UNIVERSITY OF LEEDS

This is a repository copy of *Probability and nonclassical logic*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/104891/

Version: Accepted Version

Book Section:

Probability and nonclassical logic

J. Robert G. Williams

July 22, 2015

Contents

Classical tautologies have probability 1. Classical contradictions have probability 0. These familiar features reflect a connection between standard probability theory and classical logic. In contexts in which classical logic is questioned—to deal with the paradoxes of self-reference, or vague propositions, for the purposes of scientific theory or metaphysical anti-realism—we must equally question standard probability theory.[1]

Section 1 covers the intended interpretation of 'nonclassical logic' and 'probability'. Section 2 reviews the connection between classical logic and classical probability. Section 3 briefly reviews salient aspects of nonclassical logic, laying out a couple of simple examples to fix ideas. Section 4 explores modifications of probability theory. The variations laid down will be motivated initially by formal analogies to the classical setting. In section 5, however, we look at two foundational *justifications* for the presentations of 'nonclassical probabilities' that are arrived at. Sections 6-7 describe extensions of the nonclassical framework: to conditionalization and decision theory in particular. Section 8 will consider some alternative approaches, and section 9 evaluates progress.

1   Preliminaries

Our topic is the interaction between nonclassical logic and probability. But 'nonclassical logic' and 'probability' in what sense?

In the following sections, we operate with a fairly narrow understanding of non-classicality. For present purposes a nonclassical logic is one that diverges from classical orthodoxy on which arguments (sequents of sentences) or inferential rules are valid or invalid. An example of a classically valid argument might be disjunctive syllogism: $A, \neg A \vee B \models B$. A rule of inference (in the technical sense that contrasts with this) involves a transition from one validity to another. Conditional proof is an example: this tells us that if $A, B \models C$ then also $A \models B \to C$. Nonclassical logics might declare one or the other, or both, invalid. For our purposes, the sentences in question can (usually) be thought of as drawn from a standard propositional language containing negation, disjunction, conjunction and a conditional.

A second sense of nonclassicality pertains to semantics. A theory is nonclassical in this sense if it diverges from classical orthodoxy on what truth statuses there are or how they can be distributed. Classical semantics endorses bivalence—every meaningful sentence is either true or false. A nonclassical semantics may allow for sentences which are neither true nor false; or for intermediate *degrees* of truth. This is not nonclassicality in logic strictly speaking; but the two forms of nonclassicality are intimately related and I will discuss both.[2]

---

[1]For the paradoxes of self-reference, (Field, 2008) provides a recent survey of nonclassical approaches. For vagueness, see inter alia (Williamson, 1994; Keefe, 2000; Smith, 2008). (Hughes, 1992) is a relatively accessible approach to the issues surrounding quantum logic, with chs.7 and 8 particularly pertinent to our concerns. For metaphysical anti-realism and logic, a locus classicus is Dummett (1991).

[2]Supervaluationism is perhaps the leading example of a nonclassical semantics that is paired with what might be argued to be a classical logic. Supervaluational semantics allows for truth value gaps. But, as standardly presented, across a standard propositional (or indeed quantificational) language, the associated 'global' supervaluational logic coincides with classical consequence. The issue is subtle, however. The supervaluational multiconclusion consequence relation diverges from the classical analogue. And across a minimally enriched language (including an object-language truth or definiteness operator) classical inferences rules such as conditional proof fail (Williamson,

A broader reading of 'non-classical logic' would include the logic of operators and connectives beyond the usual propositional list. Modal, temporal and conditional logics are paradigms of this kind of departure from classicality.[3] This broad reading won't be the focus of our discussion, but the interested reader is directed to the result from (Paris, 2001) quoted in section 3. To a first approximation, the result shows that unless we have nonclassicality in the narrow sense, the theory of probability is unchanged (a remarkable result in its own right!).

Like nonclassicality, what we call 'probability' can vary along several dimensions. The first dimension of variation is the *kind* of phenomenon in question—perhaps rational belief, or objective chance, or degree of evidential confirmation.[4] In the sections below we focus on a *subjective* interpretation of probability. On this picture, subjects have beliefs that come in degrees. The *probabilist* then maintains that to be ideally rational, the distribution of these degrees of belief must be *probabilistic*—i.e. satisfy the probability axioms. Alternative interpretations are considered in the penultimate section of this survey.

A second dimension of variation concerns the items that probabilities attach to. We can choose between investigating probabilities that attach to fine-grained entities such as sentences; or alternatively coarse grained entities like sets of outcomes ('events'). Here I take the fine-grained approach. Indeed, I will mostly talk of probabilities attaching to sentences. One advantage, against a standard background on which logical and alethic properties attach to sentences, is that we can cleanly formulate principles that connect logic and probability without worrying about the relationship between the relata of the logical consequence relation and the bearers of probability. We can simply say, for example, that if a sentence is tautologous, its probability must be 1. On the other hand, given the commitment to a *subjective* interpretation of probability the choice may seem odd. If ideal degrees of belief have to be probabilistic, it seems this requires the objects of propositional attitudes to be sentences—while believers in 'mentalese' should be happy with this, most others will not.

But there isn't a deep worry here. Suppose you hold that objects of attitudes are Fregean thoughts or Russellian structured propositions. You can then straightforwardly adapt the discussion below to your preferred setting. You already owe an account of the logic and truth-conditions of your favoured truth-bearers (and typically this can be a straightforward adaption of the usual treatment of the logic and semantics for sentences (cf. e.g. Soames, 1989)). This could be classical or nonclassical. That your truth bearers plausibly have their truth-conditions essentially doesn't prevent us from describing unintended interpretations and using them to characterize a logic in the usual model-theoretic way. The logic-probability connections appropriate to such settings will be a straightforward transcription of the sentence-based formulations below. (The real issue here is whether the *motivations* for nonclassicality extend to the propositional level. Some hold that propositional truth-conditions are broadly classical, with nonclassicality arising from the sentence-proposition relation. A case in point are treatments of reference failure which make some sentences truth-value gaps, but only because the sentences express no proposition—the propositions themselves remaining bivalent.)

1994, ch.5). Supervaluational logic is a genuinely hard case to categorize (cf. Williams, 2008).

[3]For an example of this usage of 'nonclassical', and an introduction to non-classical logics in both the narrow and broad sense, see (Priest, 2001).

[4]Cf. Hájek (Summer 2012), and the chapters in this volume on *The classical interpretation and indifference principles*, *Frequentism*, *The propensity interpretation*, *Best system approaches to chance* and *Subjectivism*.

One radical and minority view on the objects of belief is the 'ultra coarse-grained' treatment of propositions argued for by Lewis (1986) and Stalnaker (1984). In one version, the propositions we take attitudes to are identified with sets of possible worlds—the possible worlds at which the sentence is true. This does motivate a rather different view of the relation between logic and probability—one to which we return in the penultimate section of the paper.

2   The classical framework

Consider a colour swatch, Patchy. Patchy is borderline between red and orange. The classical law of excluded middle requires the following be true:

>   (LEM) Either Patchy is red, or Patchy is not red.

Many regard (LEM) as implausible for borderline cases like Patchy—intuitively there is no fact of the matter about whether Patchy is red or not, and endorsing (LEM) suggests otherwise. This motivates the development of nonclassical logic and semantics on which (LEM) is no longer a logical truth.[5] But if one doubts (LEM) for these reasons, one surely cannot regard it is a constraint of rationality that one be certain—credence 1—in it, as classical probabilism would insist. One does not have to be a convinced revisionist to feel this pressure. Even one who is (rationally) agnostic over whether or not logic should be revised in these situations, and so has at least some inclination to doubt LEM, should not accept that non-probablistic belief-states are irrational.[6]

We can view the distinctively classical assumptions embedded in standard probability theory from at least two perspectives. First, the standard *axiomatization* of probability (over sentences) makes explicit appeal to (classical) logical properties. Second, probabilities can be identified with convex combinations or expectations of *truth values* of sentences, where those 'truth values' are assumed to work in classical ways. We briefly review these two perspectives below in the classical setting, before outlining in the next section how these may be adapted to a nonclassical backdrop.

The following is a standard set of axioms for probability over sentences in the propositional language L:[7]

P1c. (Non-negativity) $\forall S \in L, P(S) \in \mathbb{R}_{\geq 0}$

---

[5]The literature on this topic is vast. Two representatives of the contemporary debate are (Wright, 2001) and (Smith, 2008). Williamson (1994) is the most influential critic of nonclassical approaches in this area.

[6]The connection between logic and probability in these contexts is a major theme of Hartry Field's work in recent times. See Field (2000, 2003b,a, 2009).

[7]An alternative approach to axiomatizing probability, starting from suggestions by Popper, dispenses with the appeal to consequence, and works directly on constraints on the interaction of probability with connectives. One appealing feature of this is that one could then use probability functions so characterized as a resource for characterizing consequence. This approach has been vigorously pursued, and there are a number of extensions to nonclassical settings, such as intuitionism. See (Roeper & Leblanc, 1999) for a survey of both classical and nonclassical work in this tradition. The focus on 'purely logical' axiomatizations below is in a sense the dual of this approach.

P2c. (Normalization) For T a (classical) logical truth, $P(T) = 1$

P3c. (Additivity) $\forall R, S \in L$ with R and S (classically) inconsistent, $P(R \vee S) = P(R) + P(S)$.

Various theorems of this system articulate further relations between logic and probability:

P4c. (Zero) For F a (classical) logical falsehood, $P(F) = 0$;

P5c. (Monotonicity) If S is a (classical) logical consequence of R, $P(S) \geq P(R)$;

Believers in the so-called *Regularity* constraint on probability functions endorse yet more logical constraints on probability. They endorse the converse of (Normalization) and (Zero), saying that *only* logical truths/falsehoods take extremal probability values. I won't discuss those views further here.

(Normalization) is problematic for the logical revisionist who seeks to deny the law of excluded middle: under our interpretation of probability, it says that rational agents must be fully confident in each instance of excluded middle. But it is not the only problematic principle. Advocates of some popular nonclassical settings say that (LEM) is true, but assert the following Truth Value Gap claim:

(TVG) It's not true that Patchy is red, and it's not true that Patchy is not red.

On this—supervaluation-style—nonclassical setting, a disjunction can be true, even though each disjunct is untrue.[8] This motivates allowing high confidence in 'either Patchy is red or Patchy isn't red', and yet ultra-low confidence in each disjunct. But this violates (Additivity).[9]

Still other, dialethic, nonclassical settings allow contradictions to be true. Let L be the liar sentence ("this sentence is not true"). Some argue that the following holds:

(TC) $L \wedge \neg L$

Advocates of this view presumably have reasonably high confidence in (TC). But (Zero) rules this out.[10]

---

[8]Some further details are given later. For supervaluations, see inter alia (van Fraassen, 1966; Fine, 1975; Keefe, 2000).

