \wedge \mathsf{T}) = p(a,\mathsf{T}) \cdot p(x,a)$ by right extensionality, permitting division when $p(a,\mathsf{T}) > 0$.

As remarked by Jonny Blamey (personal communication), this might be seen as too conservative. For if a has a very low positive probability – say, to fix ideas, $0 < p(a,\mathsf{T}) < 0.01$ – then a surprise occurrence of a might sometimes lead us to question whether the function $p(\cdot,\cdot)$ was really right to give $p(a,\mathsf{T})$ such a small value. We should perhaps move to a function $q(\cdot,\cdot)$ which makes the truth of a less unexpected, i.e. puts $q(a,\mathsf{T})$ well above $p(a,\mathsf{T})$; and for such a q the value of $q(x,a)$ will be $q(a \wedge x,\mathsf{T})/q(a,\mathsf{T})$, which may be quite different from $p(a \wedge x,\mathsf{T})/p(a,\mathsf{T})$.

Philosophically, this 'hyper-revisionary' proposal drives an interesting wedge between two different ways of adopting a condition a. On the one hand, we may accept it because its truth has been revealed to us; on the other hand, we may entertain it to explore its consequences. The argument above suggests grounds for sometimes abandoning $p(\cdot,\cdot)$ when we are *confronted* with the truth of a proposition a for which p gave a very low value; but it does not suggest doing so when we merely *entertain* the truth of a to determine what effect it has on our probabilities. The hyper-revisionary proposal thus has the merit of providing formal expression to a difference between *accepting* and *supposing* a condition of low probability, which tends to be neglected by the usual treatments of conditional probability.

Of course, the proposal has a practical inconvenience. In applications, there can be no universally fixed cut-off point, such as 0.01, at which we should revise the probability function before applying the relativized ratio rule. Where to draw the line would be a matter of context, purposes and subject matter, balanced in an informal judgement. This situation is reminiscent of that arising in the theory of error statistics as developed by Fisher, Neyman and Pearson, where one considers the choice between rival statistical hypotheses under evidence that is logically consistent with each of them, but highly improbable given one while not so improbable given the other.

Indeed, there may be deep connections between hyper-revisionary conditional probability and error statistics, but we do not attempt to explore them in the present paper.

What would a qualitative analogue of such hyper-revisionary conditionalization look like? It would allow that even when input *a* is logically consistent with belief set *K*, we should not always take *K*a* to be *Cn(K∪{a})*. As well as adding in *a*, we should perhaps be contracting *K*, for despite the logical consistency of the two, *a* may be so implausible in the eyes of *K* that the revelation of the truth of the former may lead us to an 'agonizing reappraisal' of the latter.

This, of course, is counter to one of the basic postulates of AGM belief revision, K*4, which puts *K*a = Cn(K∪{a})* in *every case* that *a* is consistent with *K*, where we read 'consistency' as consistency under *Cn*, which in turn is taken to be classical consequence. Under this reading, AGM does not admit any conflict less than classical consistency as forcing contraction, and so K*4 must be modified for hyper-revisionary belief change. The exact adjustments required do not appear to have been studied.

These examples – from screened revision to a counterpart for conditionalization, and from hyper-revisionary conditionalization to a corresponding kind of revision – presumably do not exhaust the possibilities for going back and forth. As a rule of thumb, given an interesting variant of AGM qualitative belief revision we should expect a corresponding variant of Hosiasson-Lindenbaum conditional probability, and vice versa.

5. Proto-probability

In 1996, Hawthorne investigated rules of uncertain inference which, while qualitative, may be given a probabilistic justification, using them to form an axiom system that he called Q. All its axioms are in a natural sense probabilistically sound, although the converse has not yet been settled. The question arises: do we need the full force of the axioms of probability in order to justify the rules of Q, or can it be done with weaker axioms? In this section we observe that considerable weakening is possible. We need only certain modest order-theoretic conditions from among those available in the system of conditional probability of van Fraassen, already the weakest of those presented in section 2.2.

5.1. Hawthorne's system Q of Uncertain Inference

First, we recall Hawthorne's axioms. They concern consequence relations |∼ (in words: snake) between formulae of classical propositional logic. There are six Horn rules O1–O6 defining a system O, and one 'almost Horn' rule of 'negation rationality' (NR) whose addition gives Q. As usual, *Cn* is classical consequence and ≈ is classical equivalence:

O1. $a \mathrel{|\sim} a$ (reflexivity)

O2. When $a \mathrel{|\sim} x$ and $y \in Cn(x)$, then $a \mathrel{|\sim} y$ (RW: right weakening)

O3. When $a \mathrel{|\sim} x$ and $a \approx b$, then $b \mathrel{|\sim} x$ (LCE: left classical equivalence)

15

O4. When $a \mathrel{|\sim} x \wedge y$, then $a \wedge x \mathrel{|\sim} y$ (VCM: very cautious monotony)

O5. When $a \mathrel{|\sim} x$, $b \mathrel{|\sim} x$ and $\neg b \in Cn(a)$, then $a \vee b \mathrel{|\sim} x$ (XOR: exclusive $\vee$+)

O6. When $a \mathrel{|\sim} x$ and $a \wedge \neg y \mathrel{|\sim} y$, then $a \mathrel{|\sim} x \wedge y$ (WAND: weak $\wedge$+).

NR. When $a \vee b \mathrel{|\sim} x$ and $\neg b \in Cn(a)$, then either $a \mathrel{|\sim} x$ or $b \mathrel{|\sim} x$.

As Hawthorne showed, the system Q is probabilistically sound in the sense that for any probability function $p(\cdot,\cdot)$ satisfying van Fraassen's postulates and 'threshold' $t \in [0,1]$, if we define a relation by putting $a \mathrel{|\sim_{pt}} x$ iff $p(x,a) \geq t$, then $\mathrel{|\sim_{pt}}$ satisfies all the rules of Q. For further information on systems O and Q see Hawthorne 1996, Hawthorne and Makinson 2007, Paris and Simmonds 2009, Simmonds 2010, Makinson (to appear). In particular, Paris and Simmonds have shown that O (i.e. the above without NR) is not complete for the class of probabilistically sound Horn rules.

5.2. Proto-probability Functions

Our question is: how much probability is really needed for the job? If we simply drop one of the van Fraassen axioms, then we admit functions that do not validate Hawthorne's system Q. So, rather than delete, we abstract. It turns out that the validation can be effected by any function into an arbitrary complete preorder with greatest and least elements, satisfying certain very modest conditions in which no arithmetical operations appear.

Let D be any non-empty set equipped with a relation $\leq$ that is transitive and complete ($d \leq e$ or $e \leq d$, for all $d,e \in D$) with a greatest element 1_D and a distinct least element 0_D. Note that we do not require that $\leq$ is anti-symmetric (and thus linear), although of course it may be so. A *proto-probability* function into D is any function $p: L^2 \to D$ satisfying the following six conditions:

P1. $p(a,a) = 1_D$

P2. $p(x,a) \leq p(y,a)$ whenever $y \in Cn(x)$

P3. $p(x,a) = p(x,b)$ whenever $a \approx b$

P4. $p(x \wedge y,a) \leq p(y,a \wedge x)$

P5. $p(x,a) \leq p(x,a \vee b) \leq p(x,b)$ whenever $p(x,a) \leq p(x,b)$ and $\neg b \in Cn(a)$

P6. $p(x,a) = p(x \wedge y,a)$ whenever $p(y,a \wedge \neg y) \neq 0_D$.

We call condition (P5) the principle of *disjunctive interpolation*. The part $p(x,a \vee b) \leq p(x,b)$ (under the stated conditions) is essentially the same as a principle of 'alternative presumption' of Koopman 1940, 1940a. The condition as a whole is essentially the finite case of a rule known as conglomerability, due to Seidenfeld et al 1998. It may also be seen as extracting the qualitative content of a principle for quantitative conditional probability that was articulated by Gärdenfors 1988. The appendix details all three connections.

If we take any proto-probability function $p(\cdot,\cdot)$ and $t \in D$, and define a relation by putting $a \mathrel{|\sim_{pt}} x$ iff $p(x,a) \geq t$, then $\mathrel{|\sim_{pt}}$ satisfies all the rules of Q. This fact may be seen as a soundness theorem for Q with respect to proto-probability functions. Indeed, each condition (Oi) follows directly from its counterpart (Pi), with (NR) also following

from (P5). The completeness of the relation $\leq$ over D is needed to derive the two postulates of Q dealing with disjunction, i.e. NR and O5 (alias XOR). The verifications are trivial, but given the novelty of the notion of proto-probability, we provide them in the appendix.

It is also easy to check (see appendix) that when D is set at [0,1] and $\leq$ as usual, then the axioms for proto-probability functions follow from those of van Fraassen, *a fortiori* from the stronger systems discussed in section 2.2. In fact, they are considerably weaker. Informally, it is clear that the left projection and product axioms of van Fraassen do not hold for all proto-probability functions since our conditions for the latter make no use of addition (which is implicit in the left projection axiom), nor of multiplication (explicit in the product axiom).

For a specific example of a proto-probability function that is not a van Fraassen one, take p: $L^2 \rightarrow \{0,1\}$ to be the characteristic function of the classical consequence relation, i.e. put $p(x,a) = 1$ when $x \in Cn(a)$, otherwise $p(x,a) = 0$. Clearly, this satisfies conditions P1 through P6, but it fails (vF2) since $p(x \vee \neg x, \mathsf{T}) = 1$ while $p(x, \mathsf{T}) = 0 = p(\neg x, \mathsf{T})$ for contingent formulae x, so that p_T is not a Kolmogorov function.

