A Defense of Preference-Based Probabilism

by

Robert Carrington

Class of 2011

A thesis submitted to the
faculty of Wesleyan University
in partial fulfillment of the requirements for the
Degree of Bachelor of Arts
with Departmental Honors in Philosophy

Middletown, Connecticut April 12, 2011

Acknowledgements

I would like to thank Professor Sanford Shieh for advising this thesis, and Professors Joseph Rouse and Steven Horst for helpful conversations. Most of all I would like to thank my parents, to whom I owe all the remarkable opportunities that I have had. To say I am grateful does not even scratch the surface.

CONTENTS

Abstract: In this paper I endeavor to make two points. First, I argue that the Dutch Book Argument does not provide a suitable foundation for probabilism. Second, I show that an interpretivist approach *can* support probabilism by way of a Representation Theorem. I end by considering a few major objections to this position, which I argue are not successful. The final result, I hope, will be a picture of probabilism which is supported by clear argumentation and not mere technical apparatus.

A Defense of Preference-Based Probabilism

Robert Carrington

Introduction

Going back to at least the early modern period, epistemology has characterized an agent's knowledge state in terms of *categorical beliefs*. Accounts about what these beliefs actually amounted to have varied – they could be internal representations, dispositions to behave, or something else entirely – but agents were consistently thought of as either *having* or *not having* any single belief. As Daniel Hunter describes the picture, "Having a categorical belief is very much like having a penny in one's pocket: the penny is either in your pocket or it isn't – there isn't any in-between."[1]

Several points motivate this binary picture of belief. For one, our ordinary talk about belief often reflects this structure. The proper way to answer a question like "Do you believe in God?" seems to be with a yes or no, and not a degree measure. To the extent that this is true in general, our usage would imply that the concept of belief is categorical. A more technical reason might be philosophy's traditional focus on deductive foundationalism. We cannot deduce certainties about the world given only partial beliefs as premises. Indeed, how can syllogisms or propositional logic even be

[1] See Hunter (1996).

applied to such things? Even if we are 70% sure of “A” and 70% sure of “A→B,” modus ponens does not justify us in being 70% sure of “B.”[2] A third motivation is that it is hard to characterize knowledge without a notion of categorical belief. If I know that the Earth is round, doesn’t this imply that I categorically *believe* that the Earth is round? As part of the recipe for knowledge, it seems that beliefs need to be categorical.

René Descartes is a classic example of this view of epistemology, with motivating points of exactly this sort. In his First Meditation, Descartes starts with a rationale for his foundationalism:

> Several years have now passed since I first realized how numerous were the false opinions that in my youth I had taken to be true… And thus I realized that once in my life I had to raze everything to the ground and begin again from the original foundations, if I wanted to establish anything firm and lasting in the sciences.

This methodological point leads him directly to his view of beliefs as being necessarily categorical. Anything less than full belief, he argues, would be useless in the pursuit of knowledge:

> [R]eason now persuades me that I should withhold my assent no less carefully from opinions that are not completely certain and indubitable than I would from those that are patently false. For this reason, it will suffice for the

[2] Fuzzy logic does aim to answer questions such as these, but this approach was not available to those in the philosophical tradition. Even today the usefulness of fuzzy logic is debatable.

> rejection of all of these opinions, if I find in each of them some reason for doubt.[3]

Descartes, in this passage, is rejecting any use of partial belief in the pursuit of truth. Largely because of his deductive foundationalism, Descartes insists that beliefs are either given "assent" or are as good as false.

Since categorical belief involves taking something to be *true,* the laws of deductive logic are usually thought to provide a constraint on rational belief. Deductive logic constrains your belief set just as it would a set of propositions. First it provides a static constraint on your belief set – a demand of *synchronic* coherence. Just as you may not simultaneously assert A and ~A, you may not believe both A and ~A. Deductive logic also provides a notion of *diachronic* coherence with its inference rules, for example modus ponens. If your belief set includes A and A→B, you should add B. These "update" constraints give us coherence across changes to the belief set.