[9]Compare (Field, 2000). I note that sometimes, it is assumed that a 'supervaluational style' approach motivates not *low* credence but an *imprecise* credence. This is an illustration of a theme I will emphasize later—that formally similar systems may allow for multiple interpretations (so 'supervaluationism' may very well pick out not a single system, but multiple such. The results below will show, however, how well the low-confidence model fits with standard supervaluational rhetoric, including the identification of truth with supertruth, and the appeal to global supervaluational consequence as the preferred consequence relation.

[10]See (Priest, 2006), who explicitly discusses the modifications of standard probability theory required to accommodate paraconsistent logics.

(Monotonicity) inherits problems both from (Zero) and (Normalization). Since $A \vee \neg A$ classically follows from anything, (Monotonicity) tells us that rational confidence in excluded middle is bounded below by our highest degree of confidence in anything. And since $L \wedge \neg L$ classically entails anything, (Monotonicity) tells us that rational confidence in the conjunction of the liar and its negation is bounded above by our lowest degree of confidence in anything. But revisionists object: one such revisionist thinks we should have higher confidence that hands exist (for example), than in sentence (LEM). Another thinks we should have lower confidence in the moon being made of green cheese than we do in the conjunction of the liar and its negation.

Finally, what of (Non-negativity)? Many revisionists will find this unproblematic; notice that it doesn't appeal to (classical) logical relations explicitly at all. But the assumption that (subjective) probabilities are non-negative *real numbers* builds in, inter alia, that rational degrees of belief are *linearly ordered*. It's not crazy to question this assumption in a nonclassical setting. To take one example: MacFarlane (2010) argues that certain graphs, rather than point-like credences, capture our doxastic states in the nonclassical setting he considers.

The obvious moral from this brief review is that it would be madness for a logical revisionist to endorse *as articulating rationality constraints on belief* a probability theory that is based on the 'wrong' logic. A natural thought is to generalise the axiomatizations by switching out the appeal to classical consequence in favour of one's favoured non-classical consequence. This is indeed the core of the approach explored below—but notice it cannot always be the whole story. The problems for (Additivity) in the supervaluational setting arise even though the relevant sentences remain inconsistent.

We turn now from axiomatics to the second perspective. This connects probabilities not to logic but directly to truth values. We presuppose an 'underlying' credence function on a maximally fine-grained partition of possibilities ("worlds"). For simplicity, we take this to be finite. The only constraints imposed on this underlying credence is that the total credence invested across all possibilities sums to 1, and that the credence in a given possibility is never less than 0. Sentences are true or false relative to these worlds. Let $|S|_w$ be a function that takes value 1 iff S is true at w, and value 0 iff S is false at w—this we call the *truth value* of the sentence at w. Each underlying division of credence c then allows us to define a function from sentences to real numbers:

$$f(S) := \sum_w c(w)|S|_w$$

It turns out that such fs are exactly the probabilities over sentences. Some terminology: when a function f is a 'weighted average' of functions g_i, with weights given by coordinates $\lambda_i \geq 0$ (with the λ_i summing to 1), we say that f is a *convex combination* of the g_i. Letting the worlds w play the role of the indices i, and setting $g_w = |S|_w$ and $\lambda_i = c(w)$, the equation above meets this condition. Classical probabilities are convex combinations of classical truth values. We can also think of the probability of S so characterized as the *expectation* of S's truth value, relative to the underlying credence defined over worlds. [11] We'll call this the

---

[11]For those familiar with probability theory: we treat the worlds as the sample space for the probability function c, and then for any sentence S consider a random variable $t(S)$ whose value at w is equal to $|S|_w$. Where there are

convex-combination characterization of probability.

In characterizing probability this way, the association of 1s and 0s with truth statuses truth and falsity is crucial. *The True* and *the False* can't themselves be arithmetically manipulated; whereas the arithmetical manipulations of 1 and 0 make perfect sense. So why call these 'truth values' (Howson, 2008)? The answer I will explore—and extend to the nonclassical case—is that this representation is justified only because they are the degree of belief that omniscient agents should invest in *S*, in situations where *S* has that truth status. They reflect the omniscience-appropriate cognitive states; the 'cognitive loading' of the classical truth statuses.

Since convex combinations of (classical) truth values lead to our familiar probability functions, all the problematic consequences for the logical revisionist arise once again. The revisionist faced with the convex-combination characterization of probabilities will pinpoint the appeal to *classical* truth value distributions as what causes the trouble. Faced with classical axiomatics, the natural strategy is to consider revised principles appealing to a nonclassical consequence relation. Faced with the convex-combination characterization, the natural strategy is to explore variations where nonclassical *truth value distributions* are appealed to.

3   Nonclassical logic and semantics

Nonclassical logic and semantics comes in a wild and wonderful variety.[12] Although the results to be discussed shortly will apply to a large variety of settings, including those with (for example) infinitely many truth values, to fix ideas I set out a three-valued setting that allows us to characterize a handful of sample logics. A *Kleene truth status assignment* involves, not a scattering of two statuses (Truth, Falsity) over sentences, but a scattering of three—call them for now T, F and O. The distribution over compound sentences must accord with the (strong) Kleene truth-tables for negation, conjunction and disjunction:

A	$\neg A$
T	F
O	O
F	T

$A \wedge B$	T	O	F
T	T	O	F
O	O	O	F
F	F	F	F

$A \vee B$	T	O	F
T	T	T	T
O	T	O	O
F	T	O	F

(In the last two tables, the horizontal headers represent the truth status of *A*, and the vertical headers the truth status of *B*, and corresponding entry the resultant truth status of the complex sentence.) We have various options for characterizing logical consequence on this basis:

---

only finitely many worlds, the full technology of a worldly probability space isn't needed. But if there are infinitely many worlds we can still appeal to expectations of truth value, relative to the underlying credence.

   [12]For a general introduction to nonclassical logics, including the Kleene logic and LP discussed below, see (Priest, 2001) and for further philosophical discussion, see (Haack, 1978).

Kleene logic:	$A \vdash_K B$ iff on every Kleene truth status assignment, if A is T, then B is T too.
LP:	$A \vdash_L B$ iff on every Kleene truth status assignment, if A is T or O, then B is T or O too.
Symmetric logic:	$A \vdash_S B$ iff on every Kleene truth status assignment, if A is T, then B is T; and if A is T or O, then B is T or O.

However we characterize consequence, logical truths (tautologies) are those sentences that are logical consequences of everything; logical falsehoods are those sentences of which everything is a logical consequence; an inconsistent set is a set of sentences of which everything is a logical consequence. The strong Kleene logic is a simple example of a nonclassical logic where excluded middle is no tautology: if A has status O, then so will $\neg A$, and looking up the truth table above, so will $A \vee \neg A$. But then this provides a Kleene-logic countermodel to the claim that $A \vee \neg A$ follows from everything, since any case where B has value T and $A \vee \neg A$ value O is a countermodel to $B \vdash_K A \vee \neg A$. By contrast, excluded middle will be a tautology on the LP understanding of consequence. $A \vee \neg A$ can never have the status F; and that suffices to ensure it follows from everything on the LP definition. But LP provides us with a simple example of a paraconsistent logic—one on which explicit 'contradictions' $L \wedge \neg L$ do not 'explode'—they do not entail everything. The symmetric characterization has both features—contradictions are not inconsistent/explosive and excluded middle is no tautology.

What shall we make of these Ts, Fs and Os? In the classical setting, we ordinarily assume that (in context) sentences are true or false *simpliciter*—that these are monadic properties that sentences (in context) either possess or fail to possess. Truth status distributions represent possible ways in which such properties can be distributed. We could regard the Kleene distributions in the same way. The picture would then be that rather than two monadic alethic properties, there are three; but we can still ask about what the *actual* distribution is, and about the nature of the properties so distributed. Perhaps such information would motivate one choice of logic over another. A nonclassical logic motivated this way we call *semantically driven*.

But one needn't buy into this picture, to use the abstract three-valued 'distributions' to characterize the relations $\vdash_K$, $\vdash_L$ and $\vdash_S$. Hartry Field has argued for such a non-semantically-driven approach to logic in recent times. Semantics for Field does not involve *representing* real alethic statuses that sentences possess. It is rather an *instrumental device* that allows us to characterize the relation that is of real interest: logical consequence (Field, 2009, passim). He doesn't propose that we eliminate truth-*talk* from our language—he favours a deflationarist approach to the notion—but such talk is not supposed to describe a range of 'semantic values' that sentences possess. For Field, the Ts, Fs and Os can remain uninterpreted, since they're merely an formal tool used to describe the consequence relation. And the question of which of these categories a sentence like (LEM) falls into would simply be nonsense.

Let's suppose that we do not go Field's way, but take our nonclassical logic to be semantically driven, so that sentences have categorical properties corresponding to (one of) T, F and O. What information would we like about these statuses, in order to further understand the view being put forward? Consider the classical case. Here the statuses were Truth and Falsity; and these statuses were each 'cognitively loaded': we could pinpoint the ideally appropriate attitude to adopt to each. In the case of a true sentence this was full belief (credence 1); and in

the case of a false sentence, utter rejection (credence 0). We'd like to know something similar about the nonclassical statuses T, F and O. If S has status O, should an omniscient agent invest confidence in S? If so, to what level? Would they instead suspend judgement? Or feel conflicted? Or groundlessly guess?

Call a semantics *cognitively loaded* when each alethic status that it uses is associated with an 'ideal' cognitive state. Nonclassicists endorsing a semantically-driven conception of logic may still not regard the underlying semantics as cognitively loaded. For example: Maudlin (2004) advocates a nonclassical three valued logic (the Kleene logic, in fact) in the face of semantic paradoxes, but explicitly *denies* that there is any cognitive loading *at all* to the middle status O. Indeed, he thinks that the distinctive characteristic of O that makes it a 'truth value gap' rather than a 'third truth value', is that it gives no guidance for belief or assertion.

The nonclassical logics we will focus on will be semantically driven, cognitively loaded, and further, will be loaded with cognitive states of a particular kind: with standard degrees of belief, represented by real numbers between 1 (full certainty) and 0 (full rejection, anti-certainty). This last qualification is yet another restriction. There's no a priori reason why the cognitive load appropriate to nonclassical statuses shouldn't take some other form—calling for some non-linear structure of degrees of belief, or suspension rather than positive partial belief, or some such. Such views motivate more radical departures from classical probabilities than the ones to be explored below.[13]

Consider the following three loads of Kleene distributions (numerical values represent the degree of belief that an omniscient agent should adopt to a sentence having that status):

Status:	T	O	F
Kleene loading:	1	0	0
LP loading:	1	1	0
Symmetric loading:	1	$\frac{1}{2}$	0

The loads differ on the attitude they prescribe for O under conditions of omniscience: utter rejection, certainty, or half-confidence respectively. They motivate informal glosses on this truth status: respectively neither true nor false; both true and false; or half-true. Furthermore, the loads correspond systematically to the logics mentioned earlier: in each case, logical consequence requires that there be no possibility of a drop in truth *value*, where the truth value is identified with the cognitive load of the truth status. (In the special case where the loads are simply 1 and 0, this corresponds to the familiar distinction between 'designated' and 'undesignated' truth statuses, and the characterization of consequence as preservation of designated status (cf. Dummett, 1959, e.g.).)

