Nonclassical probability and convex hulls

Seamus Bradley

February 9, 2016

Abstract

It is well known that the convex hull of the classical truth value functions contains all and only the probability functions. Work by Paris and Williams has shown that this also holds for various kinds of nonclassical logics too. This note summarises the formal details of this topic and extends the results slightly.

1 An informal walk through the argument

J.R.G. Williams has recently published several interesting papers on the topic of "nonclassical probability" [16, 17, 18]. Building on work by J.B. Paris [12], Williams shows that standard arguments for probabilism[1] can be extended to cover agents who have credences over nonclassical logics. The key move is to notice that two standard arguments for probabilism – Dutch book arguments and Accuracy arguments – both work by showing that a rational agent ought to have her credences "in between" the possible truth values, and that this part of the argument makes no appeal to particular facts about how the truth values are structured [16]. In the case of classical logic, it is straightforward that functions in the convex hull of the truth value functions – the formal cashing out of this "in betweenness" – are all and only the probability functions. So the hard work is then to characterise what properties these functions in between the truth values have in general. Williams notes a partial characterisation given by Paris, and says that "beyond this, it is a matter of hard graft to see whether similar completeness results can be derived for [other logics]" [18]. This note is a contribution to that hard graft.

The remainder of this section informally describes the basics of the project we are engaged in.

The basic ingredients are a language, a set of sentences, that a rational agent has degrees of belief in. These sentences can be assigned various *truth statuses*. In the classical case, we have "True" and "False" as the statuses, but a three valued logic has "True", "False" and "Neither". One might be moved towards a three valued logic by considering some proposition involving a vague predicate:

[1] By "probabilism" I mean the view that a rational agents' credences – or degrees of belief – ought to satisfy the axioms of probability (Definition 5).

"McX is tall" might be neither definitely true nor definitely false if McX is somewhat tall but not very tall. One might think that in this case the sentence "McX is tall" gets truth status Neither.[2] Different ways the world could be are identified with different assignments of these truth statuses to sentences. The language has some compositional structure and the assignment of truth statuses tracks this structure in a certain way.

On top of this, there is the question of belief. Let's imagine that the "actual" truth status of a proposition is "True": what would an ideally rational, fully informed agent believe about that proposition? Presumably she would believe it to maximal degree 1? Choice of "1" as maximal is something of a convention, but a convention that makes the presentation rather easy. I am assuming throughout that credences are real-valued, though even this assumption can be significantly weakened, if one wants.

What about the ideally informed rational credence for a proposition assigned "Neither"? What would a perfectly rational, fully informed agent believe about this proposition? Perhaps she would believe it to degree one half? Or degree zero? Or degree one? There is, here, some flexibility in how one determines what the perfectly rational belief states are once we have some non-classical truth statuses. But let's say we fix some way of determining the perfectly rational beliefs to have in sentences with these alternative truth statuses. Codify this decision in a function from truth statuses to real numbers. This gives us a composite function from sentences to real numbers that assigns to each sentence the ideal belief state for that sentence in a particular world. Call this a cognitive evaluation.

In the terminology of Williams [18], the logic of the objects of belief is *semantically driven*: it is facts about ways the truth statuses could be distributed that determine the logic. The approach is also *cognitively loaded*: facts about truth status are intimately connected to the rational beliefs of ideally informed agents.

So the question now is: how can we constrain the rational beliefs of non-ideally informed agents? Arguments based on betting, or based on basic concerns about epistemic value secure the conclusion that a rational agent's belief state should be in the convex hull of the set of possible cognitive evaluations [16]. The remainder of this paper will explore the consequences of this fact.

The hope is that this note will be fairly elementary and self-contained, providing most of the definitions required to make sense of the material discussed.

2 Logics and lattices

Let's start with the objects of belief. We have some sort of logical language L. I want to assume as little as possible about L. The idea is that L is the collection of sentences we assume you have credal attitudes towards.

[2]There are other motivations for three valued logics, but I won't discuss them here. There are also other responses to the vagueness example: I will touch on two other logics for vagueness – degree theories and supervaluationism – later on.

Definition 1 A *logical language* is a set containing finitely many elementary letters that is closed under the binary operations $\vee$ and $\wedge$. The set contains two privileged elements $\top$ and $\bot$. $\square$

So if $\varphi, \psi \in L$ then $\varphi \vee \psi \in L$ and likewise for $\wedge$. $\top$ and $\bot$ will be privileged by serving as the tautology and contradiction. Note that we don't assume that L consists *only* of elementary letters and things built up from them using the two mentioned connectives: I want to leave open the possibility that L contains, for example, negation, or quantifiers, or modal operators... All we require is that it be closed under *or* and *and*.[3] The focus on these connectives makes sense since these are the connectives that appear in the definintion of probability.

Definition 2 A *cognitive evaluation* is a function $v\colon L \to [0,1]$ such that $v(\top) = 1$ and $v(\bot) = 0$. $\square$

I'm being very broad in my understanding of evaluations. We'll add some properties to evaluations in a moment. $\top$ and $\bot$ are privileged in having their cognitive loads picked out. Each v captures some way the world could be. And every way the world could be is captured by some v.