In summary, the proto-probability functions are defined by purely order-theoretic conditions that are strictly weaker than the main systems for conditional probability as described in section 2 above. Yet they are strong enough to support the rules defining Hawthorne's system Q of probabilistic inference. In fact, we also have a representation theorem for Q in terms of proto-probability functions: given any consequence relation $|\sim$ satisfying the conditions of system Q, there is a proto-probability function p: $L^2 \rightarrow D$ with $|\sim \ = \ |\sim_{pt}$. The proof is quite trivial. Choose $D = \{0,1\}$ and $\leq$ as usual over it, take p: $L^2 \rightarrow D$ to be the characteristic function of $|\sim$ (i.e. $p(x,a) = 1$ when $a |\sim x$, else $p(x,a) = 0$), and finally put $t = 1$. It is straightforward to check that p is a proto-probability function, and we have immediately that $|\sim \ = \ |\sim_{p1}$ by the equivalences $a |\sim x$ iff $p(x,a) = 1$ iff $a |\sim_{p1} x$.

From the soundness and representation systems for the system Q we immediately have a representation theorem for proto-probability functions themselves. For any proto-probability function q: $L^2 \rightarrow E$ and any $t \in E$, there is a proto-probability function p: $L^2 \rightarrow \{0,1\}$ with $|\sim_{qt} = |\sim_{p1}$. This may also be verified directly without passing through the logic Q: given q: $L^2 \rightarrow E$ and $t \in E$, simply put $p(x,a) = 1$ when $q(x,a) \geq t$ and $p(x,a) = 0$ otherwise.

Representation theorems normally imply associated completeness theorems, though not always conversely (see Makinson 2007 for a general discussion). This is no exception. Consider any rule – whether Horn or allowing negative premises and/or conclusion – with premises $\pm_i(a_i |\sim x_i)$ and conclusion $\pm(b |\sim y)$, where $\pm$ is affirmation or denial. Suppose that it fails for some consequence relation $|\sim$ satisfying all postulates of Q. Then there is a proto-probability function p (namely the one that represents $|\sim$) such that each $p(x_i,a_i)$ is correspondingly 1 or 0 while $p(y,b)$ is contrariwise 0 or 1.

Closely related to the results above is a closure property of the class of all proto-probability functions. Let D, E be any non-empty sets equipped respectively with transitive complete relations $\leq, \leq'$, with greatest and least elements $1_D, 0_D, 1_E, 0_E.$

Consider any proto-probability function $p: L^2 \to D$ and order-preserving function h: $D \to E$ with $h(1_D) = 1_E$ and $h(0_D) = 0_E$. Then the composition $p': L^2 \to E$ defined by putting $p'(x,a) = h(p(x,a))$ is also a proto-probability function. In particular, this is the case when we choose $E = \{0,1\}$, $t \neq 0_D$ in D, and $h(d) = 1$ iff $d \geq t$ else $h(d) = 0$. The verification is straightforward; the only condition that needs attention is P5, which we give in the appendix.

In contrast, however, the class of all proto-probability functions is not closed under direct products, since the intersection of two complete relations over a set is not in general complete.

5.3. Comparison with Plausibility Measures, and Further Examples

How do proto-probability functions compare with the well-known 'conditional plausibility measures' studied by Halpern in a number of papers, e.g. Halpern 2001? A short answer is that while their conditions on the relation $\leq$ are incomparable, as are the domains of the functions, Halpern's conditions are, roughly speaking, considerably more general; a more precise comparison is given in the appendix. This is not surprising, for the motivations are not the same. We are asking how much probability is needed to validate the properties (as given by system Q) of probabilistically sound qualitative consequence relations; Halpern is looking for a 'most general' kind of conditional probability that includes all those known in the literature.

We note three further kinds of function that are simultaneously conditional plausibility measures and proto-probability functions, namely the 'conditional ranking functions' of Spohn e.g. 1986, 2009, 'conditional possibility functions' of Dubois and Prade e.g. 1988, and 'conditional quasi-measures' of Weydert 1994. We simply state some basic facts, omitting the verifications.

- Let $\kappa: L \to \mathbf{N} \cup \{\infty\}$ be a 'negative ranking function' in the sense of Spohn (expressed in the propositional rather than his field of sets mode) on the language L into the natural numbers together with ∞. Consider Spohn's associated 'conditional ranking function' defined by putting $\kappa(x/a) = \kappa(a \wedge x) - \kappa(a)$. Then if we convert the order of the ranking (so that 0 becomes the greatest element and ∞ the least), the function $\kappa(\cdot|\cdot)$ satisfies the conditions (P1) through (P6) for proto-probability functions (without needing the hypothesis that $\neg b \in Cn(a)$ for P5).

- Let $\pi: L \to [0,1]$ be a 'possibility measure' in the sense of Dubois and Prade (again in the propositional rather than their field-of-sets mode), i.e. $\pi(a) = \pi(b)$ when a,b are classically equivalent, $\pi(a) = 0$ when a is a contradiction, $\pi(a) = 1$ when a is a tautology, and $\pi(a \vee b) = \max\{\pi(a),\pi(b)\}$. Consider the associated 'conditional possibility function' defined by Dubois and Prade 1988 page 206, which sets $\pi(x/a) = \pi(a \wedge x)/\pi(a)$, except when $\pi(a) = 0$ as in the ratio definition from Kolmogorov probabilities. Then $\pi(\cdot|\cdot)$ (without conversion of the order) satisfies conditions (P1)–(P6).

- Weydert 1994 abstracted the common algebraic features of the above two examples in his 'quasi-measure spaces', and the resulting 'conditional quasi-

measures' are also both conditional plausibility measures and proto-probability functions.

Appendix

This appendix runs parallel to the main text. It contains most of the formal definitions and verifications, as well as references and historical remarks supporting the main text.

For Section 1: Why Go Beyond the Ratio Rule?

The Kolmogorov axioms

There are several modes for presenting the Kolmogorov axioms for one-place probability functions, according to what we take as their domain. It may be a field of sets (most common in mathematics and applications), or equivalently a Boolean algebra (the preferred way of algebraists), or the set of all formulae of a propositional language (whose quotient structure under classical equivalence will be a free Boolean algebra). In this paper we work in the propositional mode, with the following formulation (Makinson 2005) of the postulates.

A (one-place) *proper Kolmogorov function p*: $L{\rightarrow}[0,1]$ is any function defined on the set L of formulae of a language closed under the Boolean connectives, into the real numbers from 0 to 1, such that:

(K1) $p(x) = 1$ for some formula x
(K2) $p(x) \leq p(y)$ whenever $y \in Cn(x)$
(K3) $p(x{\vee}y) = p(x)+p(y)$ whenever $\neg y \in Cn(x)$.

Cn is classical consequence; we also write $\approx$ for classical equivalence. Thus postulate (K1) tells us that 1 is in that range of p; (K2) says that $p(x) \leq p(y)$ whenever x classically implies y; (K3), called the rule of finite additivity, tells us that $p(x{\vee}y) = p(x)+p(y)$ whenever x is inconsistent with y. It is sometimes extended so as to constrain the probability of countable unions (most easily expressed in the field of sets mode).

As remarked in the text and observed by many authors, e.g. Harper 1975 and subsequently Gärdenfors 1988, Leblanc and Roeper 1989, in comparative contexts it is convenient to regard the *unit function* (i.e. the function p that puts $p(x) = 1$ for every $x \in L$) as also being a Kolmogorov function, and we will follow this convention. It can be formalized by the simple expedient of defining a *Kolmogorov function* as one that is *either* a proper Kolmogorov function (i.e. satisfies the above postulates) *or* is the unit function. Equivalently, one could weaken axiom (K3) by putting it under the proviso that p is not the unit function. We refer to the unit function as the *improper* Kolmogorov probability function.

The ratio rule

The *ratio rule* for conditional probability uses an arbitrary Kolmogorov function p: $L{\rightarrow}[0,1]$ to define a two-place function, conventionally written as $p(x|a)$ and read as

'the probability of x given a', defined on $L\mathsf{x}\{a \in L: p(a) > 0\}$ by the rule: $p(x|a) = p(a \wedge x)/p(a)$ when $p(a) > 0$ and otherwise undefined.

Left projections

We recall the standard concept of the left projection f_a: $X{\rightarrow}Y$ of a two-place function f: $X\mathsf{x}A{\rightarrow}Y$ from point $a \in A$, defined by putting $f_a(x) = f(x,a)$ for all $x \in X$.

For Section 1.1. Doctrinal *vs* Pragmatic Considerations

The view that all probability is 'really' conditional

Such views have been expressed by a number of probabilists, notably Rényi 1955 and 1970, de Finetti 1974 and by some philosophers, e.g. Hájek 2003.

Rényi 1955 (page 286) puts it briefly: "In fact, the probability of an event depends essentially on the circumstances under which the event possibly occurs, and it is a commonplace to say that in reality every probability is conditional". The same idea recurs at greater length in his 1970 (page 35).

De Finetti 1974 (page 134) similarly remarks: "Every evaluation of probability is conditional; not only on the mentality or psychology of the individual involved, at the time in question, but also, and especially, on the state of information in which he finds himself at that moment."