We continue to use the three Kleene-based logics as worked examples. But there are many, many ways of setting up nonclassical logics. So long as the logics are semantically driven, and truth statuses are cognitively loaded with real values, then our discussion will cover them.

---

[13]Three potential examples of this are Wright's notion of a *quandary* (Wright, 2001); Macfarlane's credence profiles (MacFarlane, 2010) and whatever we should take to be the appropriate response to the partially ordered values of (Weatherson, 2005).

4   Probability, truth values and logic

Cognitive loads give a natural way to extend the convex-combination characterization of probability.[14] Recall the classical case: for an appropriate c, the probability of each S must satisfy:

$$P(S) = \sum_w c(w)|S|_w$$

Consider the limiting case where c is zero everywhere but the actual world (i.e. conditions of 'credal omniscience') The above equation then simplifies to $P(S) = |S|_w$. Under conditions of omniscience, the subjective probability matches the numerical value assigned as S's truth value; hence, that number will be the cognitive load of the truth status. In this way, the Kleene, LP and Symmetric loads induce three kinds of 'nonclassical probabilities', as convex combinations of the respective truth values.

A nice feature of this approach is that the axiomatic perspective generalizes in tandem with the convex-combination one.[15] Consider the following principles, for parameterized consequence relation $\vdash_x$:

P1x. (Non-negativity) $\forall S \in L, P(S) \in \mathbb{R}_{\geq 0}$

P2x. (Normalization) If $\vdash_x T$, then $P(T) = 1$

P3x. (Additivity) $\forall R, S \in L$ such that $R, S \vdash_x, P(R \vee S) = P(R) + P(S)$

P4x. (Zero) If $F \vdash_x$, then $P(F) = 0$;

P5x. (Monotonicity) If $R \vdash_x S$, then $P(S) \geq P(R)$;

If we pick the Kleene loads, then these five principles are satisfied by any 'nonclassical probability' (expectation of truth value), so long as we use the Kleene logic (set $x = K$). Mutatis mutandis for the LP and Symmetric loads and logics.

It's useful to add two further principles—extensions and variations on (Additivity)—which are also satisfied by convex combinations of Kleene truth values:

P6x. (IncExc) $\forall R, S \in L, P(R) + P(S) = P(R \vee S) + P(R \wedge S)$

P7x. (Dual additivity) $\forall R, S \in L$, if $\vdash_x R \vee S$, then $P(R) + P(S) - 1 = P(R \wedge S)$

---

[14]See in particular (Paris, 2001) for this strategy. Compare also (Zadeh, 1968) and (Smith, 2010).

[15]For instances of this observation in specific settings, see e.g. (Weatherson, 2003; Field, 2003b; Priest, 2006). As we shall see, (Paris, 2001) gives a particularly elegant treatment of many cases.

In the presence of (Zero) and (Normalization) respectively, plus the assumption that the conjunction of an inconsistent pair is a logical falsehood, (IncExc) will entail the original (Additivity) and (Dual Additivity). (Additivity) itself is weak in logics with few or no inconsistencies, such as *LP*; if there are no inconsistent pairs of sentences, then the antecedent is never satisfied, and the principle becomes vacuously true. Dual Additivity is correspondingly weak in logics with few or no tautologies, such as the Kleene logic. The weaknesses are combined in the Symmetry logic. But their generalization, (IncExc), makes no mention of the logical system in play, and so retains its strength throughout.

The connection between convex-combination and logical characterizations illustrated above is very general. Scatter truth statuses over sentences howsoever you wish, with whatever constraints on permissible distributions you like. Make sure you associate them with real-valued 'cognitive loads'—degrees of belief within $[0,1]$, so that we can straightforwardly define the notion of possible expected truth value, by letting $|S|_w$ be equal to the cognitive load of the status that S has at w. We consider the following logic:

No drop:
  $A \vdash B$ iff on every truth status assignment w, $|A|_w \leq |B|_w$.

It's straightforward to check that (P1x), (P2x), (P4x) and (P5x) will then hold of all the expected truth values.

The status of (Additivity) and its variants is more subtle. These principles make explicit mention of a particular connective, so it's no surprise that whether or not they hold depends how those connectives behave. (IncExc) will hold iff we have the following:[16]

$$|A|_w + |B|_w = |A \vee B|_w + |A \wedge B|_w$$

Classical logic, and many nonclassical logics, satisfy this principle. But we cannot assume this holds generally.

In the classical setting, we had more than just a grabbag of principles satisfied by probabilities: we had an axiomatization complete with respect to classical expected truth values. An obvious question is whether some subset of nonclassical variants is complete with respect to nonclassical expected truth values in a similar way.

Paris (2001) delivers an elegant result on this front. Among much else of interest, he shows that the nonclassical verisons of (Normalization), (Zero), (Monotonicity) and (IncExc) deliver complete axiomatizations of a wide range of nonclassical probabilities. The conditions for this result to hold are that: (i) truth values (in our terminology: the cognitive loads of truth statuses) are taken from $\{0, 1\}$; (ii) $A \vdash_k B$ is given the 'no drop' characterization mentioned earlier; and (iii) the following pair is satisfied:

---

[16]The 'if' direction follows by the linearity of convex combinations. The 'only if' direction holds by considering the special case of probability where the underlying credence all lies on a single world, c, and hence the probability coincides with truth values.

$$(\mathcal{T}2) \quad |A|_w = 1 \wedge |B|_w = 1 \quad \Longleftrightarrow \quad |A \wedge B|_w = 1$$
$$(\mathcal{T}3) \quad |A|_w = 0 \wedge |B|_w = 0 \quad \Longleftrightarrow \quad |A \vee B|_w = 0.$$

This applies, for example, to the Kleene and LP loads mentioned above, as well as the original classical case. Its application goes well beyond this: for example, to appropriate formulations of intuitionistic logic.[17]

(A side note: this is the theorem that delivers a direct extension of probability theory to many settings that are nonclassical in the 'broader' sense discussed in the introduction, for example, ones that contain modal, temporal and conditional operators. The standard semantics for such settings will satisfy (i-iii); and the treatment of conjunction and disjunction satisfies $\mathcal{T}2, 3$.)

Beyond this, it is a matter of hard graft to see whether similar completeness results can be derived for settings that fail the Parisian conditions (one representative of which is our Symmetric logic). Drawing on the work of Gerla (2000) and Di Nola *et al.* (1999), Paris shows that a similar result holds for a finitely valued (Łukasiewicz) setting and Mundici (2006) later extended this to the continuum-valued fuzzy setting.[18]

We have already mentioned supervaluational logics. These are widely appealed to in the philosophical literature. They arise as a generalization of classical truth values, via the assumption that the world and our linguistic conventions settle, not a *single* intended classical truth value assignment over sentences, but a set of co-intended ones. Sentences are supertrue if they are true on all the co-intended assignments, and superfalse if they are false on all of them. This allows supertruth gaps: cases where the assignments for S differ, and so it is neither supertrue nor superfalse. We shall assume that supertruth has a cognitive loading of 1, and other statuses have a loading of 0 (compare the Kleene loading earlier). The no-drop logic is then so-called global supervaluational consequence, $\vdash_s$.

This articulation of supervaluationism delivers the results mentioned earlier. For example, as a classical tautology, (LEM) is true on each classical assignment, and a fortiori true on the set of co-designated ones, so it will always be supertrue (value 1). But this is compatible with each disjunct being a supertruth gap (value 0). Invest credence in a world where this is the case, and the credences in the disjuncts can be 0 while credence in their disjunction is 1. (Additivity) and (IncExc) fail.

Axiomatizing the convex-combinations of supervaluational truth values is achieved by a theorem that Paris gives, drawing on the work of Shafer (1976) and Jaffray (1989). For the propositional language under consideration, the results show that convex combinations of such truth values are exactly the *Dempster-Shafer belief functions*. These may be axiomatized thus:

---

[17]For the intuitionistic case, compare Weatherson (2003). Paris reports the general result as a corollary of a theorem of Choquet (1953).

[18]The major difference between the 3-valued Kleene based setting and the Łukasiewicz settings is the addition of a stronger conditional—and this is crucial to the proofs mentioned. It's notable that Paris provides axiomatizations not in terms of a 'no drop' logic, but in terms of the logic of 'preserving value 1'. This is possible because the 'no drop' consequence is effectively encoded in the 1-preservation setting via tautological Łukasiewicz conditionals.

$$
\begin{aligned}
(\mathcal{DS}1) \quad & \vdash_s A \quad && \Rightarrow \quad P(A) = 1 \\
& A \vdash_s \quad && \Rightarrow \quad P(A) = 0 \\
(\mathcal{DS}2) \quad & A \vdash_s B \quad && \Rightarrow \quad P(A) \leq P(B) \\
(\mathcal{DS}3) \quad & P(\textstyle\bigvee_{i=1}^{m} A_i) \geq \sum_S (-1)^{|S|-1} P(\bigwedge_{i \in S} A_i)
\end{aligned}
$$

(where S ranges over non-empty subset of $\{1, \ldots, m\}$).[19]

These kind of completeness results are often sensitive to the exact details of the sentences we are considering. We do *not* have a guarantee that the completeness result will generalize when we add expressive resources to the language. This is one reason why the earlier Paris result, which applies to *all* languages equipped with a semantics meeting the stated conditions, is so attractive.

($\mathcal{DS}$1-2) are simply the constraints (Normalization), (Zero) and (Monotonicity) that we met earlier. But $\mathcal{DS}$3 is something new. Sometimes called *subadditivity*, it is a new, weaker, member of the (Additivity) family. It's noticeable that what goes in place of (Additivity) that is varying from setting to setting, while other principles are held constant. Why is this?