Definition 3 For a set of evaluations V, the *convex hull* of V, $co(V)$ is the set of functions, b, such that $b(\vartheta) = \sum_V w(v)v(\vartheta)$ for all ϑ, with $\sum_V w(v) = 1$ and $w(v) \geq 0$ for all $v \in V$. $\square$

The convex hull is important, because several important arguments about what it takes to have rational credences boil down to saying that you ought to have your credences be in the convex hull of the set of evaluations [12, 16, 17, 18]. In the classical case we know that the convex hull of the evaluations is the set of all and only the probability functions.

Definition 4 A *classical cognitive evaluation* is a cognititve evaluation $v\colon L \to \{0,1\}$ with:

T1 $v(\neg\vartheta) = 1$ iff $v(\vartheta) = 0$

T2 $v(\vartheta \wedge \varphi) = 1$ iff $v(\vartheta) = 1$ and $v(\varphi) = 1$

T3 $v(\vartheta \vee \varphi) = 0$ iff $v(\vartheta) = 0$ and $v(\varphi) = 0$ $\square$

Note that the truth value of any conjunction is determined by v's action on the conjuncts, likewise for disjunctions.

Definition 5 A *probability function* is a function $p\colon L \to [0,1]$ with:

P1 If $\vdash \vartheta$ then $p(\vartheta) = 1$ and if $\vartheta \vdash$ then $p(\vartheta) = 0$

P2 If $\vartheta \vdash \varphi$ then $p(\vartheta) \leq p(\varphi)$

[3] It may be that certain elements of the proof can survive the removal of this assumption by judicious insertion of phrases like "when $\vartheta \vee \varphi$ exists".

P3 $p(\vartheta \vee \varphi) + p(\vartheta \wedge \varphi) = p(\vartheta) + p(\varphi)$ □

Note that this definition of probability is relative to consequence relation $\vdash$. If $\vdash$ is classical consequence then this is equivalent to the standard definition of probability as a nonnegative, normalised, additive function on a Boolean algebra. This generalised definition of probability appears in Paris [12] and Williams [16, 17, 18]. When I say "probability" in what follows, what I will mean is generalised probability in the above sense. It will normally be clear from the context which $\vdash$ I have in mind when I talk about probability.

We are interested in theorems of the form "all $v \in V$ satistify certain properties if and only if all $b \in co(V)$ satisfy certain properties". For example, Paris shows that if the truth value functions satisfy certain properties, then the convex hull consists of all and only the probability functions.

We want some sort of consequence relation on L. The definition of probability makes reference to some relation among propositions with respect to which the probability function is monotonic, so we need to find such a relation.

Definition 6 A set of valuations V determines a relation on L, $\leq_V$ that satisfies the *No Drop* property:

$$\vartheta \leq_V \varphi \Leftrightarrow \forall v \in V,\ v(\vartheta) \leq v(\varphi)$$ □

I omit the subscript if it's clear which V is at stake. This relation is transitive and reflexive, so L is partially preordered by $\leq_V$. I use $\leq$ rather than something like $\vdash$, since there's no obvious analogue of $<$ – the irreflexive part of $\leq$ – for the turnstile, and I'll need this symbol later. Note that if V is the set of classical cognitive evaluations defined above, then $\leq_V$ is classical consequence.

What we're moving towards is having $L/{\equiv_V}$ being well behaved enough that we can do interesting things with it. We're going to impose structure on this object by requiring that V have certain nice properties. Let's now add some structure to our cognitive evaluations.

Definition 7 An evaluation v is *truth-functional for* $*$ for some connective $*$ iff for all $\varphi, \vartheta \in L$ we have $v(\varphi * \vartheta) = T_*(v(\varphi), v(\vartheta))$ for some real-valued function T_*. A set of valuations V is truth-functional for $*$ iff all $v \in V$ are truth-functional for $*$ with the same T_*. □

Definition 8 A relation $\sim$ is a *congruence relation for* $*$ iff

- $\sim$ is an equivalence relation (transitive, reflexive and symmetric)
- If $\vartheta \sim \vartheta'$ and $\varphi \sim \varphi'$ then $\vartheta * \varphi \sim \vartheta' * \varphi'$ □

Lemma 1 *If V is truth functional for $*$ then $\equiv_V$ is a congruence relation for $*$.* □

Given a relation $\leq$ on a language L, denote the symmetric part of $\leq$ by $\equiv$. Let $L/{\equiv}$ to be the set of equivalence classes of L.[4] What we're doing is turning a big

[4]For the classical case, this is the Lindenbaum–Tarski algebra.

space of sentences into a smaller space of equivalence classes of sentences. Since we're interested in functions that are monotonic with respect to the relation (P2 of Definition 5), these functions assign the same value to elements of the same equivalence class. So for our purposes, sentences that are grouped together can be treated as the same thing.

Definition 9 A truth function $T_*\colon [0,1]^2 \to [0,1]$ is a *t-norm* if it satisfies the following properties:

Associativity $T_*(x, T_*(y,z)) = T_*(T_*(x,y),z)$

Commutativity $T_*(x,y) = T_*(y,x)$

Non-decreasing If $x \leq y$ then $T_*(x,z) \leq T_*(y,z)$

Top-unit $T_*(x,\top) = x$

A truth function is a *t-conorm* if it satisfies the first three of these properties, but instead of Top-unit, satisfies:

Bottom-unit $T_*(x,\bot) = x$

Two more properties of interest:

Idempotent $T_*(x,x) = x$

Distributive Say $*$ distributes over $\circ$ when $T_*(x, T_\circ(y,z)) = T_\circ(T_*(x,y), T_*(x,z))$

□

We will say that V satisfies one of the above properties (with respect to a connective) when all v in V have the same truth function for the connective and that function satisfies the property.