More recently, Hájek 2003 writes: "...given an unconditional probability, there is always a corresponding conditional probability lurking in the background. Your assignment of 1/2 to the coin landing heads superficially seems unconditional; but really it is conditional on tacit assumptions about the coin, the toss, the immediate environment, and so on. In fact, it is conditional on your total evidence."

Carnap's regularity condition

Carnap's formulation of the additional 'regularity' condition may be found in his book of 1950 section 53 axiom C53-3 and also the paper 1971 chapter 2.7 page 101, cf also his 1952.

It may be suggested that when constructing a specific one-place probability function in an empirical investigation, the wise researcher will assign extremal values (zero and one) as seldom as possible, so as to minimize the likelihood of conditionalization problems further down the line. However, as has often been observed, such a policy would have the inconvenience of impeding the free use of Bayesian conditionalization, under which $p_a(a) = 1$ for all a, replacing it by the rather more complex Jeffrey conditionalization.

We note in passing that the concept of a 'counterfactual probability function' discussed by Boutilier 1995 (building on Stalnaker 1970) also assumes that the critical zone is empty. That concept, defined in the finite case, is a curious mixture of quantitative and qualitative ingredients. It puts $p(x,a)$, called the counterfactual

probability of x given a, to be the proportion of the 'best' a-states of the model that are x-states. The emptiness of the critical zone is assumed for the same reason as before: to ensure that the denominator is non-zero for consistent formulae a.

For Section 1.2. Some Notational Niceties

Two-place functions could alternatively be distinguished from one-place ones by different type-faces, e.g. lower case for one and upper case for the other. However that convention meshes poorly with the standard notation for left projection, which we also need to use extensively.

For Section 2.1. A Leaf from the AGM Book

How important is the critical zone?

Our view of the importance of the critical zone contrasts with its minimization by some authors. For example McGee 1994: "The problem we have been examining, how to revise one's system of beliefs upon obtaining new evidence that had prior probability 0, is not a problem that has any great practical significance."

Conditional probability in the light of counterfactual conditionals

An argument for going beyond the ratio definition of two-place probability may also be made in terms of counterfactual conditionals rather than belief revision. Indeed, that is the way in which it is usually developed in the philosophical literature, going back to Stalnaker 1970. However, in the author's view, the comparison with belief revision affords a clearer view, and also lends itself to the construction of natural formal maps, as shown in section 4.

Verifications of properties of B(p)

We verify the claims made in bullet points about belief sets for probability functions. Given a one-place function p we define the corresponding belief set $B(p) = \{x: p(x) = 1\}$. This is also sometimes called the *top* of the function. Write $B+a$ for the qualitative expansion of B by a, i.e. $B+a = Cn(B \cup \{a\})$. With $p_a(\cdot)$ understood as the left projection from a of the conditionalization $p(\cdot|\cdot)$ obtained from $p(\cdot)$ by the ratio/unit rule, we want show: (1) in all cases, $B(p) \subseteq B(p)+a \subseteq B(p_a)$ and (2) in the limiting case that $p(a) = 0$ we have belief explosion: $B(p)+a = L = B(p_a)$, where L is the set of all propositions of the language.

For (1), the first inclusion is immediate from the definition of expansion above. To check the second inclusion, note that since $B(p_a)$ is closed under consequence it suffices to show that $a \in B(p_a)$ and $B(p) \subseteq B(p_a)$. The former is immediate since when $p(a) > 0$ then $p_a(a) = 1$ by the ratio definition and the Kolmogorov postulates for one-place probability, and $p_a(a)$ is also 1 when $p(a) = 0$, by the unit part of the ratio/unit definition. For the latter, it suffices to show that whenever $p(x) = 1$ then $p_a(x) = 1$. This is immediate when $p(a) = 0$. When $p(a) > 0$ we have $p_a(x) = p(a \wedge x)/p(a) = p(a)/p(a) = 1$ since the hypothesis $p(x) = 1$ implies that $p(a \wedge x) = p(a)$. For (2), it suffices to show further that when $p(a) = 0$ we have $B(p)+a = L$. But when the hypothesis holds then $p(\neg a) = 1$, so $\neg a \in B(p)$ and thus $B(p)+a \supseteq Cn(\neg a, a) = L$.

For Section 2.2. Bird's-eye View of Available Systems

Verification of consequences of the van Fraassen axioms

Left extensionality: $p(x,a) = p(x',a)$ whenever $x \approx x'$. *Verification*: By left projection, p_a is either a proper Kolmogorov function or the unit function. In the former case, $p(x,a) = p_a(x) = p_a(x') = p(x',a)$ using the hypothesis. In the latter case, $p(x,a) = p_a(x) = 1 = p_a(x') = p(x',a)$ irrespective of the hypothesis.

When $y \in Cn(x)$ then $p(x,a) \leq p(y,a)$. *Verification*: If $y \in Cn(x)$ then $x \approx y \wedge x$ so by left extensionality and product, $p(x,a) = p(y \wedge x,a) = p(y,a) \cdot p(x,a \wedge y) \leq p(y,a)$.

When $p(\cdot)$ is defined as $p(\cdot,\mathsf{T})$, then we have the ratio rule. *Verification*: Suitably instantiating the product axiom, $p(a \wedge x,\mathsf{T}) = p(a,\mathsf{T}) \cdot p(x,a \wedge \mathsf{T}) = p(a,\mathsf{T}) \cdot p(x,a)$ using right extensionality, so if $p(a,\mathsf{T}) > 0$ we have $p(x,a) = p(a \wedge x,\mathsf{T})/p(a,\mathsf{T}) = p(a \wedge x)/p(a)$.

When a is a contradiction, then p_a is the unit function. *Verification*: $1 = p_a(a) = p(a,a) \leq p(x,a) = p_a(x)$, using left projection and an inequality already established.

The set Δ of all $a \in L$ such that p_a is the unit function is an ideal. *Verification*: To show that Δ is closed downwards, suppose $a \in \Delta$ and $a \in Cn(b)$. Then $1 = p(b \wedge x,a) = p(b,a) \cdot p(x,a \wedge b) = 1 \cdot p(x,a \wedge b) = p(x,b) = p_b(x)$, using the first supposition, product, first supposition again, second supposition respectively. To show that Δ is closed under disjunction, suppose p_a, p_b are both the unit function. To show that $p_{a \vee b}$ is also the unit function it suffices, by the left projection axiom to show that it is not a proper Kolmogorov function. Suppose it is; we get a contradiction. From the van Fraassen axioms we have $p(\bot,a \vee b) = p(\bot \wedge a,a \vee b) = p(a,a \vee b) \cdot p(\bot,a \wedge (a \vee b)) = p(a,a \vee b) \cdot p(\bot,a) = p(a,a \vee b) \cdot 1 = p(a,a \vee b)$ using the supposition that p_a is the unit function. Likewise $p(\bot,a \vee b) = p(b,a \vee b)$. By the supposition that $p_{a \vee b}$ is a proper Kolmogorov function we have $p(\bot,a \vee b) = 0$ so $p(a,a \vee b) = 0 = p(b,a \vee b)$. By the same supposition, $p(a \vee b,a \vee b) \leq p(a,a \vee b) + p(b,a \vee b) = 0 + 0 = 0$, contradicting the second part of the left projection axiom.

Finally, we check that a is abnormal iff $p(a,b) = 0$ for all normal b. *Verification*: From right to left, suppose $p(a,b) = 0$ for all normal b, but a is not abnormal. Then a is normal, so $p(a,a) = 0$, contradicting the second part of the left projection axiom. From left to right, suppose a is abnormal and b is normal. Then $a \wedge b$ is abnormal as already established, so $0 = p(\bot,b) = p(\bot \wedge a,b) = p(a,b) \cdot p(\bot,a \wedge b) = p(a,b) \cdot 1 = p(a,b)$ as desired.

Verification of the alternative axiomatization of the Popper system

For the easy half, assume the van Fraassen axioms plus (Positive); we need to show that $p(x,b) \neq 1$ for some x,b. By left projection, $p(\mathsf{T},\mathsf{T}) = 1 > 0$ so by (Positive) p_T is proper and thus $p(\bot,\mathsf{T}) = 0 \neq 1$ as desired. For the tricky half, assume the van Fraassen axioms plus $p(x,b) \neq 1$ for some x,b. Suppose $p(a,\mathsf{T}) > 0$; we need to show that p_a is proper, for which it suffices to show that it is not the unit function. First note that $p(\bot,\mathsf{T}) = p(\bot \wedge b,\mathsf{T}) = p(b,\mathsf{T}) \cdot p(\bot,\mathsf{T} \wedge b) = p(b,\mathsf{T}) \cdot p(\bot,b)$; but since $p(x,b) \neq 1$ it follows that p_b is proper so $p(\bot,b) = 0$ and thus $p(\bot,\mathsf{T}) = 0$. But also $p(\bot,\mathsf{T}) = p(\bot \wedge a,\mathsf{T}) =$

$p(a,\top){\cdot}p(\bot,a)$, so since $p(a,\top) > 0$ we have $p(\bot,a) = 0$ so that p_a is not the unit function, as desired.