Standard axiomatizations of probability feature principles of two kinds. The first are *purely logical*: they make no mention of specific logical connectives, but put constraints on probability in terms of the logical properties of sentences. (Normalization), (Zero) and (Monotonicity) are paradigms. The second kind of axioms are *immanent* to the nonclassical system, in that they impose constraints on sentences that involve particular logical connectives. Paradigms of this, imposing direct constraints on the distribution of probabilities over conjunctions and/or disjunctions, are (Additivity), (IncExc), (Dual Additivity) and ($\mathcal{DS}$3). Since the treatment of conjunction and disjunction can vary wildly from one nonclassical system to the next, one would not *expect* to find wholly general axiomatizations if one works with immanent axioms—one will have to indulge in case-by-case tailoring of the axioms to the particular system under investigation (or, as in Paris's result quoted earlier, impose general conditions on the semantics of the connectives that ensure that a particular immanent axiom is satisfied).

Are there purely logical axioms in the vicinity of the (Additivity) family? The following are promising. Say that Γ *x-partitions* a sentence S if the sum of truth values of the sentences in Γ always equals the truth value of S at any (nonclassical) x-world. And say that sets Γ and Δ are *x-recarvings* of one another if the sum of the truth values of sentences in Γ are always equal the sum of the truth values of sentences in Δ.[20] With this in hand, we can formulate the following purely logical constraint on probabilities:[21]

P8x. (Partition) If Γ is a set of sentences that x-partitions S, then $P(S) = \sum_{\gamma \in \Gamma} P(\gamma)$.

P9x. (Recarving) If Γ is a set of sentences that x-recarves Δ, then $\sum_{\delta \in \Delta} P(\delta) = \sum_{\gamma \in \Gamma} P(\gamma)$.

---

[19]Paris's initial formulation is slightly different, and uses classical logic (p.7), but as he notes this is extensionally equivalent to current version using the 'no drop' logic over the 'supervaluational' truth values (p.10).

[20]In the classical setting, this is a condition that (Joyce, 1998, 2005) calls 'isovalence'.

[21]The classical version of (Recarving) is a special case of the principle that Joyce (1998) calls 'Scott's axiom', tracing it to (Scott, 1964) and (Kraft *et al.* , 1959). The latter labeled the principle 'Generalized Additivity'.

It's easy to check that the Partition and Recarving Principles hold of all generalized probabilities.[22] Moreover (Partition) entails (Additivity) under Paris's assumptions $\mathcal{T}2$ and $\mathcal{T}3$, since R and S will then partition $R \vee S$ in the relevant sense. (Recarving) entails (IncExc) under the same assumptions, as they ensure that the set $\{R \vee S, R \wedge S\}$ recarves $\{R, S\}$.

(Partition) and (Recarving) neatly capture the *logical* structure that lies behind (Additivity), (IncExc) and the like. What additional power or generalizations one gains from the purely logical version of these axioms remains to be seen—their power depends on what partitions or recarvings are available in the particular nonclassical setting under investigation. A good test would be: can we somehow extract $\mathcal{DS}3$ in a supervaluational setting from these more abstract principles? A partial result is given in a footnote.[23]

While Paris-style completeness proofs are interesting and elegant, from a philosophical perspective, the identification of a reasonably rich body of principles that hold good of nonclassical probabilities is of philosophical interest even if we can't show them complete. Only the most radical Bayesians think that satisfying probabilistic coherence is *all* that there is to rationality; and so even if satisfying the axioms sufficed for probabilistic coherence, it would be contentious to conclude that it sufficed for rationality. On the other hand, so long as probabilistic coherence *is* a constraint on rational belief in the nonclassical setting, then what we learn from the above is that *violating* certain principles suffices for *irrationality*. A natural next question, therefore, is whether the 'nonclassical probabilities' that we have identified so far, have the same claim as classical probabilities in the classical setting to provide constraints on rational belief.

5   Foundational considerations

In many well-behaved nonclassical settings, we have seen a nice generalization of probability theory in prospect. But is this just coincidence? Or can we argue that this is the *right* way to theorize about subjective probabilities in such a setting?

I will focus on two arguments for 'probabilism' familiar from the classical case: the

---

[22]We argue for the recarving principle, of which the Partition Principle is a special case. Recall that an arbitrary generalized probability of any proposition is a convex combination of its truth values, say with parameters λ_w. Then $\sum_{\delta \in \Delta} P(\delta) = \sum_{\delta \in \Delta}[\sum_w \lambda_w |\delta|_w] = \sum_w \lambda_w[\sum_{\delta \in \Delta} |\delta|_w]$. By a parallel argument, $\sum_{\gamma \in \Gamma} P(\gamma) = \sum_w \lambda_w[\sum_{\gamma \in \Gamma} |\gamma|_w] =$. But by the assumption that Γ recarves Δ, we have for each w: $\sum_{\delta \in \Delta} |\delta|_w = \sum_{\gamma \in \Gamma} |\gamma|_w$. So the two sums are identical, and the identity betwen the probabilities is ensured.

[23]Suppose we are working with a language with a supervaluational semantics, which includes the supervaluational 'definitely' operator D. DS is false when S is false, and also when it is neither true nor false. Otherwise it is true. First, note that since the D operator 'screens off' the nonclassical behaviour of the sentences it attaches to, we can rerun a standard classical 'inclusion-exclusion' argument from the partition principle for the special case of D-prefixed sentences, obtaining $P(\bigvee_{i=1}^m DA_i) = \sum_S (-1)^{|S|-1} P(\bigwedge_{i \in S} DA_i)$ for S a non-empty subset of $\{1,\dots,m\}$. But it turns out in the supervaluational setting that an arbitrary conjunction $\bigwedge_{i \in S} DA_i$ is logically equivalent to the unprefixed $\bigwedge_{i \in S} A_i$. So by monotonicity twice, the RHS of the above can be written $\sum_S (-1)^{|S|-1} P(\bigwedge_{i \in S} A_i)$. On the other hand, $\bigvee_{i=1}^m DA_i$ certainly supervaluationally-entails $\bigvee_{i=1}^m A_i$, so by monotonicity we have the LHS bounded above by $P(\bigvee_{i=1}^m A_i)$. Putting these together, we have $\mathcal{DS}3$.

This result only holds for a language featuring the operator D, whereas Paris's completeness result was specific to a propositional language lacking such an operator. It's possible to investigate strengthened interpretations of probabilistic constraints that bridge this gap; but for reasons of space I won't explore this here.

Dutch-bookability of credences that violate the axioms of probability, and the accuracy-domination arguments advocated by Joyce (1998, 2009). Jeff Paris has shown how the first can be generalized, showing that credences that are not *convex combinations of truth value* in the relevant sense are Dutch-bookable, and conversely.(Paris, 2001). And for similar formal reasons, such credences are also 'accuracy dominated' (a converse is sometimes available).[24]

In the classical case, accuracy domination arguments consist in taking a belief state *b*, and assessing it at each world *w* for its degree of 'accuracy'. How accuracy should be measured is the leading issue for this approach; but in all cases, the starting point is to compare a degree of belief (within $[0, 1]$) to the 'truth value' of the sentence in question. But comparing a number with *Truth* or *Falsity* is not terribly tractable. So one standardly compares a given degree of belief with the cognitive loading of the truth status—how close, overall, the degrees of belief are to the 1s and 0s that an omniscient (perfectly accurate) agent would have in that world.

The result (relative to many plausible ways of measuring accuracy) is the following: if one's beliefs *b* are not (classical) probabilities, then it is always possible to construct a rival probabilistic belief state *c* such that *c* is more accurate than *b no matter which world is actual*. If such accuracy-domination is an epistemic flaw, then only probabilistic belief states can be flawless. This is offered as a rationale for why subjective *probabilities* in particular should be constraints on ideally rational belief.

What's important for our purposes is that the argument generalizes.[25] As earlier, suppose the cognitive loading of some nonclassical truth statuses are real numbers between $[0, 1]$. We use the very same accuracy measures as previously, to measure closeness of beliefs to these nonclassical truth values. And it turns out that if the belief state is not representable as a convex combination of truth values then it will be accuracy-dominated. The accuracy-based arguments for probabilism thus offer a justification for the claim that nonclassical probabilities, as characterized in the previous section, should indeed play the role of constraints on rational partial belief. (Of course, whether it's a good justification in either setting is contested. See (Hájek, 2008a).)

Perhaps the most familiar foundational justification for the claim that rational partial beliefs must be probabilistic comes from Dutch book arguments. The key claim is that one's degrees of belief are fair betting odds in the following sense: if offered a bet that pays £1 if *A*, and £0 if ¬*A*, then if you believe *A* to degree *k*, then you should be prepared to buy or sell the bet for *k*. Suppose that degrees of belief do play this role. Then if *b* is an improbabilistic belief state, there is a set of bets—a 'Dutch book'—such that you are prepared to buy each bet within it, but which ends up giving you a loss *no matter what*. Pragmatically viewed, a set of beliefs that open you up to sure-losses may seem flawed. Alternatively, one might think that the belief

---

[24]See (De Finetti, 1974) for the formal background to both results (in the latter case, with a quite different interpretation of its significance). (Williams, 2012a) examines the relation between the two results—it is essentially identical for the leading 'Brier Score' explication of accuracy. In that setting, and in others where accuracy is explicated by what are known as 'proper scoring rules', a converse to the accuracy-domination result is available; no probability will be accuracy-dominated. For discussion of the philosophical significance of converses to Accuracy Domination arguments, see (Williams, 2012b). A rather different connection between Dutch book foundations for probability and nonclassical (intuitionistic) logic is argued for in (Harman, n.d.).

[25]At least, the results in (Joyce, 1998) generalize. As discussed in (Williams, 2012b) the situation is much more complex for the main argument in (Joyce, 2009). The variation in proofs is significant, since different assumptions about the accuracy measure are involved in each case.

state is flawed because it commits you to viewing the book as both to be accepted (since consisting of bets that your belief state makes you prepared to accept) and also obviously to be rejected (since it leads to sure loss). So you are committed to inconsistency.

Dutch book justifications for probabilism, like accuracy domination arguments, are controversial (see (Hájek, 2008a,b) for a review and extension of criticisms). But independently of whether they persuade, are they adaptable to our case?

Suppose one has bought a bet that pays out 1 if A and 0 otherwise. If one is in a nonclassical setting, one can be faced with a situation where A takes some nonclassical truth status. The returns on such a bet then depends on how the bookie reacts. Call a real number $k \in [0,1]$ the *pragmatic loading* of a truth status X, just in case the right way for the bookie to resolve such a bet, given that A has status X, is to give the gambler £k.[26] Clearly the pragmatic loading of classical truth should be 1, and the pragmatic loading of classical falsehood is 0. Just as with cognitive loads of nonclassical truth statuses, there are many many ways one might consider assigning pragmatic loads (and just as with cognitive loads, there are pragmatic loads for truth statuses that don't fit the above description—the option of 'cancelling the bet' for example—as well as the option to deny that truth statuses *have* any identifiable pragmatic loading).