Note that the classical truth function for conjunction is a (idempotent) t-norm, and for the disjunction is a (idempotent) t-conorm and further, conjunction and disjunction distribute over each other. For any n, an n-valued fuzzy logic satisfies these properties iff $T_\vee = \max$ and $T_\wedge = \min$. Other many valued logics do not satisfy idempotency, but satisfy the rest of the properties. The three cognitive loadings of the Kleene truth tables that Williams discusses [18] likewise satisfy the above properties.

Lemma 2 *Consider a set of valuations V and a relation defined by the No Drop condition $\leq_V$.*

- *If V satisfies Non-decreasing and Top-unit for $\wedge$ then $\varphi \equiv_V \varphi \wedge \top$*
- *If V satisfies Non-decreasing and Bottom-unit for $\vee$ then $\varphi \equiv_V \varphi \vee \bot$*
- *If V satisfies Idempotent for $\wedge$ then $\varphi \equiv_V \varphi \wedge \varphi$ and likewise for $\vee$*
- *If V satisfies Non-decreasing for $\wedge$ then, if $\varphi \leq_V \psi$ then $\varphi \wedge \vartheta \leq_V \psi \wedge \vartheta$ and likewise for $\vee$*

- *If V satisfies Associativity for $\wedge$ then $(\varphi \wedge \vartheta) \wedge \psi \equiv_V \varphi \wedge (\vartheta \wedge \psi)$*

- *If V satisfies Commutativity for $\wedge$ then $\varphi \wedge \psi \equiv_V \psi \wedge \varphi$ and likewise for $\vee$*

- *If $*$ distributes over $\circ$ then $\vartheta * (\varphi \circ \psi) \equiv_V (\vartheta \circ \varphi) * (\vartheta \circ \varphi)$* □

The proofs are all pretty easy so they have been ommitted. Note that associativity and commutativity mean that we can abbreviate long (finite) conjunctions or disjunctions by $\bigwedge_i \vartheta_i$ and $\bigvee_i \vartheta_i$ without ambiguity.

Like Williams, we are interested in *cognitively loaded logics*: logics where the structure comes from the set of cognitive evaluations V. We want V to be well-behaved enough that $\leq_V$ puts some interesting structure on L that we can exploit. In particular, we want the structure to be lattice-theoretic. Let's briefly summarise the definition of a lattice.[5]

Definition 10 A *poset* $\langle X, \preceq \rangle$ is a set X with a binary relation $\preceq$ that is reflexive, antisymmetric and transitive. Let $\prec$ be the irreflexive part of $\preceq$. $\preceq$ is a *partial order* on X, and I will sometimes talk about "the lattice X" where it is clear what relation is at stake. A *lower bound* for $A \subset X$ is a $\vartheta \in X$ such that $\vartheta \preceq \varphi$ for all $\varphi \in A$. The *infimum* – greatest lower bound – of $A \subset X$, called $\inf A$, is a lower bound of A such that if ϑ is a lower bound of A, then $\vartheta \prec \inf A$ or $\vartheta = \inf A$. Note that infima needn't exist in general, but if X is finite and non-empty, then $\inf A \in X$. Define *upper bounds* and *greatest upper bounds* (suprema, or sup) in the obvious way. Infima and suprema, if they exist, are unique.

A *lattice* is a poset that contains infima and suprema for all pairs of elements (and hence for all finite subsets). A lattice is *bounded* if it has top and bottom elements $\top, \bot$[6] such that $\vartheta \preceq \top$ and $\bot \preceq \vartheta$ for all ϑ in X. A lattice is *finite* if X is. A lattice is *distributive* if $\inf\{\vartheta, \sup\{\varphi, \psi\}\} = \sup\{\inf\{\vartheta, \varphi\}, \inf\{\vartheta, \psi\}\}$. □

A couple of shorthands will be useful: let $\inf\{\vartheta, \varphi\} = \vartheta \curlywedge \varphi$, and $\sup\{\vartheta, \varphi\} = \vartheta \curlyvee \varphi$.

Definition 11 A *filter* in a lattice X is an up-set of X closed under infima, or a subset of $F \subseteq X$ such that:

- If $\vartheta \in F$ and $\vartheta \preceq \varphi$ then $\varphi \in F$ (up-set)

- If $\vartheta, \varphi \in F$ then $\inf\{\vartheta, \varphi\} \in F$ (closed under infima)

An *ideal* of X is a down-set closed under suprema. The *principal filter generated by ϑ* is the smallest filter containing ϑ, likewise for the principal ideal. All filters (and ideals) of finite lattices are principal filters (ideals) of some element. A filter (ideal) is called *proper* if it is not identical to X. Order the proper filters

[5] See [1, 2, 5].

[6] I'm reusing this notation here, because $\top$ and $\bot$ will be privileged precisely in being the bounds on our logical lattices.

by set inclusion. The maximal elements of this ordering are the *ultrafilters.* Alternatively, a filter F is an ultrafilter iff: if $\vartheta \curlyvee \varphi \in F$ then $\vartheta \in F$ or $\varphi \in F$.