Example of a 'mixed' function

Leblanc and Roeper 1989 gave an example of a two-place function satisfying the Popper postulates, whose treatment of formulae with probability zero is a mix of the expansionary and revisionary policies. They presented it rather enigmatically as an 8·8 table (their Table 5). We provide it with a more transparent rule-based presentation, which for convenience we express with a field of sets.

Take the field F of all subsets of the three-element set $S = \{\alpha,\beta,\gamma\}$. For motivation, think of α,β,γ as being of increasing levels of importance beginning from α, which has no importance at all. For $a,x \subseteq S$, put $p(x,a) = 1$ unless there is some item of positive importance in a and the item of greatest importance in a is not in x. More precisely, we define $p: S^2 {\to} [0,1]$, in fact into $\{0,1\}$, as follows:

1. If $\gamma \in a$ then $p(x,a) = 1$ if $\gamma \in x$, otherwise $p(x,a) = 0$
2. If $\gamma \notin a$ but $\beta \in a$ then $p(x,a) = 1$ if $\beta \in x$, otherwise $p(x,a) = 0$
3. If $\gamma \notin a$ and $\beta \notin a$ then $p(x,a) = 1$.

This function is a mix of the two kinds of conditional probability: $p(\{\beta\},S) = 0 = p(\{\alpha\},S)$ applying the first clause, but $p(\varnothing,\{\beta\}) = 0$ applying the second while $p(\varnothing,\{\alpha\}) = 1$ by the third. On the other hand, it is straightforward to check that it satisfies the Popper axioms.

Historical development of conditional probability

We review the historical steps in the construction of axioms for two-place probability functions, working backwards from Popper 1959. For ease of comparison, we consider them all in the propositional mode, and treat each as defined on the whole of L^2, but comment on particularities of the original formulations each as we go.

Popper's original postulates for two-place probability functions, contained in an appendix of Popper 1959 (recalled e.g. in Leblanc and Roeper 1989 and more accessibly Koons 2009) were in the propositional mode. They reflected a desire for the autonomy of probability theory from logic, abstract algebra and set theory and so avoided any use of concepts from those areas. But if we are happy to use concepts of classical logic in our presentation then, as shown by subsequent writers, Popper's axioms may be given more perspicuously. The following formulation of Hawthorne 1996 requires that for $p: L^2 {\to} [0,1]$:

(P0) $p(x,a) \neq 1$ for some formulae a, x
(P1) $p(x,a) = p(x,b)$ whenever $a \approx b$
(P2) $p(x,a) = 1$ whenever $x \in Cn(a)$
(P3)   either $p(x{\vee}y,a) = p(x,a)+p(y,a)$ whenever $\neg(x{\wedge}y) \in Cn(a)$, or p_a is the unit function
(P4) $p(x{\wedge}y,a) = p(y,a){\cdot}p(x,y{\wedge}a)$

23

Of course, if we are working in the context of fields of sets, (P1) becomes vacuous. Warning: The term 'Popper function' is sometimes used rather loosely, to refer to almost any primitive two-place probability function defined over the critical zone. For example, Lindström and Rabinowicz 1989 use the term to refer to the narrower class of Hosiasson-Lindenbaum functions, defined below.

Our modular presentation takes from Rényi 1955, 1970, 1970a his leading idea that for 'most' values of a, the left projection from a will be a proper Kolmogorov function giving a the value 1, and so is very similar in gestalt. But in its details, Rényi's system is rather different from any of those we have considered. Formulated in the field-of-sets mode, it treats the right domain as a *parameter*, allowing it to be chosen as any subset of the left domain that is consistent with the axioms. These axioms are just the product rule and the principle that p_a is a *proper* one-place Kolmogorov function with $p_a(a) = 1$, both formulated under the restriction that the right argument takes a value in the restricted right domain. For values of the right argument outside that subset, the probability functions are left undefined. We are thus given a scheme for a *family* of axiom sets, one for each choice of right domain.

This yields the Popper axioms if we constrain the right domain to include $\{a: p(a,S) > 0\}$, where S is the set on which the field is based, and carry out the following editing: (a) put $p(x,a) = 1$ for all a outside the right domain, (b) ensure consistency by allowing in the left projection axiom that p_a may be improper (as in the axiom (vF2) of section 2.2), (c) for the one-place Kolmogorov functions mentioned in the left projection axiom, weaken Rényi's assumption of countable to finite additivity, and finally (d) translate from the field-of-sets mode to the propositional one.

The system of Hosiasson-Lindenbaum 1940 concerned what she called 'confirmation' functions, writing them as $c(x,a)$ rather than $p(x,a)$ and working in the propositional mode. This ground-breaking work has been comparatively neglected, despite its accessible and respected place of publication. In particular, the paper is not mentioned in any of Rényi 1955, 1970, 1970a, nor in the wide-ranging discussion of Harper 1975 or the comprehensive study of Roeper and Leblanc 1999. Popper 1959 does mention Hosiasson-Lindenbaum in passing, but with respect to other questions and without citing her 1940 paper. This contrasts with his explicit acknowledgement (note 12 in new appendix iv) of the influence of Rényi 1955 on his thinking.

We remark that the Hosiasson-Lindenbaum system reappears in field-of-sets form in Dubins 1975, a paper that has been particularly influential among statisticians. However, Dubins' definition (of 'full conditional probability' in his section 3) appears to have been devised independently: he does not mention Hosiasson-Lindenbaum's paper, and the manner of presentation suggests the influence of Rényi.

Hosiasson-Lindenbaum excluded inconsistent propositions from the right domain (likewise Dubins excluded the empty set in his later version). Restoring the inconsistent propositions to make that domain full, we get the following axioms:

(HL1) $p(x,a) = 1$ whenever $x \in Cn(a)$

(HL2) $p(x \vee y,a) = p(x,a) + p(y,a)$ whenever $\neg(x \wedge y) \in Cn(a)$, provided a is consistent

(HL3) $p(x \wedge y,a) = p(x,a) \cdot p(y,a \wedge x)$ for all formulae a, x, y

24

(HL4) $p(x,a) = p(x,b)$ whenever $a \approx b$.

Axiom (HL2) thus broadens the conditions under which the left projection of a two-place function satisfies additivity and is thus a proper Kolmogorov function, from the narrower case $p(a,\mathsf{T}) > 0$ to the wider one that a is consistent. The system may be obtained fron Rényi's scheme by putting the right domain to be the set of all non-empty sets of S and editing by first putting $p(x,\varnothing) = 1$ and then as for Popper's system.

In what respect can it be said that Rényi's formulation was an advance on that of Hosiasson-Lindenbaum? For working mathematicians and statisticians, its use of the field-of-sets mode made application to practical problems more transparent. The variability of the right domain may have made it more flexible. But at a deeper level, the step forward from earlier formulations was *conceptual* − the realization that a rather arbitrary-looking axiom system becomes natural if we build it around the idea that for 'most' values of the right argument, the left projection will be a proper one-place probability function. As Rényi put it: "a conditional probability space is nothing else than a set of ordinary probability spaces which are connected with each other by [the product axiom]" (Rényi 1955 pp 289-290).

Mini-note: We reverse a correction made by Hailperin 1991 (page 75) to the effect that since Hosiasson-Lindenbaum's formulation is in the propositional mode, it needs a left companion to (HL4) stating that $p(x,a) = p(y,a)$ whenever $x \approx y$. In fact, this follows from the postulates as given. In the limiting case that a is inconsistent we have $p(x,a) = 1 = p(y,a)$ by (HL1), so suppose a is consistent and $x \approx y$. Then $\neg(x \wedge \neg y) \in Cn(a)$, so by the additivity axiom (HL2) we have $p(x \vee \neg y, a) = p(x,a) + p(\neg y, a)$. But the supposition also gives us LHS = 1 by (HL1), so $p(x,a) + p(\neg y, a) = 1$. Moreover, (HL1) and (HL2) imply that $p(\neg y, a) = 1 − p(y,a)$, and so by arithmetic $p(x,a) = p(y,a)$. Essentially this point was already made by Tarski with regard to the earlier axiomatization of Mazurkiewicz 1932 (discussed below), and was acknowledged in footnote 1 of that paper.

Hosiasson-Lindenbaum 1940 states that her axioms for two-place probability are "analogous" to still earlier ones of Mazurkiewicz 1932. In fact, they considerably simplify and clarify his quite complex system, which requires the left domain to contain *individual* propositions, while the right one contains consistent *sets* of propositions closed under classical consequence – the two kinds of proposition drawn, moreover, from intersecting and not very clearly defined languages. In his only example, Mazurkiewicz considers a game: the left argument of $p(x,A)$ can be filled by a proposition describing a state of play, while the right one can be occupied by a closed set of propositions containing the rules of the game, the current state of play, and any mathematical apparatus needed for deductions.

In turn, Mazurkiewicz states that he is taking as his starting point the axioms of Bohlmann 1909. However, Bohlmann's postulates are for *one*-place probability in a mode of unanalysed items called events and occurrences, which he supplements with an 'axiom' defining conditional probability by the ratio rule.

For some late nineteenth-century uses of conditional probability (without any attempt at axiomatization) see Hailperin 1988.

Thus our trail into the history of axiomatizations of two-place probability that cover the critical zone appears to end with Mazurkiewicz 1932 as first serious attempt, Hosiasson-Lindenbaum 1940 as the first really successful one, and Rényi 1955 for providing a clear gestalt.