Suppose we have real-valued pragmatic loads for truth statuses, however. Then we can make sense of resolving bets in a nonclassical setting, and can consider what kinds of belief states are immune from Dutch books. Happily, the answer is just as you would expect: immunity from Dutch books is secured when (and only when) the belief state is a 'nonclassical probability'—a convex-combination of the relevant truth values.[27]

It's worth noting that in this last result, the 'truth value' of a sentence refers to the *pragmatic loading* of the relevant truth status, whereas in the previous results it referred to the *cognitive loading* of the truth statuses. If they differed, then we might have inconsistent demands—for example, if the cognitive loading of the 'other' status was 0.5 (omniscient agents are half-confident in A, when A is O), but its pragmatic loading was zero (one doesn't receive any reward for a bet on A, given that it is half-true) then being 0.5 confident in $A \land \neg A$ might be entirely permissible from the an accuracy-domination point of view, but one that makes you Dutch-bookable. The way to avoid this, of course, is to have cognitive and pragmatic loads coincide. It is interesting to speculate on *whether* they should coincide, and if so *why*. I can imagine philosophers taking cognitive value as primary, and arguing on this basis that the *right* way to resolve bets accords with the pragmatic loading; but I can equally envisage philosophers arguing that pragmatic loads are primary, and that these *give the reasons* why a particular cognitive loading attaches to a truth status. I can also imagine someone who takes both as coprimitive, but argued ('transcendentally') that they must coincide, otherwise rationality would place inconsistent demands on agents.

Both Dutch book and accuracy arguments—and much of the debate between their advocates and critics—can be replayed in a nonclassical setting. This should bolster our confidence that

---

[26]Compare the stipulations in (Paris, 2001) on the returns of bets in a nonclassical setting. On this description there's room for a kind of meta-uncertainty about what the pragmatic loading is—which could be modelled by allowing a wider class of 'truth value' distributions over worlds corresponding to all the possible pragmatic loading distributions the agent is open to. I won't explore this further here.

[27]The result follows from Dutch book arguments for expectations in (De Finetti, 1974) and is interpreted in the way just mentioned in (Paris, 2001). For more discussion, see (Williams, 2012a).

we have the *right generalization* of probability theory for the cases under study. And none of the results just mentioned make *any* assumptions about the particular kind of truth value distributions or logical behaviour of the connectives in question—other that the truth values, in the relevant sense, lie within $[0, 1]$. These are extremely general results.

6   Conditional probabilities and updating

Subjective probability without a notion of conditional probability would be hamstrung. If we are convinced (at least pro tem) that we have a nonclassical generalization of probability, then the immediate question is how to develop the theory of conditional probability within this setting. Three approaches suggest themselves. The first is simply to carry over standard characterizations of conditional probabilities, the ratio formula (restricted to cases where $P(A) \neq 0$; I often leave such constraints tacit in what follows):

$$P(B|A) := P(B \wedge A)/P(A)$$

The second is to investigate axiomatizations of probability in which conditional probability is the basic notion (of course, if left unchanged, these lead to classical probabilities). One investigates variations of these axioms, much as we did for the unconditional case above (compare Roeper & Leblanc, 1999). Third, we can look to the work we want conditional probability to do, and try to figure out what quantity is suited, within the nonclassical setting, to play that role. It is this third approach we adopt here, with a focus initially on the role of conditional probability in *updating* credences.

Conditional probabilities will be two-place functions from pairs of propositions to real numbers, written $P(\cdot|\cdot)$. The key idea will be that this should characterize an update policy: when one receives total information A, one's updated unconditional beliefs should match the old beliefs conditional on A: $P_{\text{new}}(\cdot) = P_{\text{old}}(\cdot|A)$. If updating on information isn't to lead us into irrationality, then a minimal constraint on conditional probabilities fit to play this role is that the result of 'conditioning on A' as above, should be a probability. (It turns out, incidentally, that straightforwardly transferring the 'ratio formula' treatment of conditional probabilities can violate this constraint.)[28]

Classical conditionalization on A can be thought of as the following operation: one first sets the credence in all $\neg A$ worlds to zero, leaving the credence in A-worlds untouched. This, however, won't give you something that's genuinely a probability (for example, the 'base credences' no longer sum to 1). So one renormalizes the credences to ensure we do have a probability, by dividing each by the total remaining credence $P(A)$.

We could generalize this in several ways, but here is the one we will consider. Take the first step in the classical case: we wipe out credence in worlds where the proposition is false (truth

---

[28]Suppose that we are working within the 'symmetric/half truth' nonclassical setting, suppose $P(A) = P(\neg A) = P(A \wedge \neg A) = 0.5$—which is certainly permitted by the relavant nonclassical probabiltiies. Now consider $P(A \wedge \neg A|A)$. By the ratio formula, this would be $P(A \wedge \neg A \wedge A)/P(A) = P(A \wedge \neg A)/P(A) = 0.5/0.5 = 1$. So $P_{\text{new}}(A \wedge \neg A) = 1$. But no probability (convex-combination of truth values) in this setting can have this exceed 0.5.

value 0) and leave alone credence in worlds where the proposition is true (truth value 1). Another way to put this is that the updated credence in w, $c_A(w)$ (prior to renormalization) is given by $c(w)|A|_w$: the result of multiplying the prior credence in w by the truth value of A at that possibility. Since we have real-valued truth values in our nonclassical settings, we simply transfer this across. The credence is scaled in proportion to *how true A* is at a given possibility. Renormalizing is achieved just as in the classical setting, by dividing by the prior credence in A.[29] Notice that by focusing on how the underlying credence c is altered under conditionalization, we have guaranteed that the function $P_A(X)$ defined by this procedure will be a convex combination of truth values, and so a nonclassical probability in our sense. We set $P(X|A) := P_A(X)$.

The characterization of the update procedure can be set down as follows:

$$P_A(X) = \sum_{w \in W} \frac{c_A(w)}{P(A)} |X|_w$$

Expanding the right hand side, this gives the following fix on conditional probability:

$$P(X|A) = \frac{\sum_{w \in W} c(w)|A|_w |X|_w}{P(A)}$$

Now, if we have a connective $\circ$ such that for arbitrary A and B, at any w:

$$|A|_w |B|_w = |A \circ B|_w$$

then it follows:

$$P(X|A) = \frac{\sum_{u \in W} c(u)|A \circ X|_u}{P(A)} = \frac{P(X \circ A)}{P(A)}$$

Certain nonclassical settings already contain the connective $\circ$—in the classical, supervaluational, Kleene and LP settings, $\wedge$ plays this role, and so the familiar ratio formula with relatively familiar conjunctive connectives is derived. A more exciting example is the main conjunctive connective of the product fuzzy logic (cf. Hájek, 1998) In other settings, it is a well defined truth function, but would require an extension of the language to introduce. But it is not automatic that the conditions for $\circ$ can be met by a truth function, in arbitrary nonclassical systems. Consider, for example, the Symmetric loading of the Kleene assignments. A sentence that has the truth status O gets the truth value 0.5. So $A \circ A$ would by construction have to take the truth value 0.25; $(A \circ A) \circ A$ would have to take the value 0.125, and so on. But in the symmetric setting, there are no truth statuses that have these loads. (A fortiori, $\circ$ is clearly *not* the Symmetric connective $\wedge$.) The process of conditionalization that was described works perfectly well in the Symmetric setting as a way to shift from one

---

[29] To see this, note that the result of the process is to give a 'base credence' over worlds which may add to less than 1. The sum total is given by $\sum_{w \in W} c_A(w) = \sum_{w \in W} c(w)|A|_w$. But by construction this is exactly $P(A)$. Hence dividing by $P(A)$ will renormalize the base credence, making it sum to 1, after the procedure described above.

nonclassical probability to another on receipt of the information that *A*. It's just that it doesn't have a neat formulation that mirrors the ratio formula.[30]

Much more on these nonclassical conditional probabilities in the fuzzy logic setting is available in (Milne, 2007, 2009)—who cites (Zadeh, 1968) as the source for the conception. Milne shows how to provide a synchronic Dutch book argument for this characterization of conditional probability, relative to the assumption (i) that conditional probabilities give fair betting odds for conditional bets; (ii) that nonclassical conditional bets are to be resolved a certain way (in particular, that they are 'progressively more and more called off' as the truth value of the condition gets lower and lower).[31] As Milne emphasizes, the assumption (ii) is crucial; in principle, there are many ways one might consider handling conditional bets in this setting, which would vindicate different conceptions of conditional probability. (Williams, 2012a) gives a nonclassical generalization of the Teller-Lewis diachronic Dutch book argument, but (although it gives us non-trivial information) it is even worse at giving leverage on the crucial case of conditionalizing on sentences that at some worlds take nonclassical values.

The real test for a proposed generalization of conditional probabilities lies in its applications—as an update procedure and elsewhere. To give a flavour of some important ways it generalizes classical conditional probability, we show how some key results generalize.

The analogue of Bayes' theorem is immediate:

$$P(A|B) = \frac{P(A \circ B)}{P(B)} = \frac{P(A \circ B)}{P(A)} \frac{P(A)}{P(B)} = P(B|A) \frac{P(A)}{P(B)}$$

Further key classical results also carry over:

A. **Lemma.** Assume that $\forall w, |\neg A|_w = |1 - A|_w$. Then $P(C) = P(C \circ A) + P(C \circ \neg A)$.

---

[30]Suppose that we have another connective $A \oplus B$ that is dual to $\circ$ in the following sense: $|A \oplus B|_w = |A|_w + |B|_w - |A \circ B|_w$. Then, by an earlier note, (IncExc) will hold for probabilities involving $\circ$ and $\oplus$ as the conjunction and disjunction symbols. It's worth noting that even if a setting has the $\circ$ connective, it needn't have its dual. In supervaluational logic, $\wedge$ coincides with $\circ$, but $\vee$ does not coincide with $\oplus$—as we can see by noting that $A \oplus \neg A$, unlike the supervaluational $\vee$, always takes truth value zero when both A and its negation have truth value zero. An alternative way to introduce a disjunction via $\circ$ is through the De Morgan identity: $|A \vee B|_w = |\neg(\neg A \circ \neg B)|_w$. In a supervaluational setting, this will be the normal supervaluational disjunction, for which we already know that (IncExc) does not hold. The general moral is that one can introduce an (IncExc) supporting dual of $\circ$, but there's no guarantee that it exists in a given setting; and one can find a corresponding notion of disjunction that is definable in any system with $\circ$ and $\neg$, but there's no guarantee that it is dual to $\circ$ in the way that (IncExc) requires.