A *cover* of an element ϑ is an element φ such that $\vartheta \prec \varphi$ and there does not exist a ψ such that $\vartheta \prec \psi \prec \varphi$. A cover of ϑ is an element "immediately above" ϑ. A cover of ϑ is a minimal element of the up-set of ϑ. A *join-irreducible* element is one that covers only one element. Or equivalently, j is join-irreducible if $j = \sup\{\vartheta, \varphi\}$ implies $j = \vartheta$ or $j = \varphi$. Meet-irreducible elements are defined dually. □

Definition 12 A *No Drop System* $\langle L, V\rangle$ is a logical language L and a set of evaluations on L, V, such that $\langle L/{\equiv_V}, \leq_V\rangle$ is a lattice where $\varphi \wedge \psi = \inf\{\varphi, \psi\}$ and $\varphi \vee \psi = \sup\{\varphi, \psi\}$ for all $\varphi, \psi \in L$. □

We're now going to describe a set properties for a set of valuations V such that $\langle L, V\rangle$ is a No Drop system. Note that these are jointly sufficient, but are unlikely to be necessary conditions.

Theorem 1 *If V is a truth-functional set of valuations with an idempotent t-norm for $\wedge$ and an idempotent t-conorm for $\vee$ then $\langle L/{\equiv_V}, \leq_V, V\rangle$ is a No Drop System.*

Proof It is immediate that $\leq_V$ and V satisfy No Drop. We omit the subscript V throughout. That $L/{\equiv_V}$ is partially ordered by $\leq_V$ follows from $\leq_V$ being a congruence relation for the connectives. $\psi \leq \top$, so since V satisfies Non-decreasing and Top-unit $\varphi \wedge \psi \leq \varphi \wedge \top = \varphi$. Likewise for φ and ψ swapped. So $\varphi \wedge \psi$ is *a* lower bound on $\{\varphi, \psi\}$. Therefore, we just need to show that it is the greatest lower bound. Assume that ϑ is a lower bound; that is $\vartheta \leq \varphi, \psi$. Since V satisfies Idempotent and Non-decreasing for $\wedge$ we have $\vartheta = \vartheta \wedge \vartheta \leq \varphi \wedge \psi$. The proof for $\vee$ is similar. Thus, infima and suprema for pairs of elements exist and so $\langle L/{\equiv_V}, \leq_V\rangle$ is a lattice. ■

Further, if V has $\wedge$ and $\vee$ distribute over each other, then the lattice is distributive. Without idempotency, it is still the case that for any lower bound ϑ on $\{\varphi, \psi\}$, we have $\vartheta \wedge \vartheta \leq \varphi \wedge \psi \leq \vartheta \leq \inf\{\varphi, \psi\}$ (if this latter exists). So without idempotency you might not have $\wedge$ and inf coinciding perfectly, but they are still required to be somewhat close. A similar fact holds for sup and $\vee$. I leave this as a somewhat inchoate suggestion for how to generalise the above result.

The above theorem demonstrates that the kinds of systems arrived at by Williams [17, 18] are of this sort (if one assumes some reasonable properties of V). The logic needs at least this structure for the definition of probability to make sense. That is, what we have done up to now is provide a rigorous discussion of what it takes to be a well behaved semantically driven logic. What we show in the next section is that this is sufficient for the convex hull to contain all and only the (nonclassical) probabilities.

Might it be that some logic of interest in the context of rational belief fails to satisfy the properties discussed above? One example of this will be discussed in the second half of the paper. Other examples might crop up, and seeing what can be said about them in this context is, echoing Williams, a matter of hard graft.

3 Convex hull of evaluations

This section summarises the relevant results of Gustave Choquet's *Theory of Capacities* [4]. In particular, the following is a summary of some material from sections 39–42. We extract from Choquet an argument that the convex hull of the cognitive evaluations is precisely the set of probabilities (relative to the lattice defined by $\leq_V$). The first step is to show a more general result for all additive monotonic functions on a lattice.

Let L be a distributive lattice. Consider the set of real valued functions on this lattice F. Consider $f + g$ defined pointwise, and λf likewise for $\lambda \in \mathbb{R}$. This makes this set of functions a vector space.

Definition 13 Call an element e of a convex set[7] of functions an *extremal element* iff $e = e_1 + e_2 \Rightarrow e_1 = \lambda_1 e$ and $e_2 = \lambda_2 e$ where $\lambda_i > 0$. □

Consider the set of positive real valued functions on the lattice that are monotonic and satisfy: $f(\varphi \curlywedge \psi) + f(\varphi \curlyvee \psi) = f(\varphi) + f(\psi)$. These form a positive cone, call it A.

Theorem 2 *A function f is extremal in A iff $f = f_{P,\lambda}$ where this is defined:*

$$f_{P,\lambda}(x) = \begin{cases} 0 & \text{if } x \in P \\ \lambda & \text{if } x \notin P \end{cases}$$

P is a downset and λ a real number.

Proof We first show that $f_{P,\lambda}$ is extremal. Suppose that $f_{P,\lambda} = f_1 + f_2$. For $x \in P$ we have $f_1(x) = f_2(x) = 0$. Consider $u, v \notin P$, and take some $w \geq u$. Now,

$$f_1(u) + f_2(u) = \lambda = f_1(w) + f_2(w)$$

therefore

$$[f_1(w) - f_1(u)] + [f_w(w) - f_2(u)] = 0$$

However, we know from the fact that f_1 and f_2 are positive and monotonic that $f_1(w) \geq f_1(u) \geq 0$ and thus that $f_1(w) = f_1(u)$, and likewise for f_2. There is a w such that $w \geq u, v$. Thus $f_1(u) = f_1(v)$. Hence, f_1 and f_2 take constant values on the complement of P. Let these values be α_1 and α_2. We know that $\alpha_1 + \alpha_2 = \lambda$. Thus, $f_1 = \frac{\alpha_1}{\alpha} f$ and $f_2 = \frac{\alpha_2}{\alpha} f$. Hence, f is extremal.