For Section 3.1. Hosiasson-Lindenbaum *vs* Popper *vs* van Fraassen

Bayesian update of a conditional probability function

As mentioned in the text, it may be suggested that while Bayesian update as defined by the rule $p_{\wedge b}(x,a) = p(x,a \wedge b)$ is suitable for an expansionary notion of conditional probability given by the ratio or ratio/unit definition, it is quite inappropriate for a revisionary one. The class of Hosiasson-Lindenbaum (or of Popper) functions should indeed be closed under conditionalization, but that operation should be understood differently. Once again, the point may be appreciated by comparing with the situation for qualitative belief change. The counterpart of Bayesian update for qualitative expansion is the equality $(K+a)+b = K+(a \wedge b)$, which is trivially correct by classical logic. But the counterpart for revision would be $(K*a)*b = K*(a \wedge b)$, which is quite inappropriate, conflicting with the principle of conservation of consistency of input (i.e. that $K*x$ is consistent whenever x is consistent). It is acceptable only in the special case that b is consistent with $K*a$, where the equality $(K*a)*b = (K*a)+b$ is given by the AGM supplementary postulates $(K*7)$ and $(K*8)$.

So how should we define conditionalization of a revisionary two-place probability function $p(\cdot,\cdot)$ under an input proposition a? The short answer is that there is no such definition, because there is not a unique operation of this kind. Just as on the qualitative level there are many ways of revising a belief set and thus in particular of revising its revision, so too on the quantitative level there are many ways of revising a conditional probability function $p(\cdot,\cdot)$ given an input proposition a. In both contexts we may articulate interesting regularity conditions, but there is no formally justified choice of a unique and universally applicable operation taking conditional probability function $p(\cdot,\cdot)$ and propositional input a to a conditional probability function $p_{*a}(\cdot,\cdot)$ that is waiting to be expressed as a definition. Moreover, just as the task of settling on suitable regularity conditions for iterated qualitative belief revision is notoriously difficult (much more so than in the case of one-shot AGM revision) and is still under debate, so too we may expect that the task of articulating consensual regularity conditions on the passage from $p(\cdot,\cdot)$ and a to $p_{*a}(\cdot,\cdot)$, following input of proposition a, will not be easy.

Changing the underlying consequence operation

If one is working in the mode of fields-of-sets, or of Boolean algebras as carriers for the probability functions, then one can similarly express van Fraassen functions as Hosiasson-Lindenbaum ones by passing to the quotient algebra determined by the same filter as in the propositional mode. Essentially this construction was used for different purposes by Harper 1976 (section 6).

For Section 4.1. Correspondence between AGM and HL

A map from HL into AGM

Lindström and Rabinowicz 1989, building on work of Gärdenfors 1988 chapter 5, already constructed a map from the class of all Gärdenfors probability-revision operations into the class of AGM belief revision operations. The construction below essentially translates it (with some simplifications and an explicit verification of surjectivity in the finite case) into a map from the class of Hosiasson-Lindenbaum probability functions to the AGM operations.

Given any HL function p: $L^2 \to [0,1]$ as defined in section 2.2 or equivalently in its appendix, we construct the associated belief set $K = B(p)$, called the *top* of p, as follows:

- $B(p) = *_p(\mathsf{T}) = \{x: p(x,\mathsf{T}) = 1\}$.

We define an AGM belief revision function with this set K fixed, i.e. as the two-place operation $*_p$: $\{K\} \times L \to 2^L$ with singleton left domain, or more briefly the one-place operation $*_p$: $L \to 2^L$, by putting:

- $*_p(a) = \{x: p(x,a) = 1\}$.

We need to show that for every HL function p: $L^2 \to [0,1]$:

- $B(p)$ is a consistent belief set.
- The operation $*_p$: $L \to 2^L$ satisfies the full set of AGM postulates (K$*$1) through (K$*$8) with respect to $K = B(p)$.

First, recall from section 2.2 that for HL functions $p(\cdot,\cdot)$, the left projection p_a from a is a proper Kolmogorov one-place probability function whenever a is consistent, so we can apply well-known properties of the one-place functions without detailed justification, as well as the HL axioms themselves.

To show that $K = B(p)$ is a belief set, suppose $y \in Cn(K)$; we need to check that $y \in K$. By compactness, $y \in Cn\{\wedge x_i: i \le n\}$ for some $x_1,.., x_n \in B(p)$, so each $p(x_i,\mathsf{T}) = 1$, so $p(\wedge x_i,\mathsf{T}) = 1$ and thus $p(y,\mathsf{T}) = 1$ so that $y \in B(p)$. To show that $B(p)$ is consistent we need then only note that $p(\bot,\mathsf{T}) = 0$.

We now check that the function $*_p$: $L \to 2^L$ satisfies each of the AGM postulates (K$*$1) through (K$*$8) with respect to $K = B(p)$. Some general remarks before the details:

- The AGM postulates for revision were first formulated in Gärdenfors 1984 and a convenient overview may be found in Peppas 2007, whose presentation we follow. We note in passing that the classic account in Alchourrón, Gärdenfors and Makinson 1985 focused on contraction, and its axiomatization of revision contains a confusion: it omits postulate (K$*$3) below, and treats the definition of contraction from revision via the Harper identity as a postulate.

- We are not verifying satisfaction with respect to an *arbitrary* belief set K, but with respect to a *specific* belief set depending on the choice of p, namely $K = B(p) = \{x: p(x,\mathsf{T}) = 1\}$. This specification is needed for (K$*$3) and (K$*$4),

though not for the other postulates, where *K* does not appear in unrevised form.

- Our result corrects the claim made by Spohn 1986 and 2009 that the AGM postulates correspond to the Popper axioms for conditional probability. When *Cn* is understood as classical consequence, the specific Hosiasson-Lindenbaum axiom (HL) (see section 2.2) is needed to ensure that the function $*_p\colon L{\to}2^L$ satisfies the AGM postulate (K∗5), as noted in the verification below.

(K∗1): *K∗a = Cn(K∗a)*. *Verification*: Same as the above for *B(p) = Cn(B(p))*), but replacing ⊤ by *a*.

(K∗2): *a ∈ K∗a*. *Verification*: We need *p(a,a)* = 1, immediate from axiom (vF2).

(K∗3): *K∗a ⊆ Cn(K∪{a})*. *Verification*: Suppose *y* ∈ LHS, so that *p(y,a)* = 1. We need to show that *y ∈ Cn(K∪{a}) = Cn(B(p)∪{a}) = Cn({x: p(x,⊤) = 1}∪{a})*, so it suffices to show that ¬*a*∨*y* ∈ {*x: p(x,⊤)* = 1}, i.e. that *p*(¬*a*∨*y*,⊤) = 1. Now *p*(¬*a*∨*y*,⊤) = *p*(¬*a*∨(*a*∧*y*),⊤) = *p*(¬*a*,⊤)+*p*(*a*∧*y*,⊤). But *p*(*a*∧*y*,⊤) = *p*(*a*,⊤)·*p*(*y,a*) = *p*(*a*,⊤) since by supposition *p(y,a)* = 1. Thus *p*(¬*a*∨*y*,⊤) = *p*(¬*a*,⊤)+*p*(*a*,⊤) = *p*(⊤,⊤) = 1 as desired.

(K∗4): *Cn(K∪{a}) ⊆ K∗a* whenever *a* is consistent with *K*. *Verification*: Suppose *y ∈ Cn(K∪{a})* and *a* is consistent with *K*; we need to show *p(y,a)* = 1. By the first supposition, ¬*a*∨*y ∈ Cn(∧xᵢ: i ≤ n)* for some $x_1,.., x_n \in K = B(p)$ with each *p(xᵢ,⊤)* = 1, so that *p*(∧*xᵢ*,⊤) = 1 and thus *p*(¬*a*∨*y*,⊤) = 1. Hence *p(a,⊤)* = *p*(*a*∧(¬*a*∨*y*),⊤) = *p*(*a*∧*y*,⊤). But also we have *p*(*a*∧*y*,⊤) = *p*(*a*,⊤)·*p(y,a)*. Putting these together, *p(a,⊤)* = *p*(*a*,⊤)·*p(y,a)*. But by supposition, ¬*a ∉ K = B(p)* so *p*(¬*a*,⊤) ≠ 1 so *p(a,⊤)* ≠ 0, so by arithmetic *p(y,a)* = 1 as desired.

(K∗5): *K∗a* is consistent whenever *a* is consistent. *Verification*: Suppose *a* is consistent; we need *p*(⊥,*a*) = 0, which is immediate given the distinctive axiom for for HL functions.

(K∗6): If *a ≈ b* then *K∗a ≈ K∗b*. *Verification*: Suppose *a ≈ b*; we need *p(x,a)* = 1 iff *p(x,b)* = 1, again immediate.

(K∗7): *K∗(a∧b) ⊆ Cn((K∗a)∪{b})*. *Verification*: Suppose *x* ∈ LHS, so that *p(x,a∧b)* = 1. It suffices to show that ¬*b*∨*x ∈ K∗a*, i.e. that *p*(¬*b*∨*x,a*) = 1. When *a* is abnormal, this is immediate, so suppose that *a* is normal. From the supposition, *p*(¬*b*∨*x, a*∧*b*) = 1. Now *p*(*b*∧*x,a*) = *p*(*b*∧(¬*b*∨*x*),*a*) = *p*(*b,a*)·*p*(¬*b*∨*x,a*∧*b*) = *p*(*b,a*)·1 = *p(b,a)*. Since *a* is normal, the left projection of *p* from a is a proper Kolmogorov function, so we may conclude that *p*(*b*∧¬*x, a*) = 0 and thus *p*(¬*b*∨*x, a*) = 1 as desired.