This traditional picture of belief contributes to an overall understanding of *rational action*. Making sense of a person's choices inevitably involves facts about their beliefs. If John Wayne reaches for his revolver, a rational interpretation (on this theory) might involve his belief that he is facing off with an unsavory character. The reconstruction of his decision might look something like this:

> The man in front of me is a villain.
>
> > (P1) If X is a villain, then X will attempt to kill me.
>
> So the man in front of me will attempt to kill me.
>
> > (P2) If X is going to attempt to kill me, I should stop the threat from X.

[3] See Ariew, Roger and Watkins, Eric, eds (2009), p 41.

So I should stop the threat from the man in front of me.

(P3) If I shoot X, I will stop the threat from X.

I should shoot the man in front of me.

On this picture, Wayne's categorical beliefs play an important role in explaining action because they are part of a line of reasoning which we can ascribe to him.

In the last century, Bayesian epistemology has come to offer a rival characterization of belief and rationality. On this picture, *degrees of belief* capture how the agent takes the world to be. In contrast to traditional epistemology, an agent might, for example, have 42% confidence in a proposition and legitimately make a decision on that basis. To be clear, the theory does not claim that probabilities are somehow literally inside our heads, "written in neurons." While accounts differ on how to cash out this notion, most probabilists take degree of belief to be a disposition, a functional role, or some other notion which can be realized in multiple ways. What is essential is that what we take to be the case can be expressed in partial beliefs rather than beliefs full-stop.

There are intuitive motivations for this view. For one, statements of the form "I think he's a Mormon, but I'm not sure" or "I am fairly certain the Steelers will win the Superbowl" are common enough in ordinary language to suggest that partial belief is a concept we use naturally. But more importantly, it seems to be more accurate in describing the rationality behind our decisions. When you decide to run to catch the train in the morning, this action does not seem to be the result of a categorical belief like "The train will come at 8:00am." After all, if someone stopped you midstride and asked if you were certain about said bus arriving exactly at that time, you would probably admit that the bus is often late. A better description might be that you are *wagering* that the bus will come on time. As probabilist pioneer Frank Ramsey phrased it, "whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home."[4] In short, our decisions are often the result of *how likely* we take something to be, not just the propositions we assume to be true.

The explanation of rational action, on this view, runs quite differently than on the traditional epistemology account. In this version, John Wayne takes the proposition that he's facing off with a nasty character as *very probable,* and this plays a role in explaining why he fires his pistol. The reconstruction of the decision might go like this:

> If I shoot the man in front of me,
>
> (i). I avoid being shot (0.90 utility) by a villain (0.90 confidence).
>
> (ii). I risk the unjust killing (-0.50 utility) of a non-villain (0.10 confidence).
>
> Overall expected value = [(0.90)(0.90) + (-0.50)(0.10)] = 0.76
>
> If I *do not* shoot the man in front of me,
>
> (i). I risk being shot myself (-0.90 utility) by a villain (0.90 confidence).
>
> (ii). I avoid the unjust killing (0.50 utility) of a non-villain (0.10 confidence).
>
> Overall expected value = [(-0.90)(0.90) + (0.50)(0.10)] = -0.76
>
> *So I should shoot the man in front of me.*

On this picture, it is not a deductive line of reasoning that makes Wayne rational, but a weighing of utility based on subjective probabilities. Note that on this

[4] See Ramsey (1990).

account, a more nuanced picture can emerge. It turns out that, although Wayne isn't certain he's shooting a bad guy, the undesirability of getting shot in the gut himself ultimately gives him reason to pull the trigger.

The most contentious part of the probabilist picture is that it claims to have an additional notion of coherence for degrees of belief, over and above that of deductive logic. Probabilists offer several arguments (which we will soon consider) for the view that *the laws of probability* are normative over our degrees of belief as well. As in the deductive case, the laws of probability offer a notion of *synchronic* coherence. A static set of beliefs should obey the laws of probability internally – it may not, for example, give 60% confidence to A and 60% confidence to ~A. The probability of a logically exhaustive set of cases must add up to exactly 100%. Some probabilists also offer arguments for *diachronic* coherence, arguing that the laws of probability give us rules for updating our degrees of belief. As we will see later, diachronic coherence is a significantly more ambitious thesis than the synchronic case.