[31]With classical presuppositions, the assumption is that a bet on *A* conditional on *C* with prize 1 and price β will return the prize if *AC* is the case; will return nothing if $A\bar{C}$ is the case; and the bet is called off (with a return of the initial stake) if the condition *C* is false. In the general case, we need to consider what happens to the bet in situations when *C* is partially true. Here is the stipulation: a conditional bet on *A* given *C* at prize 1, price β will have part of the stake is returned, and the potential prize decreased, in proportion to the falsity of *C*. Modulo this, the returns depend on *A*s truth value as in the categorical case. The overall return of the unit bet above is therefore $|C|_w(|A|_w - \beta)$ at w. The philosophical premise we need is that the fair price for a conditional bet so construed is exactly the conditional probability of *A* on *C*.

Proof. First note that relative to arbitrary w,

$$|C| = |C|(|A| + 1 - |A|) = |C|(|A| + |\neg A|) = |C||A| + |C||\neg A| = |C \circ A| + |C \circ \neg A|$$

For arbitrary nonclassical probability P, there's an underlying credence-over-worlds c such that $P(A) = \sum_w c(w)|A|_w$. So in particular

$$P(C) = \sum_w c(w)|C|_w = \sum_w c(w)(|C \circ A| + |C \circ \neg A|)$$

But this in turn is equal to:

$$\sum_w c(w)|C \circ A| + \sum_w c(w)|C \circ \neg A| = P(C \circ A) + P(C \circ \neg A)$$

as required.

B. **Corollary.** $P(C) = P(C|A)P(A) + P(C|\neg A)P(\neg A)$. Follows immediately from the above by the $\circ$-ratio formula for conditional probability.

Recall that Γ is a *nonclassical partition* if in each world, the sum of the truth values of the propositions in Γ is 1 (thus our assumption that $|\neg A| = 1 - |A|$ ensured that $A, \neg A$ was a partition). Then replicating the above proof delivers:

A. **Generalized Lemma.** $P(C) = \sum_{\gamma \in \Gamma} P(C \circ \gamma)$, so long as Γ is a partition.

B. **Generalized Corollary.** $P(C) = \sum_{\gamma \in \Gamma} P(C|\gamma)P(\gamma)$, so long as Γ is a partition.

It's nice to have this general form since there are some settings (supervaluational semantics for example) where the truth values of A and $\neg A$ don't sum to 1; the partition-form is still applicable even though the first result is not.

Another useful result is that, if P_C is the probability that arises from P by updating on C, then $P_C(A|B) = P(A|B \circ C)$. This follows straightforwardly from the ratio formula. For:

$$P_C(A|B) = P_C(A \circ B)/P_C(B) = \frac{\frac{P(A \circ B \circ C)}{P(C)}}{\frac{P(B \circ C)}{P(C)}} = \frac{P(A \circ B \circ C)}{P(B \circ C)} = P(A|B \circ C)$$

As an application, we can use the above results together with the basic convex-combination characterization of nonclassical probabilities to derive the following two 'expert principles' :[32]

$$P(S|t(S) = x) = x$$
$$P(S) = \sum_x x \cdot P(t(S) = x)$$

---

[32]Compare Reflection (van Fraassen, 1984) and the Principal Principle (Lewis, 1980). Recall that $|t(S) = x|_w = 1$ iff $|S|_w = x$; otherwise it takes value 0. The second displayed summation makes sense if x takes finitely many values in $[0,1]$; otherwise we will have to switch to integral formulations.

For the first, the characterization of conditional probabilities, and expansion by the convex combination characterization and the definition of $\circ$ gives:

$$P(S|t(S) = x) = \frac{P(S \circ t(S) = x)}{P(t(S) = x)} = \frac{\sum_w c(w)|S \circ t(S) = x|_w}{\sum_w c(w)|t(S) = x|_w} = \frac{\sum_w c(w)|S|_w|t(S) = x|_w}{\sum_w c(w)|t(S) = x|_w}$$

But $|t(S) = x|_w$ takes value 1 at worlds where the truth value of S is x, and 0 otherwise. So we can rewrite the above summing over those worlds where its truth value is indeed x. And of course, at all those worlds, $|S|_w = x$, from which we get our result:

$$= \frac{\sum_{w:|S|_w=x} c(w)|S|_w}{\sum_{w:|S|_w=x} c(w)} = \frac{\sum_{w:|S|_w=x} c(w)|S|_w}{\sum_{w:|S|_w=x} c(w)} = \frac{x \cdot \sum_{w:|S|_w=x} c(w)}{\sum_{w:|S|_w=x} c(w)} = x$$

The second of the two expert principles follows from the first by the generalized corollary above, applied to the partition given by all sentences of the form $t(S) = x$. The net result is a derivation of the 'expert' principle that, for rational believers, one's degree of belief in S should match one's expectation of its truth value. (In an earlier slogan, we used a similar 'expectational' gloss on nonclassical probabilities—but that required appeal to the underlying credence distribution over worlds. The result here is formulated purely in terms of the sentential probability P, and matches one standard classical way of calculating the expected value of a random variable t.)

This all looks promising. On the other hand, there are some surprising divergences from how we might expect a conditional probability to behave. Conditional probability so generalized does not guarantee that $P(A|A) = 1$. Consider the Symmetric loaded Kleene setting. Suppose all credence is invested in a world w in which A has truth value 0.5. It turns out by the recipe above that $P(A|A) = 0.5$.[33] This is surprising! Here is an application: probabilistic Independence is standardly defined as holding between A and B when $P(A|B) = P(A)$. But in the case just mentioned, $P(A|A) = 0.5 = P(A)$, so A will be probabilistically independent of itself. This shows that the epistemic significance of independence (so defined) and the use of conditional probabilities in confirmation theory more generally will need to be looked at carefully.[34]

7 Jeffrey-style Decision theory

An important application of probability is within the theory of rational decision making. We want to say something about a decision situation taking the following form: there are a range of actions A. There are factors $S \in \Gamma$, which fix the consequences of the action. Γ form a partition, and we are uncertain which element of that partition obtains. We are in a position to

---

[33]Multiplying the underlying credence by the truth value of A in any world other than w gives 0; at w it gives 0.5. Renormalizing takes this back up to 1—the underlying credence distribution is unchanged, so $P(A|A) = P_A(A) = P(A) = 0.5$.

[34]Thanks to Al Hájek for this example. Note, however, that even in the classical case, A can be probabilistically independent of itself if it has probability 1. Even in the classical case, this is a little strange—-what we see here is that in the nonclassical case the phenomenon cannot be confined to cases where A initially has an extremal value.

judge the desirability of the total course of events, representable by $A \wedge S$. But our uncertainty over which S obtains means that we have work to do to in order to figure out the desirability of A itself.

Jeffrey's decision theory (Jeffrey, 1965) provides a way to calculate the desirability D of a course of action, from the desirability of the outcomes, plus one's subjective probabilities. The desirability of the action is a weighted average of the desirability of the outcomes, with the weights provided by how likely the outcome is to obtain, given you take the action. The recipe is:

$$D(A) = \sum_{S \in \Gamma} P(S|A)D(A \wedge S)$$

Notice the crucial role given to conditional probabilities.

The application of probabilities within the theory of decision making is important, and if we couldn't recover a sensible account, this would render the whole enterprise of nonclassical (subjective) probability less interesting. As a proof of principle, I'll show that Jeffrey's recipe can indeed be generalized. I do not claim here to *justify* this as the *right* theory of decision in the nonclassical setting, but just to show that such theories are *available*.

Here's one way in which the Jeffrey decision rule can arise. Start by introducing a valuation function from worlds to reals, v—intuitively, a measure of how much we'd like the world in question to obtain. Then the desirability of an arbitrary *sentence A*, is defined as follows:

$$D(A) := \sum_{w} P(w|A)v(w)$$

Now, one might wonder if this is the *right* way to define desirability; but there is no question that it is *well-defined* in terms of the specific underlying valuation v. Now take any partition Γ of sentences (in the generalized sense of partition of the previous section). By the corollary in that section, applied to the nonclassical probability P_A that arises from conditioning on A, we have:

$$P(w|A) = P_A(w) = \sum_{S \in \Gamma} P_A(S)P_A(w|S)$$

Using another fact noted there, $P_A(w|S) = P(w|S \circ A)$, and putting these two together and substituting for $P(w|A)$ in the definition above, we obtain:

$$D(A) = \sum_{w}[\sum_{S \in \Gamma} P(S|A)P(w|S \circ A)]v(w)$$

Rearranging gives:

$$D(A) = \sum_{S \in \Gamma} P(S|A)[\sum_w P(w|S \circ A)v(w)]$$

But the embedded sum here is by construction equal to $D(S \circ A)$. Thus we have:

$$D(A) = \sum_{S \in \Gamma} P(S|A)D(S \circ A)$$

This is the exact analogue of the Jeffrey rule. So valuations over worlds allow us to define a notion of desirability that satisfies the generalized form of Jeffrey's equation.

What of the axioms for Jeffrey-Bolker decision theory? I won't investigate these here, but I will note that at least some require modification. For example, the 'averaging' axiom of Jeffrey-Bolker decision theory tells us that if A and B are inconsistent, then if $A \succ B$, then $A \succ (A \vee B) \succ B$. But on a supervaluational model of decison theoretic utility constructed as above, this can fail. Consider a situation where the value assigned to w and u is 1 in each case, and where $|A|_w = |\neg A|_w = 0$ and $|A|_u = 1, |\neg A|_u = 0$. In a supervaluational setting, $|A \vee \neg A|$ is 1 at any world. Then $D(A) = 1, D(\neg A) = 0$ but $D(A \vee \neg A) = 2$, a violation of the axiom. The underlying issue, I suspect, is that the axiom is written on the assumption that when A and B are mutually exclusive, they form a partition of $A \vee B$—which is perfectly true in a classical setting, but fails in the supervaluational setting. What can be read off the definitions above is that whenever (i) C is such that A and B form a partition of it (i.e. $\forall w, |C|_w = |A|_w + |B|_w$), and (ii) $A \circ C \succ B \circ C$, then we have $A \circ C \succ C \succ B \circ C$. This has the standard axiom as a special case, once classical assumptions are added.

8 Alternative approaches

Having explored the interaction of nonclassicality and probability under the understanding of those notions identified in section 1, this section takes a step back and considers briefly how things might look if we varied our starting assumptions.