(K∗8): *Cn((K∗a)∪{b}) ⊆ K∗(a∧b)* whenever *b* is consistent with *K∗a*. *Verification*: Suppose that *y* ∈ LHS and *b* is consistent with *K∗a*; we need to show that *p(y,a∧b)* = 1. By the second supposition, ¬*b ∉ K∗a* so *p*(¬*b,a*) ≠ 1 and thus *a* is normal and moreover *p(b,a)* ≠ 0. By the first supposition, ¬*b*∨*y ∈ K∗a*, i.e. *p*(¬*b*∨*y,a*) = 1. Hence *p(b,a)* = *p*(*b*∧(¬*b*∨*y*),*a*) = *p*(*b*∧*y,a*). But also *p*(*b*∧*y,a*) = *p*(*b,a*)·*p(y,a∧b)*. Putting these together, *p(b,a)* = *p*(*b,a*)·*p(y,a∧b)*. Since as noted *p(b,a)* ≠ 0, arithmetic gives us *p(y,a∧b)* = 1 as desired.

Failure of injectivity

28

For the failure of injectivity it suffices to find two distinct HL functions $p \neq p'$ with $*_p = *_{p'}$, i.e. with $*_p(a) = \{x: p(x,a) = 1\} = \{x: p'(x,a) = 1\} = *_{p'}(a)$ for all $a \in L$, i.e. with $p(x,a) = 1$ iff $p'(x,a) = 1$ for all $a,x \in L$. For simplicity we do this with Boolean algebras rather than propositional languages. Take any finite Boolean algebra with $n \geq 2$ atoms, and two distinct probability distributions f,f' to these atoms with each atom getting a non-zero probability; extend them to one-place probability functions (for simplicity using the same names) on the entire algebra. Noting that every non-zero element of the algebra receives a non-zero probability under each of these functions, we can define two-place functions p,p': $L^2 \rightarrow [0,1]$ by the ratio rule for non-zero right arguments and putting $p(x,0) = p'(x,0) = 1$. These are HL probability functions, in fact they are Carnap functions. Then for all a,x we have $p(x,a) = 1$ iff $p(a \wedge x) = p(a)$ iff $a \leq x$ and likewise for p', and so $p(x,a) = 1$ iff $p'(x,a) = 1$ as desired.

Surjectivity in the finite case

We now show that the map is surjective for consistent belief sets and under the condition of finiteness (i.e. that the propositional language has only finitely many mutually non-equivalent formulae). That is, in such a language, for every belief set K and every revision operation $*$: $L \rightarrow 2^L$ satisfying the AGM postulates with respect to K, there is a HL function p: $L^2 \rightarrow [0,1]$ with $* = *_p$ and such that if K is consistent then $K = B(p)$.

The construction is quite straightforward. Given $*$ and K, we define p: $L^2 \rightarrow [0,1]$ as follows:

- In the limiting case that a is inconsistent, put $p(x,a) = 1$ for all $x \in L$

- In the principal case that a is consistent, put $p(x,a)$ to be the proportion of $(K*a)$-worlds that are x-worlds.

Here, a *world* is a maximal consistent set of formulae, and an *X-world*, for $X \subseteq L$, is a world Y with $X \subseteq Y$. We need to show that (1) p satisfies the HL axioms, (2) $* = *_p$, and (3) if K is consistent then $K = B(p)$.

For (1) it is convenient to check the HL axioms in the form given to them by Hosiasson-Lindenbaum 1940 (see appendix to section 2.2), as follows.

(HL1) $p(x,a) = 1$ whenever $x \in Cn(a)$. *Verification*: If a is inconsistent then we have $p(x,a) = 1$ by the definition for that case, so we may suppose that a is consistent. By AGM, $a \in K*a$ so if $x \in Cn(a)$ we have $x \in Cn(K*a) = K*a$. Thus when $x \in Cn(a)$, all $(K*a)$-worlds are x-worlds, i.e. the proportion of $(K*a)$-worlds that are x-worlds is 1, so $p(x,a) = 1$ as required.

(HL2) $p(x \vee y,a) = p(x,a)+p(y,a)$ whenever a is consistent and $\neg(x \wedge y) \in Cn(a)$. *Verification*: Suppose a is consistent and $\neg(x \wedge y) \in Cn(a)$. By the first supposition, we need to consider proportions, and by the second the proportion of $(K*a)$-worlds that are $(x \vee y)$-worlds is the sum of the proportions of $(K*a)$-worlds that are, separately, x-worlds or y-worlds, and we are done.

(HL3) $p(x \wedge y, a) = p(x,a) \cdot p(y, a \wedge x)$. *Verification*: If a is inconsistent then so is $a \wedge x$ and LHS = 1 = RHS. Suppose a is consistent. If $a \wedge x$ is inconsistent then LHS = 0 while RHS = $0 \cdot 1 = 0$ and again we are done. If $a \wedge x$ is consistent then LHS is the proportion of $(K*a)$-worlds that are $(x \wedge y)$-worlds, while RHS is the proportion of $(K*a)$-worlds that are x-worlds multiplied by the proportion of $(K*a \wedge x)$-worlds that are y-worlds. If x is inconsistent with $K*a$ then both LHS and RHS equal 0, so we may suppose that x is consistent with $K*a$. Then by AGM axioms (K*7) and (K*8) the $(K*a \wedge x)$-worlds are just the $(K*a)$-worlds that are x-worlds. Hence RHS is the proportion of $(K*a)$-worlds that are x-worlds multiplied by the proportion of *those* that are y-worlds, which equals the proportion of $(K*a)$-worlds that are $(x \wedge y)$-worlds, equalling the LHS and we are done.

(HL4) $p(x,a) = p(x,b)$ whenever $a \approx b$. *Verification*: If a is inconsistent then so is b, so LHS = 1 = RHS. If a is consistent, then if $a \approx b$ the a-worlds are just the b-worlds, and the proportion of a-worlds that are x-worlds is the same as the proportion of b-worlds that are x-worlds.

To show that (2) $* = *_p$, consider first the principal case that a is consistent, where we need only note that by the definition of $*_p$ we have $x \in *_p(a)$ iff $p(x,a) = 1$ while, by the definition of p, also $p(x,a) = 1$ iff every $(K*a)$-world is an x-world, i.e. iff $x \in Cn(K*a) = K*a$. In the limiting case that a is inconsistent, $p(x,a) = 1$ for every x and by the AGM postulates, $x \in K*a$ for every x, so again we are done.

Finally, we check (3) that if K is consistent then $K = B(p)$. For this, we need only show that $x \in K$ iff $p(x, \top) = 1$. But if K is consistent, the AGM postulates tell us that $K = K*\top$, and the equivalence $p(x,a) = 1$ iff $x \in K*a$ just established may be applied substituting $\top$ for a, completing our proof.

We conjecture that surjectivity fails in the infinite case. Evidently its present proof breaks down there, since one cannot meaningfully speak of proportions of infinite sets, thus blocking the definition of $p(\cdot, \cdot)$ above. Nor is it possible to repair the proof by replacing proportionality by some probability distribution that gives each world a non-zero value. For if the set of formulae is countable, there are continuum many worlds and as is well known, there is no probability distribution on a non-countable set that gives a non-zero value to each element.

One may wonder whether it is possible to get rid of the condition that K is consistent when verifying (3). It can be done – at the cost of fiddling with the definitions of both AGM revision and HL functions for this case. When K is inconsistent, we need to take $K*a = K$, and consequently modify the AGM postulate K*5 to read, a little redundantly given (K*2): K is consistent iff both K and a are consistent. At the same time, we need to add the unit two-place function $1(\cdot, \cdot)$ to the class of HL functions, and consequently replace both the Popper and HL postulates by the following: p_a is a proper Kolmogorov function iff $p(\bot, \top) \neq 1$ and a is consistent. We omit verifications.

For Section 5.1. Hawthorne's system Q of Uncertain Inference

It is interesting that some of Hawthorne's postulates are echoed in the theory of data mining, in axioms for redundancy among association rules, despite the fact that these

deal with a context without any propositional connectives. See Balcázar 2010 for more information.

For Section 5.2. Proto-probability Functions

Disjunctive interpolation

As remarked in the text, the principle of disjunctive interpolation is essentially a restriction to the finite case of a rule known as conglomerability, due to Seidenfeld et al 1998. That rule says that if $r \leq p(y|a_i) \leq s$ for every cell a_i of a partition $\{a_i: i \in I\}$ of the probability space, then $r \leq p(y) \leq s$. This principle is uncontroversial in the finite case, following from the Kolmogorov axioms, but poses difficulties in the case that I is countable, where it can fail unless countable additivity is assumed. Our disjunctive interpolation is the anodyne case that $\#(I) = 2$, from which of course the other finite cases may be obtained by induction.