Now assume f is extremal. If f only takes one non-zero value then it is clear that f is extremal, so assume without loss of generality that we have $0 < f(a) = \nu < \mu = f(b)$ for some a, b in L and ν, μ in the reals. Let $f_1(x) =$

[7] A set is convex iff it is identical to its convex hull

$f(x \curlyvee a) - f(a)$ and $f_2(x) = f(x \curlywedge a)$. f_1 and f_2 are in F (because L is distributive) and $f = f_1 + f_2$. $f_1(a) = 0$ but $f(b) \neq 0$, and $f(a) \neq 0$ so there is no $\lambda > 0$ such that $f_1 = \lambda f$. ∎

This theorem is basically a combination of Choquet's theorems 40.1 and 41.1.

Consider M, the set of positive additive monotonic functions on L with $m(\bot) = 0$. $M \subset F$. Conversely, every $f \in F$ is of the form $f(\bot) + m$ for some $m \in M$. Consider M_1, the subset of M such that $m(\top) = 1$. What are the extremal elements of M_1?

Theorem 3 *The extremal elements of M_1 are indicator functions of ultrafilters.*

Proof Consider an extremal element of M_1 called m. Let $B(m)$ be the set of elements of E such that $m(x) = 1$. $B(m)$ is an ultrafilter. It's clear that $B(f)$ for any $f \in M_1$ is a filter. To show that $B(m)$ is an ultrafilter, one assumes that $m(\vartheta \vee \varphi) = 1$ but $m(\vartheta), m(\varphi) < 1$ and then by the same strategy as in the proof above one shows that m can't be extremal.

Conversely consider an ultrafilter B and its indicator function f_B. $B(f_B) = B$ and f_B is an extremal element of M_1. Both claims are pretty obvious. ∎

So the set of probability functions (the functions in M_1) have, as their extremal elements the indicator functions of the ultrafilters. In the classical case, the indicator functions of ultrafilters are precisely the cognitive evaluations. In the nonclassical case, there can obviously be evaluations that aren't indicator functions of ultrafilters: for example, any evaluation that assigns any number in $(0, 1)$ to some sentence is not an indicator function of anything. But, by the same token, those evaluations will be non-extremal elements: they will be in the convex hull of the indicator functions of ultrafilters. So, as long as the indicator functions of ultrafilters are among the admissible evaluations, the convex hull of the evaluations will be equal to the convex hull of the indicator functions of ultrafilters. What's interesting to note is that as far as non-classical probability goes, it is only the lattice structure (encoded in the ultrafilters) that matters to what counts as probabilistically coherent.

So we have that the extremal elements of the set of probabilities are some of the cognitive evaluations. We now just need to show that the convex hull of this set of extremal elements is indeed the set of probabilities we started with. This follows from the Krein-Milman theorem [6], which says precisely that the convex hull of the extremal elements of a convex set is that set.[8] Note that we didn't need to assume that the lattice L was finite for any of the above.

There is an alternative approach to a related result that makes appeal to the structure of *MV algebras* [7, 10, 11]. This framework is more general in the sense that the logical algebraic structure may fail to be a lattice, but it is more restricted in the sense that an MV algebra requires a negation that behaves classically.[9] Exploring further relations between MV algebras and the current project would be an interesting avenue for further work.

[8] It is actually the closure of that set, but the set of probabilities is closed.

[9] In the sense that $\neg\neg\varphi = \varphi$.

This concludes the first half of the paper. The main conclusion so far is that semantically driven, cognitively loaded logics with truth functional cognitive evaluations for conjunction and disjunction that are an idempotent, distributive t-norm and t-conorm respectively have all and only the (generalised) probabilities as convex combinations of the evaluations.

4 Supervaluationism and filter evaluations

As Williams notes, there is an important and popular class of logics that are not truth-functional: supervaluations. Since truth-functionality was a vital part of getting the above results off the ground, it might seem that we cannot accommodate these kinds of logics. But it turns out that we can, as Paris commented.

From now on we will have to assume that the lattice L is finite. It does not suffice for this to require that the set of elementary letters is finite. What is actually required to ensure finiteness of the logical lattice is a tricky question: it is essentially a kind of "word problem". Let me point to a couple of cases where we do know that the lattice will be finite. A classical propositional logic with finitely many elementary letters will have a finite No Drop System (essentially, the Lindenbaum-Tarski algebra). This is so since the evaluation of a sentence is a function of the evaluations of the elementary letters, and there are only finitely many distributions of classical evaluations to the elementary letters. An n-valued Łukasiewicz logic with finitely many elementary letters has a finite No Drop System for the same reason. It follows from work in [8] that an S5 modal logic with finitely many elementary letters has a finite No Drop System. A logical language with a modal operator where there was some fixed maximum degree (in the sense of [3, p.74]) of sentences would also have a finite No Drop System.[10]

And let me point to a couple of cases where the lattice is not finite: an intuitionistic logic could have $\vartheta, \sim\vartheta, \sim\sim\vartheta, \sim\sim\sim\vartheta \dots$ be distinct propositions. So the Heyting algebra for one elementary letter is countably infinite. Likewise, for modal logics weaker than S5 $\vartheta, \Box\vartheta, \Box\Box\vartheta \dots$ could all be distinct.