The first variation we will consider concerns the items to which probabilities attach. We have been talking as if both probabilities and logical properties attach to sentences. We emphasized earlier that the focus on linguistic bearers is inessential—all the above could be transferred to probabilities and logic pertaining to Fregean thoughts, Russellian structures, and similar fine-grained 'propositions'. However, a very common approach to probability theory takes the objects of probabilities to be *coarse-grained*. For example, one finds a probability defined in terms of a triple (Ω, F, P), where Ω is a set (the 'sample space'), the event space F is an algebra of subsets of Ω, and P is a function from F to real numbers. We then have the familiar Kolmogorov axioms:

P1. (Non-negativity) $\forall E \in F, P(E) \in \mathbb{R}_{\geq 0}$

P2. (Normalization) $P(\Omega) = 1$

P3. (Additivity) $\forall D, E \in F$ with D and E disjoint, $P(D \cup E) = P(D) + P(E)$

On one reading, Ω could be the set of possible worlds, and then F would be coarse grained *propositions* in the sense of Lewis (1986) and Stalnaker (1984): sets of possible worlds. Lewis and Stalnaker argue that coarse-grained entities such as these are the objects of attitudes—a radical and minority position that means that even ordinary agents cannot take different attitudes to necessarily equivalent propositions. (An intermediate position here is to allow fine-grained entities as the objects of attitudes, so that different attitudes to necessarily equivalent claims are possible, but insist that *ideal* agents' degrees of belief across mentalese sentences should be such that it induces a probability function across the coarse-grained propositions such sentences express.) The interesting thing about this setting from our perspective is that *logic* seems to have disappeared from view. To be sure, we have analogues of 'inconsistency' (disjointness)—but the classicality is not explicit.

A natural diagnosis is that classical logic is tacitly built into this framework—by the assumption that the event space F forms a Boolean algebra. *Quantum* probabilities are developed exactly on this diagnosis on the (highly contentious) quantum logic approaches to quantum mechanics. Quantum events are held to form a non-distributive lattice, corresponding to the structure of subspaces of Hilbert space, rather than to the structure of a subsets of a classical sample space (Hughes, 1992, ch.8,9). Probabilities attaching to these quantum events will be distinctively 'nonclassical'.[35] More generally, to generate nonclassical probabilities we consider all sorts of alternative algebraic structures for the event space, and study analogues of the standard probability axioms in that setting. Algebraic logic provides a rich set of resources for those interested in pursuing this approach (Jansana, 2011). The variation of the algebra can be radical (replacing Boolean by Heyting algebras, for example) or it could be more minor (minor variations in the standard closure conditions of the algebra, for example).

In the discussions in previous sections, at various points I assumed an underlying space of 'worlds' at which sentences take truth values, one of which is 'actual', and determines a definite truth-value distribution. The final formulations are often statable without making this assumption (hence those formulations are available as a piece of formalism even to one who rejects the above), but it was freely used in their motivation and justification. It's worth bearing in mind that some of the more radical motivations for nonclassicism might make this assumption questionable. For example, the advocates of quantum logic interpret the Kochen-Specker theorems as ruling out this kind of picture; a Dummettian anti-realist about the past might find it hard to accommodate; and in his work developing a nonclassical logic to evade the liar paradox, Field (2008) specifically argues against the idea there is a definite distribution of actual (even nonclassical) truth values. Perhaps the algebraic approach, together with a philosophical interpretation of what the items in the 'event space' are to be, will seem to these theorists a more attractive philosophical foundation than the one sketched here.

We've just considered alternative approaches to our question that vary the objects of probabilities. Another alternative is to vary what 'probability' is to mean. Now, some interpretations of probability are tightly connected with epistemic issues in a way that allows

---

[35]The Hilbert space structure and the numbers assigned are common ground in quantum mechanics; but other approaches will offer alternative interpretations of the formalism—for example, identifying some Boolean subalgebra as the 'real' event space, with standard classical probabilities defined across it, with numbers assigned to points outside this subalgebra given an alternative physical intepretation.

much of the above discussion to go through. Evidential probabilities, as in (Williamson, 2000) are a case in point. Prima facie we expect that our evidence can discriminate between necessarily equivalent propositions. I may have strong evidence for the liquid in my glass being water, without having evidence that it is H_2O. This makes a fine-grained setting natural (though a coarse-grained theorist may respond instead by allowing 'possible worlds' into the sample space that are not *metaphysically* possible). The nonclassicist who rejects Patchy being red or not red certainly seems to need a treatment on which this does not get assigned evidential probability 1. There's still some plausibility in the claim that assignments of evidential probability can't be accuracy dominated; and that setting betting odds according to the evidence shouldn't lead to a sure loss. So while the material certainly needs to be examined carefully and reworked under the new interpretation, it still seems directly relevant.

By contrast, under an interpretation of probability as objective chance, the parameters of the discussion are changed more radically. First, one has to decide what the vehicles of chance are. Unlike belief, there's little motivation to think that necessarily equivalent propositions should be able to take different probabilities. That makes the coarse-grained formulation of probability, on which chances attach to sets of possible worlds, a natural one. This takes us back to the issues discussed earlier in this setting, whereby nonclassicality has to impact, if at all, in the algebraic structure of event space. On the other hand, we may still want to make sense of thought and talk about the objective chance of vaguely-specified events. In ordinary life, we might want to know about the chance of a ball in a bag being red. In special sciences, chances may attach to macro-events *e*, such that some physically possible course of micro-events will make it vague whether *e* obtains. The vagueness blocks a straightforward translation into the obvious coarse-grained setting. Likewise, theses that connect chance to subjective belief, such as the Principal Principle, are often stated in a way that presupposes a common domain of entities to which both degrees of belief and objective chances attach. I take the moral of this to be that *even if* the underlying chancy phenomena are best thought of as pertaining to coarse-grained events or propositions, we may owe an account of an induced chance-function across fine-grained propositions or sentences. At this point the main thread of discussion in this article is again relevant. The nonclassicist who rejects the claim that patchy is red or not red, certainly won't want to say that the objective chance of patchy being red or not red is one. So if this chance-attribution is well-formed at all, it seems to require a nonclassical objective chances. Of course, the earlier motivations for the particular formulations of nonclassical probabilities (such as the Dutch book and Accuracy arguments) presupposed the subjective interpretation. We might hope nevertheless to motivate nonclassical probabilities as formulations of objective chances by way of the Principal Principle, plus nonclassical probabilism about degrees of belief; the nonclassical analogue of the project carried out in (Lewis, 1980).

There are plenty of other reinterpretations of probability to consider. In each case, the nonclassical probabilities we've been looking at are an obvious resource, but as has been illustrated, the details matter.

9   Conclusions

To recap the main points of our discussion:

A. Nonclassical probabilities can be viewed as convex combinations of nonclassical truth values, and standard principles of probability can often be carried over to the nonclassical case if we substitute an appropriate nonclassical logic for the classical one. The appropriate nonclassical logical entailment can be generally characterized as one guaranteeing no drop in truth value. In section 6, we showed that truth values behave as 'experts' relative to these probability functions.

B. The truth values concerned should be thought of as the cognitive loads of the nonclassical truth status. But the general recipe may break down if (a) one does not endorse a semantically driven conception of logic (one will not have 'truth statuses' to play with); (b) one does not regard the statuses as cognitively loaded; or (c) the cognitively loads are not representable as real-valued degrees of belief. As discussed in section 7, this approach is also undermined if one insists on defining probabilities not over fine-grained truth-bearers (whether sentences, thoughts or structured propositions) but over coarse-grained, algebraically structured events or truth conditions.

C. The nonclassical probabilities so defined can be justified as constraints on rational belief via analogues of Dutch book and accuracy-domination arguments.

D. A notion of conditional probability can be characterized, that preserves important features of classical conditional probability. It satisfies a ratio formula, using a connective that is available in some but not all nonclassical settings.

E. An analogue of Jeffrey's recipe for calculating desirability of actions with respect to an arbitrary partition of states is available.

F. Though the discussion is conducted in the context of a subjective interpretation of probability, the theory of nonclassical probabilities that emerges is a resource for probability theory under other interpretations, though each application raises new philosophical issues.

This provides a rich field for further investigation:

A. Studying *axiomatizations* of nonclassical probabilities is an open-ended task. Can we extend the results of Paris, Mundici et al, and get a more general sense of what set of axioms are sufficient to characterize convex combinations of truth values? One key step was outlined earlier: to switch from 'immanent' axiomatizations (like Additivity) to purely logical ones (like Recarving).

B. We have focused on cases where 'truth values' (the cognitive loading of nonclassical truth statuses) take a particularly tractable form: represented by reals in $[0, 1]$. Can we get a notion of nonclassical probability in a more general setting, where the cognitive loads are not linear ordered, or where some truth statuses are missing such loading altogether? Perhaps the notion of expectations of non-real valued random variables may provide a lead here.

C. What is the relation between the cognitive loading of a nonclassical truth status (appealed to directly in the accuracy-domination argument) and the pragmatic loading (relevant to the generalized Dutch book argument). Must they coincide? If so, why?

D. How much of the theory of conditional probability transfers to the nonclassical setting? Can confirmation theory, in particular, be preserved in the nonclassical setting?

E. Many foundational and formal questions about nonclassical decision theory deserve exploration. Are analogues of classical representation theorems available, relative to a set of rational constraints on qualitative preference? What are the appropriate generalizations of the qualitative constraints of Jeffrey-Bolker theory? And can other forms of decision theory find expression in the nonclassical setting?

F. There are many possible interpretations of probability for which we could raise questions analogous to those discussed here. The chance and evidential interpretations have only briefly been discussed—and there's plenty more to say about these, and the relation between different forms of probability in the nonclassical setting. One particularly interesting interpretation to study is the *logical* intepretations of probability, on which the aim is to articulate the degree to which one propositions supports another, as a putative generalization of the total support captured by ordinary logical consequence. Surely *this* should interact directly with a nonclassical logic.

It's worth emphasizing a remark we made right at the start. It is not only convinced revisionists that need to be concerned about these issues. Anyone not dogmatically opposed to logical revision needs to take interest. For, prima facie, if one is open to the possibility that excluded middle fails in some cases, for example, then one shouldn't invest full confidence in each of its instances. And yet, it does not seem that one is irrational in harbouring such doubts, as the interpretation of classical probabilities as constraints on rational belief would suggest.Interpreting specific nonclassical probabilities as constraints on rational belief is likely to be problematic for analogous reasons.

Now, perhaps one could argue that in the end, such doubts manifest a lack of perfect, ideal rationality—so at least, the dogmatist must argue. I find this somewhat implausible. To begin with, it may be (as Putnam argued long ago) that the issue of which logic is correct is a broadly empirical one. Whatever one thinks of quantum logic as an putative exemplar of this, I would be surprised if we were never faced with scientific theory choice between total theories embedding incompatible logico-semantic packages. There's certainly a proliferation of such packages available in metaphysics; and many of us are naturalists enough to think that empirical evidence just as much as apriori reflection is holistically relevant to theory choice in metaphysics. It would seem to me a strike against a theory of rational belief if it can't represent rational uncertainty between such physical or metaphysical systems, and the gradual accumulation of evidence for one or the other. But set such considerations aside. The dogmatists need to argue, not just that there is an *absence* of possible evidence in favour of unfavoured logics, but that in some ideal limit there is *positive* evidence for their favoured system, sufficiently strong to require rational certainty. Why think that our total evidence, and superhuman processing power, would convey *total* conviction in the correctness of classical logic, or indeed any other logical system?