One half of our rule was articulated and discussed by Koopman 1940, 1940a under name of 'alternative presumption'. Recall that disjunctive interpolation states that $p(x,a) \leq p(x,a \vee b) \leq p(x,b)$ whenever $p(x,a) \leq p(x,b)$ and $\neg b \in Cn(a)$. Koopman's rule of alternative presumption says that $p(x,a) \leq p(y,c)$ whenever both $p(x,a \wedge b)$, $p(x,a \wedge \neg b) \leq p(y,c)$. If one assumes that the order $\leq$ is complete (as we do, although Koopman does not), alternative presumption is in fact equivalent to the right half of disjunctive interpolation (i.e the assertion that $p(x,a \vee b) \leq p(x,b)$ under the same conditions).

To obtain Koopman, suppose both $p(x,a \wedge b) \leq p(y,c)$ and $p(x,a \wedge \neg b) \leq p(y,c)$. By the completeness of $\leq$, either $p(x,a \wedge b) \leq p(x,a \wedge \neg b)$ or conversely. In e.g. the former case we have by the right part of disjunctive interpolation that $p(x,(a \wedge b) \vee (a \wedge \neg b)) \leq p(x,a \wedge \neg b)$, so by the second supposition with right extensionality and transitivity of $\leq$, we are done. In the converse direction, suppose $p(x,a) \leq p(x,b)$ and $\neg b \in Cn(a)$, we want to show that $p(x,a \vee b) \leq p(x,b)$. We need only note that $p(x,(a \vee b) \wedge b) = p(x,b)$ and, since $\neg b \in Cn(a)$, also $p(x,(a \vee b) \wedge \neg b) = p(x,a) \leq p(x,b)$, so we can apply Koopman to get $p(x,(a \vee b)) \leq p(x,b)$.

Disjunctive interpolation may also be seen as extracting the qualitative content of Gärdenfors' principle (P*M) in section 5.8 of his 1988. That principle may be formulated in the language of two-place conditional probability, as follows: $p(x,a \vee b) = p(x,a) \cdot k + p(x,b) \cdot (1-k)$ where $k = p(a,a \vee b)$, whenever $\neg b \in Cn(a)$. Disjunctive interpolation follows immediately. For if $\neg b \in Cn(a)$ and $p(x,a) \leq p(x,b)$ then by (P*M) $p(x,a \vee b) \leq p(x,b) \cdot k + p(x,b) \cdot (1-k) = p(x,b) \cdot (k+(1-k)) = p(x,b)$ and likewise $p(x,a \vee b) \geq p(x,a) \cdot k + p(x,a) \cdot (1-k) = p(x,a) \cdot (k+(1-k)) = p(x,a)$.

Actually, Gärdenfors presented his principle (P*M) in terms of operations $*$ that revise one-place probability functions: $p*(a \vee b) = (p*a) \cdot k + (p*b) \cdot (1-k)$ where $k = (p*(a \vee b))(a)$, whenever $\neg b \in Cn(a)$. The two versions translate directly.

Verification that Q is proto-probabilistically sound

We check that when we take any proto-probability function $p(\cdot,\cdot)$ and $t \in D$, and define a relation by putting $a \mathrel{|\!\sim_{pt}} x$ iff $p(x,a) \geq t$, then $\mathrel{|\!\sim_{pt}}$ satisfies all the rules of Q. For (O1) we need $p(a,a) \geq t$, which is immediate from (P1). For (O2), we need that when $p(x,a) \geq t$ and $y \in Cn(x)$ then $p(y,a) \geq t$, which is immediate from (P2) and transitivity. For (O3), we need that when $p(x,a) \geq t$ and $a \approx b$ then $p(x,b) \geq t$, which is immediate from (P3). For (O4), we need that when $p(x \wedge y,a) \geq t$ then $p(y,a \wedge x) \geq t$, which is immediate from (P4) and transitivity. For (O6) alias WAND, we need that when $p(x,a) \geq t$ and $p(y,a \wedge \neg y) \geq t$ then $p(x \wedge y,a) \geq t$. If $t = 0_D$ then this is immediate, and if $t \neq 0_D$ it is given by (P6).

It remains to obtain (O5) and the non-Horn rule NR of negation rationality, which we do from the two parts of (P5), making use of the completeness of the relation $\leq$. Suppose for both that $\neg b \in Cn(a)$. For (O5) we need that when $p(x,a) \geq t$, $p(x,b) \geq t$ then $p(x,a \vee b) \geq t$. Since the order on D is complete, either $p(x,a) \leq p(x,b)$ or conversely; consider e.g. the former. Then by the left part of (P5), $p(x,a) \leq p(x,a \vee b)$ and we are done by transitivity of $\leq$. For NR we need to show that when $p(x,a \vee b) \geq t$ then either $p(x,a) \geq t$ or $p(x,b) \geq t$. Since the order on D is complete, either $p(x,a) \leq p(x,b)$ or conversely, consider e.g. the former. Then by the right part of (P5), $p(x,a \vee b) \leq p(x,b)$ so by transitivity $p(x,b) \geq t$ as desired.

Note that the only conditions on $\leq$ that are used in this verification are those imposed: transitivity, completeness, largest and least elements. In particular, we did not need anti-symmetry.

The modularity of this verification shows that it can also be run for subsystems. In particular, if on the semantic side we omit the conditions (P4) and (P6) we can still obtain the postulates of system Q that do not refer explicitly to conjunction, i.e. Q less (O4: VCM) and (O6: WAND). Perhaps more interestingly, when we omit condition (P5) (and this time also, if desired, the requirement that $\leq$ is complete) we can still obtain the postulates of Q that do not refer to disjunction, i.e. Q less (O5: XOR) and (NR). Representation theorems then follow by the same construction as used in the text. Such systems may be worth exploring further, but we do not pursue the matter here.

Verification that van Fraassen functions satisfy the proto-probability conditions

The verifications of (P1) through (P4) are trivial; we give those for (P5) and (P6). The latter is the shorter. Suppose $p(y,a \wedge \neg y) \neq 0$; we want to show $p(x,a) = p(x \wedge y,a)$. If p_a is the unit function this is immediate, so suppose that it is a proper Kolmogorov function, so that $p(y \wedge \neg y,a) = 0$. Now by (vF3), $p(y \wedge \neg y,a) = p(\neg y,a) \cdot p(y,a \wedge \neg y)$ so one of the two factors is zero, so by the initial supposition the left one is, so in turn $p(x \wedge \neg y,a) = 0$. By (vF2) we have $p(x,a) = p(x \wedge y,a) + p(x \wedge \neg y,a)$ so $p(x,a) = p(x \wedge y,a)$ as desired.

For (P5), we have already observed that it follows immediately from Gärdenfors' principle (P*M), which is known to follow from the van Fraassen axioms. For a direct verification, however, we can argue as follows. Suppose $\neg b \in Cn(a)$ and $p(x,a) \leq p(x,b)$; we want to show that $p(x,a) \leq p(x,a \vee b) \leq p(x,b)$. In the case that $p_{a \vee b}$ is the unit function we have the left inequality and moreover, since the set of abnormal elements

of D is an ideal (see section 2.2 and its appendix), p_b is also the unit function giving us also the right one. So suppose that $p_{a\vee b}$ is not the unit function, and so is a proper Kolmogorov function; we show first the right inequality. By (vF2), $p(x,a\vee b) = p(x\wedge a,a\vee b) + p(x\wedge\neg a,a\vee b)$. By the product rule (vF3) the left summand equals $p(a,a\vee b)\cdot p(x,a) \leq p(a,a\vee b)\cdot p(x,b)$ since $p(x,a) \leq p(x,b)$, while the right one equals $p(\neg a,a\vee b)\cdot p(x,\neg a\wedge b) = p(\neg a,a\vee b)\cdot p(x,b)$ since $\neg b \in Cn(a)$. Putting these together, $p(x,a\vee b) \leq [p(a,a\vee b)\cdot p(x,b)] + [p(\neg a,a\vee b)\cdot p(x,b)] = p(x,b)$ by arithmetic distribution and (vF2) again. A similar argument (interchanging a and b and converting $\leq$) gives us $p(x,a\vee b) \geq p(x,a)$, and we are done.

Note that in this verification we have appealed to all three of the axioms of van Fraassen (the first one implicitly, the other two explicitly). If we simply drop any one of these three axioms, we allow in non-proto-probability functions. However, there might be interesting ways of weakening the second or third of the van Fraassen axioms that leave us within the class.

Verification of closure property of the class of all proto-probability functions

Conditions P1–P4 and P6 are immediate; P5 is a little less so, as follows. Suppose $\neg b \in Cn(a)$ and $hp(x,a) \leq' hp(x,b)$; we need to show that $hp(x,a) \leq' hp(x,a\vee b) \leq' hp(x,b)$. If $p(x,a) \leq p(x,b)$ we have $p(x,a) \leq p(x,a\vee b) \leq p(x,b)$ since p is a proto-probability function, and we need only apply order-preservation. Otherwise, by completeness of $\leq$ we have $p(x,b) \leq p(x,a)$ so, again since p is a proto-probability function, $p(x,b) \leq p(x,b\vee a) = p(x,a\vee b) \leq p(x,a)$ and so by order-preservation $hp(x,b) \leq' hp(x,a\vee b) \leq' hp(x,a)$. So by the supposition $hp(x,a) \leq' hp(x,b)$, transitivity of $\leq'$ gives us $hp(x,a) \leq' hp(x,a\vee b) \leq' hp(x,b)$ as desired.