The core idea of a supervaluational logic is to consider *sets* of evaluations of some underlying base logic (typically classical logic). The evaluations in the set typically represent various "sharpenings" or "precisifications" of some vague predicate. Something is "supertrue" if it is true on all sharpenings: true for all evaluations in the set. Consider some set of evaluations $X \subset V$ and define $v_X(\vartheta)$ iff $v(\vartheta) = 1$ for all $v \in X$ and 0 otherwise.

Lemma 3 *For every $X \subset V$, there exists a ϑ such that $v_X(\varphi) = 1$ iff $\vartheta \leq \varphi$.*

Proof The set of $\vartheta \in L$ that are such that $v(\vartheta) = 1$ form a filter. This follows from No Drop and that v satisfies Non-decreasing. So for each v there is a ϑ which is the principal element of the filter corresponding to v (since L is finite, every filter is principal). Since the intersection of filters is a filter and every

[10] I suspect that there is some relationship between the finite model property [3, p. 146 ff.] and having a finite No Drop System, but I'm not quite sure what.

filter is principal, the intersection filter has a principal element. In particular, consider the intersection of the filters corresponding to elements of X. The indicator function for this set is precisely v_X. Call the principal element of this filter ϑ. It is clear that $v_X(\varphi) = 1$ iff $\vartheta \leq \varphi$. ■

Definition 14 A *filter evaluation* is the indicator function of a filter. □

The above lemma and definition show that supervaluational truth value functions (supertruth evaluations) are precisely the filter evaluations. Since, in the finite case, there's a one-to-one correspondence between filters and their principal elements, we can talk about filters on $L/\equiv_V$ or elements of $L/\equiv_V$ interchangeably. This is not true for the infinite case.

Consider v_ϑ which is such that $v_\vartheta(\varphi) = 1$ iff $\vartheta \leq \varphi$ and 0 otherwise. Take the collection of these for all $\vartheta \in L$ and call it V^{SV}. If we take the base logic to be classical then every $v \in V$ is also in V^{SV} and it corresponds to an ultrafilter.

Jaffray [9] points out the following. Consider $b \in co(V^{SV})$.

$$\begin{aligned} b(\vartheta) &= \sum_{V^{SV}} w(v)v(\vartheta) \\ &= \sum_{L} w(v_\varphi)v_\varphi(\vartheta) \\ &= \sum_{\varphi \leq \vartheta} w(v_\varphi) \end{aligned}$$

It turns out that b defined as above are not probabilities, but they do satisfy the slightly weaker requirements for being *Dempster-Shafer belief functions.*

Definition 15 A *Dempster-Shafer belief function* satisfies:

DS1 If $\vdash \vartheta$ then $b(\vartheta) = 1$ and if $\vartheta \vdash$ then $b(\vartheta) = 0$

DS2 If $\vartheta \vdash \varphi$ then $b(\vartheta) \leq b(\varphi)$

DS3 $b(\bigvee_{i=1}^{m} \vartheta_i) \geq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|-1} b(\bigwedge_{i \in I} \vartheta_i)$ □

Let $m(\varphi) = w(v_\varphi)$ for all φ. Note that m so defined is a mass function:

Definition 16 A *mass function* satisfies:

M1 $m(\bot) = 0$

M2 $\sum_L m(\vartheta) = 1$

M3 $m(\vartheta) \geq 0$ for all ϑ □

Recall that the w in terms of which m is defined are the "weights" involved in something's being in the convex hull of the vs, and these are non-negative and sum to 1. Now all we need to do is show that b satisfies DS1–3 iff m satisfies M1–3. This turns out to be quite involved; the next two sections are devoted to laying out this result in detail.

5 Incidence algebras and the Möbius inversion

So now we need to go on a detour into some basics of incidence algebras on posets. This section summarises the relevant results of Rota [13]. First, let $[\vartheta, \varphi]$ be the set of ψ such that $\vartheta \leq \psi \leq \varphi$. Define $[\vartheta, \varphi)$ etc in the obvious way. Since L is finite, all such "intervals" will be finite. Consider the function $\zeta\colon L \times L \to \mathbb{R}$ defined by $\zeta(\vartheta, \varphi) = 1$ iff $\vartheta \leq \varphi$. Set ζ to 0 everywhere else. Consider also the function $\delta\colon L \times L \to \mathbb{R}$ defined by $\delta(\vartheta, \varphi) = 1$ iff $\vartheta = \varphi$ (and 0 everywhere else). Now, we want to find the μ such that:

$$\sum_{\psi \in [\vartheta, \varphi]} \mu(\vartheta, \psi)\zeta(\psi, \varphi) = \delta(\vartheta, \varphi)$$

That is, we want the μ which is the "inverse" of ζ where composition of functions is by "convolution" and the identity is the delta function (this is the incidence algebra of a poset).

Theorem 4 *The μ we need is:*

$$\mu(\vartheta, \varphi) = \begin{cases} 1 & \text{if } \vartheta = \varphi \\ -\sum_{\psi \in [\vartheta, \varphi)} \mu(\vartheta, \psi) & \text{if } \vartheta < \varphi \\ 0 & \text{if } \vartheta \not\leq \varphi \end{cases}$$

PROOF First consider the case where $\vartheta = \varphi$. Then we have $\mu(\vartheta, \vartheta)\zeta(\vartheta, \vartheta) = 1$. Therefore, $\mu(\vartheta, \vartheta) = \frac{1}{\zeta(\vartheta, \vartheta)} = 1$.