A note about "calculation"

Prima facie it may seem like the probabilist picture is bound to fail because we can't, as a matter of neurological fact, make all our decisions based on probability calculations. We aren't probabilistic robots, and it would be foolish to even think that we should be. Imagine that you are at a travel agency and have to choose a vacation destination for your family. Does probabilism demand that you get out your scratch paper so you can estimate the likelihood of success and relative gain of each possible destination across the globe? Since a probabilist is committed to the view that rational decisions are based on the likelihood of the various outcomes, this might seem at first to be a counterexample to probabilism.

However this "counterexample" only refutes a caricature of the probabilist view. As a matter of fact, this situation fits quite well within the probabilist framework, and is even illuminated by it. The essential point to notice here is that there are actually *two* decisions being made in this scenario. While you are making the decision about where to go, you are simultaneously deciding how much time this decision is worth. Fully calculating out the possible benefits of the Bahamas or the Cayman Islands might result in the best choice between the two vacation options, but it certainly won't result in the best overall decision about how to use your time. The marginal gains from doing such an absurd calculation would be greatly outweighed by the loss of time and the frustration you cause your travel agent and family members. As Patrick Maher puts it, "doing a calculation to determine what act maximizes expected utility is itself an act; and this act need not maximize expected utility."[5] Probabilism does not require that you make optimal choices on every level – that just isn't possible. The probabilist advocates taking the best *overall* choice available to you.

[5] See Maher (1993), p.5.

Part 1: The Dutch Book and its Shortcomings

Although approaches involving probabilities are often called "Bayesian," one does not need to be committed to as much as the name suggests to support the picture involving degrees of belief. The more minimal view I advocate is often called subjective probabilism.[6] Subjective probabilism is committed to the following two theses:

1) We have *degrees of belief*, or *credences,* which reflect how we take the world to be.[7]

2) These degrees of belief should be coherent with each other at any given time – in other words, *synchronically* coherent in accordance with the laws of probability.

Any full-blooded Bayesian view would be committed to at least one additional thesis:

3) These degrees of belief should be coherent with each other across time – in other words, *diachronically* coherent in accordance with the laws of probability.

As we look at the arguments for the probabilist's picture of epistemology, I'll point out why I think we should restrict ourselves to the first two theses. We'll first examine the Dutch Book Argument, which is by far the most commonly used defense for these three points. I'll give an account of this argument, and then some reasons for rejecting it in favor of a different approach.

---

[6] "Subjective" because degrees of belief are taken as personal opinions, and "probabilism" because it takes the laws of probability to be a norm for these degrees of belief.

[7] Like most authors in the literature, I will use these terms interchangeably.

The Dutch Book Argument

The Dutch Book Argument was first given in “Truth and Probability,” a 1926 paper by the polymath Frank Ramsey. The main thesis of the argument are that our levels of confidence should be characterized as probabilities. That is to say, our degrees of belief should be constrained by the axioms of probability. The basic idea of the argument goes as follows:

Suppose I am watching a Yankees-Red Sox game and I feel that both teams have a very good chance of winning. Granting that belief has degrees, but could I legitimately be 70% confident in both teams? Or even just 51% confident in both teams? Let’s consider the consequences. If I believe each team has a better-than-half chance of winning, I should in each case be willing to bet *more than a dollar* (say, $1.02) against every dollar that you bet that the team will lose. For either team, I should expect to make money in the long run. But if I accept both of these bets simultaneously, which I should given my expectations, I am guaranteed to lose money. If the Yankees won, I would win a dollar but lose a larger sum, and the same would hold if the Red Sox won. Jumping into such a bet seems to be deeply irrational since we can determine *a priori* that we will lose money. Can an ideally rational person want to bet, but simultaneously know he will lose? Since a nasty set of bets like this can be constructed for anyone who does not obey the laws of probability, it seems like ideally rational agents must keep within these laws. In other words, the laws of probability are normative over our degrees of belief.