This uncertainty challenges all the forms of probabilism we have been discussing. Uncertainty over what the right logic is can lead to attitudes to individual sentences that are condemned by *all* the probability theories based on logics one is open to. Suppose that I divide my credence 50/50 over L-worlds and L^* worlds. And suppose S is true at all L worlds and false at all L^* worlds. Then I should be 50/50 in S. But this is condemned by the 'rational constraints'

associated by L and by L^*—L says I should have credence 1; and L^* says I should have credence 0.

So what can we offer an anti-dogmatist? One possibility is to drop the assumption that there is a space of truth value distributions (classical or otherwise) over which to define probabilities, independent of one's doxastic state. Perhaps the theory of subjective probability should be developed relative to a set of truth value distributions Z, *that the ideal agent regards as open possibilities*. The arguments above can be used to characterize convex combinations of truth value over the possibilites in Z, and a consequence relation defined via the no drop characterization used before. If the open possibilities include as many varieties of truth value distributions as has been suggested, then the Z-logic will be weak indeed, and the constraints on rational degrees of belief also weak. But it is a virtue of the framework we've been using that it applies even to this radically minimal setting, and many of the results carry over.[36]

Suppose we decided to theorize about ideally rational belief in this way. Even if the logic and constraints on rational belief are radically minimal, perhaps the majority of a sensible person's credence will be devoted to some $C \subset Z$ which contains—say—only classical truth value distributions. And if rational degrees of belief have to be convex combinations of truth value, then we do get the non-trivial result that the degrees of belief *conditional on C* have to meet the classical constraints. Mutatis mutandis for other interesting regions of the open possibilities Z—the Kleene possibilities K, say. So even though the stable constraints on ideally rational unconditional belief may be rather minimal, it implicitly inherits much richer constraints of rationality, in that it must be such that the updated probabilities $P_C(\cdot)$ be classical probabilities; $P_K(\cdot)$ be Kleene probabilities and so forth.

References

Choquet, G. 1953. 'Theory of capacities'. *Annales d'Institute Fourier*, **V**, 131–295.

De Finetti, Bruno. 1974. *Theory of probability: vol.1*. New York: Wiley.

Di Nola, A., Georgescu, G., & Lettieri, A. 1999. 'Conditional states in finite-valued logic'. *Pages 161–174 of:* Dubois, D, H., Prade, & Klement, E.P. (eds), *Fuzzy sets, logics, and reasoning about knowledge*. Kluwer.

Dummett, Michael A. E. 1959. 'Truth'. *Pages 1–24 of: Truth and other enigmas*. London: Duckworth. First published in *Proceedings of the Aristotelian Society* (59):141–62.

Dummett, Michael A. E. 1991. *The logical basis of metaphysics*. The William James lectures; 1976. Cambridge, Mass.: Harvard University Press.

Field, Hartry. 2009. 'What is the normative role of logic?'. *Proceedings of the Aristotelian society*, **83**(3), 251–268.

---

[36]Here is one important limitation (thanks to Mark Jago and Mike Caie for pressing me on this). I earlier mentioned and set aside nonclassical settings where truth-values are not linearly ordered, or not representable by real numbers. It's not so clear what a convex combination of truth-values is to be in that setting; nor whether our results generalize to it. But if these more radical nonclassical probabilities are included in Z, then these issues arise for what is rational to believe relative to Z itself.

Field, Hartry H. 2000. 'Indeterminacy, degree of belief, and excluded middle'. *Nous*, **34**, 1–30. Reprinted in Field, *Truth and the Absence of Fact* (Oxford: Oxford University Press, 2001) pp. 278-311.

Field, Hartry H. 2003a. 'No fact of the matter'. *Australasian journal of philosophy*, **81**, 457–480.

Field, Hartry H. 2003b. 'Semantic paradoxes and the paradoxes of vagueness'. *Pages 262–311 of:* Beall, J. C. (ed), *Liars and heaps*. Oxford: Oxford University Press.

Field, Hartry H. 2008. *Saving truth from paradox*. Oxford: Oxford University Press.

Fine, Kit. 1975. 'Vagueness, truth and logic'. *Synthese*, **30**, 265–300. Reprinted with corrections in Keefe and Smith (eds) *Vagueness: A reader* (MIT Press, Cambridge MA: 1997, pp.119-50).

Gerla, B. 2000. 'MV-algebras, multiple bets and subjective states'. *International journal of approximate reasoning*, **25**(1), 1–13.

Haack, Susan. 1978. *Philosophy of logics*. Cambridge: Cambridge University Press.

Hájek, Alan. 2008a. 'Arguments for—or against—probabilism?'. *British journal for the philosophy of science*, **59**.

Hájek, Alan. 2008b. 'Dutch book arguments'. *Pages 173–195 of:* Anand, Paul, Pattanaik, Prasanta, & Puppe, Clemens (eds), *The oxford handbook of rational and social choice*. Oxford University Press.

Hájek, Alan. Summer 2012. 'Interpretations of probability'. *In:* Zalta, Edward N. (ed), *The stanford encyclopedia of philosophy*. This article first written in winter 2002 and was last updated in Spring 2012 `http://plato.stanford.edu/archives/sum2012/entries/probability-interpret/`.

Hájek, Petr. 1998. *Metamathematics of fuzzy logic*. Springer.

Harman, Gilbert. 'Problems with probabilistic semantics.'. *Pages 243–237 of:* Orenstein, Alex, & Stern, Rafael (eds), *Developments in semantics*. New York: Haven.

Howson, C. 2008. De finetti, countable additivity, consistency and coherence. *The british journal for the philosophy of science*, **59**(1), 1–23.

Hughes, R.I.G. 1992. *The structure and interpretation of quantum mechanics*. Harvard University Press.

Jaffray, J-Y. 1989. 'Coherent bets under partially resolving uncertainty and belief functions'. *Theory and decision*, **26**, 90–105.

Jansana, Ramon. 2011. Propositional consequence relations and algebraic logic. *In:* (ed.), Edward N. Zalta (ed), *Stanford encyclopeadia of philosophy (spring 2011 edition)*.

Jeffrey, Richard C. 1965. *The logic of decision*. 2nd edn. Chicago and London: University of Chicago Press. Second edition published 1983.

Joyce, James M. 1998. 'A non-pragmatic vindication of probabilism'. *Philosophy of science*, **65**, 575–603.

Joyce, James M. 2005. 'How probabilities reflect evidence'. *Philosophical perspectives*, **19**.

Joyce, James M. 2009. 'Accuracy and coherence: prospects for an alethic epistemology of partial belief'. *Pages 263–297 of:* Huber, Franz, & Schmidt-Petri, Christoph (eds), *Degrees of belief*. Springer.

Keefe, Rosanna. 2000. *Theories of vagueness*. Cambridge: Cambridge University Press.

Kraft, C, Pratt, J, & Seidenberg, A. 1959. Intuitive probability on finite sets. *Annals of mathematical statistics*, **30**(408-419).

Lewis, David K. 1980. 'A subjectivist's guide to objective chance'. *Pages 263–93 of:* Jeffrey, Richard C. (ed), *Studies in inductive logic and probability*, vol. II. Berkeley, CA.: University of California Press. Reprinted with postscript in Lewis, *Philosophical Papers vol. II* (Oxford University Press, Oxford, 1986, pp. 83–113).

Lewis, David K. 1986. *On the plurality of worlds*. Oxford: Blackwell.

MacFarlane, John G. 2010. 'Fuzzy epistemicism'. *In:* Dietz, Richard, & Moruzzi, Sebastiano (eds), *Cuts and clouds*. Oxford: Oxford University Press.

Maudlin, Tim. 2004. *Truth and paradox: Solving the riddles*. Oxford: Oxford University Press.

Milne, Peter. 2007. 'Betting on fuzzy and many-valued propositions (long version)'. *In: ms.*

Milne, Peter. 2009. 'Betting on fuzzy and many-valued propositions'. *Pages 145–154 of:* Michal, Peliš (ed), *The logica yearbook 2008*. London: College Publications.

Mundici, D. 2006. 'Bookmaking over infinite-valued events'. *International journal of approximate reasoning*, **43**(3), 223–240.

Paris, J.B. 2001. 'A note on the Dutch Book method'. *Pages 301–306 of: Proceedings of the second international symposium on imprecise probabilities and their applications, isipta.* Ithaca, NY: Shaker.

Priest, G. 2006. *In contradiction: a study of the transconsistent*. USA:: Oxford University Press.

Priest, Graham. 2001. *An introduction to non-classical logics*. Cambridge: Cambridge University Press.

Roeper, P., & Leblanc, H. 1999. *Probability theory and probability logic*. University of Toronto Press.

Scott, Dana. 1964. 'Measurement structures and linear inequalities'. *Journal of mathematical psychology*, **1**, 233–47.

Shafer, G. 1976. *A mathematical theory of evidence*. Princeton: Princeton University Press.

Smith, Nicholas J. J. 2008. *Vagueness and degrees of truth*. Oxford: Oxford University Press.

Smith, Nicholas J. J. 2010. 'Degree of belief is expected truth value'. *In:* Dietz, Richard, & Moruzzi, Sebastiano (eds), *Cuts and clouds*. Oxford: Oxford University Press.

Soames, Scott. 1989. 'Semantics and semantic competence'. *Philosophical perspectives*, **3**, 575–596.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

van Fraassen, Bas. 1966. 'Singular terms, truth-value gaps, and free logic'. *The journal of philosophy*, **63**(17), 481–95.

van Fraassen, Bas. 1984. 'Belief and the will'. *The journal of philosophy*, **81**(5), 235–256.

Weatherson, Brian. 2003. 'From classical to constructive probability'. *Notre dame journal of formal logic*, **44**, 111–123.

Weatherson, Brian. 2005. 'True, truer, truest'. *Philosophical studies*, **123**, 47–70. Available at `http://brian.weatherson.net/ttt.pdf`.

Williams, J. Robert G. 2008. 'Supervaluations and logical revisionism'. *The journal of philosophy*, **CV**(4).

Williams, J. Robert G. 2012a. 'Generalized probabilism: dutch books and accuracy domination'. *Journal of philosophical logic*, **41**(5), 811–840.

Williams, J. Robert G. 2012b. 'Gradational accuracy and non-classical semantics'. *Review of symbolic logic*, **5**(4), 513–537.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Williamson, Timothy. 2000. *Knowledge and its limits*. Oxford: Oxford University Press.

Wright, Crispin. 2001. 'On being in a quandary'. *Mind*, **110**, 45–98.

Zadeh, L. A. 1968. Probability measures of fuzzy events. *Journal of mathematical analysis and applications.*, 421–427.