Now consider $\vartheta < \varphi$. Then:

$$\sum_{\psi \in [\vartheta, \varphi]} \mu(\vartheta, \psi)\zeta(\psi, \varphi) = \sum_{\psi \in [\vartheta, \varphi)} \mu(\vartheta, \psi)\zeta(\psi, \varphi) + \mu(\vartheta, \varphi)\zeta(\varphi, \varphi) = 0$$

Rearranging this, we see that:

$$\mu(\vartheta, \varphi) = -\frac{1}{\zeta(\varphi, \varphi)} \sum_{\psi \in [\vartheta, \varphi)} \mu(\vartheta, \psi)\zeta(\psi, \varphi)$$

So:

$$\mu(\vartheta, \varphi) = -\sum_{\psi \in [\vartheta, \varphi)} \mu(\vartheta, \psi)$$

Let $\mu(\vartheta, \varphi) = 0$ otherwise (i.e. if it is not the case that $\vartheta \leq \varphi$). ■

μ is also a right inverse for ζ, since the convolution operation is associative.

Theorem 5 *Let b be defined as follows:*

$$b(\vartheta) = \sum_{\varphi \leq \vartheta} m(\vartheta)$$

then

$$m(\vartheta) = \sum_{\varphi \leq \vartheta} b(\varphi)\mu(\varphi, \vartheta)$$

Proof

$$\begin{aligned}
\sum_{\varphi\le\vartheta} b(\varphi)\mu(\varphi,\vartheta) &= \sum_{\varphi\le\vartheta}\sum_{\psi\le\varphi} m(\psi)\mu(\varphi,\vartheta)\\
&= \sum_{\varphi\le\vartheta}\sum_{\psi} m(\psi)\zeta(\psi,\varphi)\mu(\varphi,\vartheta)\\
&= \sum_{\psi} m(\psi) \sum_{\varphi\in[\psi,\vartheta)} \zeta(\psi,\varphi)\mu(\varphi,\vartheta)\\
&= \sum_{\psi} m(\psi)\delta(\psi,\vartheta) = m(\vartheta)
\end{aligned}$$

■

So for any b defined in terms of m, we can use this Möbius inversion technique to get an expression for m in terms of b.

6 Mass functions and belief functions

We are now in a position to prove that b satisfies DS1–3 iff m satisfies M1–3.

Theorem 6 *Let b and m be related as in Theorem 5. If m satisfies M1–3 then b satisfies DS1–3*

Proof DS1 and DS2 are pretty straightforward. To show DS3, we need the following fact: if A is a nonempty finite set then

$$\sum_{B\subseteq A} (-1)^{|B|} = 0$$

This requires some combinatorics and the binomial theorem. Note also that if A is empty, the above quantity is 1, since $(-1)^0 = 1$.

Fix some collection of n sentences in L, ϑ_i. And let $I(\varphi) = \{i, \varphi \le \vartheta_i\}$.

$$\begin{aligned}
\sum_{\emptyset\ne I\subseteq\{1,\dots,n\}} (-1)^{|I|+1} b\left(\bigwedge_I \vartheta_i\right) &= \sum_I (-1)^{|I|+1} \sum_{\varphi\le\bigwedge\vartheta} m(\varphi)\\
&= \sum_{\substack{\varphi\in L\\ I(\varphi)\ne\emptyset}} m(\varphi) \sum_{\emptyset\ne I\subseteq I(\varphi)} (-1)^{|I|+1}\\
&= \sum m(\varphi)\left(1 - \sum_{I\subseteq I(\varphi)} (-1)^{|I|}\right)\\
&= \sum_{\substack{\varphi\in L\\ I(\varphi)\ne\emptyset}} m(\varphi)\\
&= \sum_{\varphi\le\vartheta_i \text{ for some } i} m(\varphi)
\end{aligned}$$

This is clearly less than $b(\bigvee \vartheta_i) = \sum_{\varphi \leq \bigvee \vartheta_i} m(\varphi)$. This concludes the proof of DS3. Note this last step requires that if $\varphi \vdash \vartheta_i$ for some i, then $\varphi \vdash \bigvee \vartheta_i$. This is obvious, since if $\varphi \leq \vartheta$ then $\varphi \leq \vartheta \vee \psi$ (this requires Non-decreasing and Bottom-unit). ∎

For the other direction of the proof, we'll need the following result.

Lemma 4 *The down-set of ϑ, $D(\vartheta)$ is a sublattice of L.*

Proof The only non-trivial part of showing that $D(\vartheta)$ is a lattice is showing that if $\varphi, \psi \in D(\vartheta)$ then $\sup\{\varphi, \psi\} \in D(\vartheta)$. But since $\varphi, \psi \leq \vartheta$ we have that $\varphi \vee \psi \leq \vartheta \vee \vartheta = \vartheta$. ∎

We'll also need the following well known fact:

Lemma 5 *An element of a distributive lattice can be written as the supremum of a set of meet-irreducible elements in a unique way.* □

This is basically lemma 3 of chapter 11 of [2], see also theorem 3.4.1 of [15] and the surrounding discussion. Let's now do the other direction.