For Section 5.3. Comparison with Plausibility Measures, and Further Examples

Comparison of proto-probability functions with Halpern's conditional plausibility measures

Comparison is rendered tricky by the fact that there are inter-connected differences regarding the *relation* over the target set, the *domain* of the function, and the *regularities* imposed on the function itself. For the relation $\leq$ over the target set D, we have required it to be transitive and connected (thus also reflexive) but not necessarily anti-symmetric, while Halpern constrains it to be a partial ordering (reflexive, transitive and anti-symmetric) but not necessarily connected. Regarding the domain, a trivial difference is that conditional plausibility measures are defined in the field-of-sets rather than propositional mode; this is essentially a matter of presentation which we can ignore. However, Halpern also follows Rényi in allowing that the right argument need not range over the whole of the field. To be sure, we can complete the right domain by giving the function value 1 for the omitted right argument values, but that forces reconsideration of Halpern's first axiom which, in the propositional mode, says that $p(\bot,a) = 0$. To maintain consistency, that axiom needs to be restricted to 'normal' values of a, i.e. those whose left projection is not the one-place unit function. That done, if we confine attention to those relations $\leq$ on the target set that satisfy both sets of conditions, i.e. to linear relations, then Halpern's conditions are considerably weaker than ours, in that neither P5 nor P6 is required, and only part of P4, giving him a broader class.

Acknowledgements

References

Alchourrón, Carlos, Peter Gärdenfors and David Makinson 1985. 'On the logic of theory change: partial meet contraction and revision functions' *The Journal of Symbolic Logic* 50: 510-530.

Arló Costa, Horacio 2001. 'Bayesian epistemology and epistemic conditionals: on the status of export-import laws' *The Journal of Philosophy* 98: 555-593.

Arló Costa, Horacio and Rohit Parikh 2005. 'Conditional probability and defeasible inference' *Journal of Philosophical Logic* 34: 97-119.

Balcázar, José L. 2010. 'Redundancy, deduction schemes, and minimum-size bases for association rules' *Logical Methods in Computer Science* 6: 1-33.

Bohlmann, G. 1909. 'Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung' pages 244-278 of *Atti del IV Congresso Internazionale dei Matematici, Roma 6-11 Aprile 1908* (Roma: Accademia dei Lincei), vol. III.

Boutilier, Craig 1995. 'On the revision of probabilistic belief states' *Notre Dame Journal of Formal Logic* 36: 158-183.

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

Carnap, Rudolf 1952. *The Continuum of Inductive Methods* (Chicago University Press).

Carnap, Rudolf 1971. 'A basic system of inductive logic' pages 33-165 of Carnap and Jeffries eds *Studies in Inductive Logic and Probability* (University of California Press).

de Finetti, Bruno 1974. *Theory of Probability* (New York: Wiley).

Dubins, Lester E. 1975. 'Finitely additive conditional probabilities, conglomerability and disintegrations' *Annals of Probability* 3: 89-99.

Dubois, Didier and Henri Prade 1998. 'Possibility theory: qualitative and quantitative aspects' pages 169-226 of D.M. Gabbay and P.Smets eds *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol I*' (Dordrecht: Kluwer).

Gärdenfors, Peter 1984. 'The dynamics of belief as a basis for logic' *British Journal for the Philosophy of Science* 35: 1-10.

Gärdenfors, Peter 1986. 'The dynamics of belief: contractions and revisions of probability functions' *Topoi* 5: 29-37.

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

Hájek, Alan 2003. 'What conditional probability could not be' *Synthese* 137: 273-323.

Hailperin, Theodore 1988. 'The development of probability logic from Leibniz to MacColl' *History and Philosophy of Logic* 9: 131-191.

Hailperin, Theodore 1991. 'Probability logic in the twentieth century' *History and Philosophy of Logic* 12: 71-110.

Halpern, Joseph 2001. 'Plausibility measures: a general approach for representing uncertainty' *Proceedings of the 17th Joint Conference on AI (IJCAI 2001)* (San Francisco: Morgan Kaufmann).

Halpern, Joseph (to appear). 'Lexicographic probability, conditional probability, and nonstandard probability' *Games and Economic Behaviour*. Meanwhile, see http://www.c.s.cornell.edu/home/halpern/papers/lex.pdf

Harper, W.L. 1975. 'Rational belief change, Popper functions and counterfactuals' *Synthese* 30: 221-262. Reprinted with minor editorial changes in pages 73-115 of W.L. Harper and C.A. Hooker eds *Foundations of Probability, Statistical Inference, and Statistical Theories of Science, Vol I* (Dordrecht: Reidel).

Harper, W.L. 1976. 'Rational conceptual change' pages 462-494 of *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1976, Vol. II* (University of Chicago Press).

Hawthorne, James 1996. 'On the logic of nonmonotonic conditionals and conditional probabilities' *Journal of Philosophical Logic* 25: 185-218.

Hawthorne, James and David Makinson 2007. 'The quantitative/qualitative watershed for rules of uncertain inference' *Studia Logica* 86: 249-299.

Hosiasson-Lindenbaum 1940. 'On confirmation' *The Journal of Symbolic Logic* 4: 133-148.

Koons, Robert 2009. 'Supplement to Defeasible Reasoning' *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu

Koopman, Bernard O. 1940. 'The axioms and algebra of intuitive probability' *The Annals of Mathematics* 41: 269-292.

Koopman, Bernard O. 1940a. 'The bases of probability' *Bulletin of the American Mathematical Society* 46: 763-774. Accessed 05 April 2009.

Leblanc, Hugues 1989. 'Popper's formal contributions to probability theory' pages 341-367 of M.A. Notturno ed. *Perspectives on Psychologism* (Leiden: E.J. Brill).

Leblanc, Hugues and Peter Roeper 1989. 'On relativizing Kolmogorov's absolute probability functions' *Notre Dame Journal of Formal Logic* 30: 485-512.

Lindström, Sten and Wlodzimierz Rabinowicz 1989. 'On probabilistic representation of non-probabilistic belief revision' *Journal of Philosophical Logic* 18: 69-101.

Makinson, David 1997. 'Screened revision' *Theoria* 63: 14-23.

Makinson, David 2005. *Bridges from Classical to Nonmonotonic Logic* (London: College Publications).

Makinson, David 2007. 'Completeness theorems, representation theorems: what's the difference?' In *Hommage à Wlodek: Philosophical Papers dedicated to Wlodek Rabinowicz*, ed. Rønnow-Rasmussen et al. www.fil.lu.se/hommageawlodek

Makinson, David (to appear). 'Logical questions behind the lottery and preface paradoxes: lossy rules for uncertain inference' *Synthese*, issue commemorating Henry Kyburg.

Mazurkiewicz, S. 1932. 'Zur Axiomatik der Wahrscheinlichkeitsrechnung' *Comptes rendus des séance de la Société des Sciences et des Lettres de Varsovie* (Sprawozdania z posiedzeń Towarzystwa Naukowego Warszawskiego) 25: 1-4.

McGee, Vann 1994. 'Learning the impossible' pages 179-199 of E. Eells and Brian Skyrms eds *Probability and Conditionals: Belief Revision and Rational Decision* (Cambridge University Press).

Paris, Jeff & Richard Simmonds, R. (2009). 'O is not enough' *Review of Symbolic Logic*, 2: 298-309.

Peppas, Pavlos 2007. 'Belief revision' chapter 8 of F. van Harmelen, V. Lifschitz, B. Porter eds *Handbook of Knowledge Representation* (Amsterdam: Elsevier).

Popper, Karl 1959. *The Logic of Scientific Discovery* second edition (New York: Basic Books).

Rényi, Alfréd 1955. 'On a new axiomatic theory of probability' *Acta Mathematica Academiae Scientiae Hungaricae* 6: 268-335.

Renyi, Alfréd 1970. *Foundations of Probability Theory* (San Francisco: Holden-Day).

Rényi, Alfréd 1970a. *Probability Theory* (Amsterdam: North-Holland).

Roeper, Peter and Hugues Leblanc 1999. *Probability Theory and Probability Logic* (Toronto: University of Toronto Press).

Seidenfeld, T, M.J. Schervish and J.B. Kaldane 1998. 'Non-conglomerability for finite-valued, finitely additive probability' *Sankya* Series A 60.3: 476-491.

Simmonds, Richard 2010. 'On Horn closure conditions for probabilistic consequence relations' (PhD thesis, University of Manchester).

Spohn, Wolfgang 1986. 'The representation of Popper measures' *Topoi* 5: 69-74.

Spohn, Wolfgang 2009. 'A survey of ranking theory', pages 185-228 of Franz Huber et al eds *Degrees of Belief* (New York: Springer).

Stalnaker, R.C. 1970. 'Probability and conditionals' *Philosophy of Science* 37: 64-80, reprinted in pages 107-128 of W.L. Harper et al eds *Ifs* (Dordrecht: Reidel, 1981).

van Fraassen, Bas 1976. 'Representation of conditional probabilities' *Journal of Philosophical Logic* 5: 417-430.

van Fraassen, Bas 1995. 'Fine-grained opinion, probability, and the logic of full belief' *Journal of Philosophical Logic* 24: 349-377.

Weydert, Emil 1994. 'General belief measures' *Uncertainty in Artificial Intelligence 94*: 575-582 (San Francisco: Morgan Kaufmann).

Department. of Philosophy, Logic and Scientific Method
London School of Economics
Houghton Street, London WC2A 2AE, United Kingdom
david.makinson@gmail.com

Note: This version contains some minor corrections and additions to the text as published