Theorem 7 *Let b and m be related as in Theorem 5. If b satisfies DS1–3 then m satisfies M1–3*

Proof Note that $m(\bot)$ is actually undefined by the above expression, so let's set it to be 0. $\sum_L m(\vartheta) = \sum_{\vartheta \vdash \top} m(\vartheta) = b(\top) = 1$. That takes care of M1 and M2.

Consider again the down-set of ϑ, $D(\vartheta)$. Since $D(\vartheta)$ is a lattice, we can consider the meet-irreducible elements of $D(\vartheta)$. Call these the φ_i and assume there are n of them. Lemma 5 entails that every element of $D(\vartheta)$ is a supremum of these meet-irreducibles. In other words: if $\psi < \vartheta$ then $\psi = \bigwedge_I \varphi_i$ for some $I \subseteq \{1, \ldots, n\}$. Note that every element of the set $[\bigwedge_I \varphi_i, \vartheta]$ corresponds to a subset of I and therefore, by Proposition 3.1 of [19], we have:

$$\mu\left(\bigwedge_I \varphi_i, \vartheta\right) = (-1)^{|I|}$$

Thus:

$$\begin{aligned} m(\vartheta) = \sum_{\psi \leq \vartheta} \mu(\psi, \vartheta) b(\psi) &= b(\vartheta) + \sum_{I \subset \{1,\ldots,n\}} \mu\left(\bigwedge_I \varphi_i, \vartheta\right) b\left(\bigwedge_I \varphi_i\right) \\ &= b(\vartheta) + \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|} b\left(\bigwedge_I \varphi_i\right) \\ &= b(\vartheta) - \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|-1} b\left(\bigwedge_I \varphi_i\right) \end{aligned}$$

Now, $\vartheta \equiv \bigvee \varphi_i$, so by DS3 we have that this last expression is greater than 0. ∎

This is a modest improvement on the result in Paris [12] and Williams [17] in that we don't need to assume that the "base" logic is classical.[11]

7 Conclusion

I had two goals in writing this note. The first goal was to explicitly lay out, in detail, all the formal apparatus and results that lie behind the remarks Williams makes about the convex hull of cognitive evaluations in his discussion of nonclassical probability. To this end I have explictly set out the important results from Choquet, Rota, Jaffray and others that are necessary ingredients in Williams' dialectic. The second goal was to attempt to get clearer on the scope of those remarks. To this end I have carefully defined the concept of a No Drop System and shown under what conditions we can define probabilities over this system.

References

[1] J.L. Bell and A.B. Slomson. *Models and Ultraproducts: An Introduction.* North-Holland, 1969.

[2] Garrett Birkhoff and Saunders Maclane. *A Survey of Modern Algebra.* 4th. Macmillan, 1977.

[3] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic.* Cambridge University Press, 2002.

[4] Gustave Choquet. "Theory of Capacities". In: *Annales d'Institute Fourier* V (1954), pp. 131–295.

[5] J. Michael Dunn and Gary M. Hardegree. *Algebraic Methods in Philosophical Logic.* Oxford University Press, 2001.

[6] Ky Fan. "On the Krein–Milman Theorem". In: *Convexity, Volume 7 of Proceedings of Symposia in Pure Mathematics.* Ed. by V. Klee. 1963, pp. 211–219.

[7] Brunella Gerla. "MV-algebras, multiple bets and subjective states". In: *International Journal of Approximate Reasoning* 25 (2000), pp. 1–13.

[8] Alfred Horn. "Free S5 Algebras". In: *Notre Dame Journal of Formal Logic* XIX (1978), pp. 189–191.

[9] Jean-Yves Jaffray. "Coherent bets under partially resolving uncertainty and belief functions". In: *Theory and Decision* 26 (1989), pp. 90–105.

[10] Daniele Mundici. "Averaging the Truth-Value in Łukasiewicz Logic". In: *Studia Logica* 55 (1995), pp. 113–127.

[11] Paris appeals to a result of Shafer's [14] which is somewhat classical in nature, and Williams assumes that the evaluation of a conjunction is the product of the evaluation of the conjuncts.

[11] Daniele Mundici. *MV Algebras: A short tutorial.* http://www.matematica.uns.edu.ar/IXCongresoMonteiro/Comunicaciones/Mundici_tutorial.pdf. 2007.

[12] J.B. Paris. “A note on the Dutch book method”. In: *Proceedings of the Second International Symposium on Imprecise Probabilities and their Applications*. 2005 [2001], pp. 301–306.

[13] Gian-Carlo Rota. “On the foundations of combinatorial theory I theory of Möbius functions”. In: *Zeitschrift für Wahrscheinlichkeitstheorie un Verwandte Gebiete* 2 (1064), pp. 340–368.

[14] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[15] Richard P. Stanley. *Enumerative Combinatorics*. 2nd. Vol. 1. Cambridge University Press, 2012.

[16] J. R. G. Williams. “Generalised Probabilism: Dutch Books and Accuracy Domination”. In: *Journal of Philosophical Logic* (2012).

[17] J. R. G. Williams. “Gradational accuracy and non-classical semantics”. In: *Review of Symbolic Logic* (2012).

[18] J. R. G. Williams. “Non-classical logic and probability”. In: *Oxford Companion to Philosophy of Probability*. Ed. by Alan Hájek and Christopher Hitchcock. Oxford University Press, forthcoming.

[19] Chunlai Zhou. “Belief functions on distributive lattices”. In: *Artificial Intelligence* 201 (2013), pp. 1–31.