Now that we have a first gloss of the argument, we can examine things more closely. Clearly this argument hinges on two key points. First – that degrees of belief are linked to our betting dispositions. Second – that ending up in a Dutch Book is always irrational. These two key points correspond to the two central theses of probabilism. We need to have measurable degrees of belief, and these degrees must be constrained by the axioms of probability. To state the assumptions of the Dutch Book Argument more clearly, let us introduce a bit of technical notation. If $E \in X$, a bet on E (with betting quotient p and stake s) is a function f(x) which pays out (1 – p)s if $x \in E$, and –ps if $x \notin E$. Such functions are denoted bet(E,p,s). In addition, let Pr(A) to denote a person's degree of belief in A:

> (DB 1): A person has Pr(A) = p iff she would accept bet(A, p, s) for all stakes s.

> (DB 2): If a person accepts a bet b where b(x) < 0 for any state of affairs x, that person must have an irrational set of degrees of beliefs.

If these assumptions are granted, the Dutch Book Theorem proves as a matter of mathematical fact that anyone whose degrees of beliefs violate the axioms of probability can be Dutch Booked, and so is irrational. This theorem is also known as the Ramsey-De Finetti Theorem, since these authors are responsible for the proof. To present the theorem, a bit more formalism is necessary.

Let X be an algebra on the set of possible states of affairs – that is, each possible state of affairs and all the possible unions, intersections and other set-theoretic combinations. We can consider these elements as propositions. The element $A \in X$, for example, is the proposition that the true state of affairs is included in A.

The Dutch Book Theorem

Axiom 1 ($Pr(A) \geq 0$ for all A): Suppose $Pr(A) = r < 0$. Since you believe A to be less likely than a logical impossibility, you should be willing to risk your money for payouts which are less than zero. So then any bet on A with stakes s, where $s < 0$, will result in a loss for you. If A is true, then you receive $(1 - r)s$ dollars, which is a negative amount. If A is false, you must pay $-rs$ dollars, which is the loss of a positive quantity. So we've reached a Dutch Book. (Note that similarly, if $Pr(A)= r > 1$, choosing $s > 0$ for stakes on a bet inevitably leads to a loss, since you should be willing to pay more for a bet than even its highest possible payout.)

Axiom 2 ($Pr(A) = 1$ when A is a tautology): Where A is a tautology, suppose that $Pr(A) = r \neq 1$. By the proof of the previous axiom, $0 \leq r \leq 1$ or else we reach an immediate Dutch Book. Choosing $S > 0$ then leads to a loss in case A is true, which is the only possible case since it is a tautology. This is a guaranteed loss, so we have a Dutch Book.

Axiom 3 (Pr(A v B) = Pr(A) + (Pr(B) where A and B are mutually exclusive):

Suppose that you buy bets on two mutually exclusive propositions *a* and *b,* each paying one dollar for the prices *p* and *q* respectively. Then your net gain is as below:

a	*b*	Net gain
T	F	1 – p – q = 1 – (p + q)
F	T	–p + 1 – q = 1 – (p+q)
F	F	–p –q = – (p + q)

This is equivalent to:

a v *b*	Net gain
T	1 – (p + q)
F	– (p + q)

So your bets on *a* and *b* determine a bet on the disjunction *a* v *b* paying one dollar and with betting quotient *p+q.* If you were to, in addition, bet *against* this disjunction with a betting-quotient *r* not equal to *p+q*, stakes for this bet can be chosen such that you would consistently lose. If r < (p+q), then set the stakes for the new bet to also be a dollar. Then the net result is, in any scenario, r – (p +q) and this will be a negative quantity (since r is smaller than p+q). If r > (p+q), then set the stakes for the new bet to be less than (p+q)/r. Call this quantity X. The net results will inevitably be X – (p+q), which by our assumptions will be a negative quantity. So a Dutch Book results in any case where the betting quotient for the two bets are not equal.

For r < (p+q) :

a v *b*	Bet #1	Bet #2	Net gain
T	1 – (p + q)	– (1 – r)	r – (p + q)
F	– (p + q)	r	r – (p + q)

For r > (p+q):

a v *b*	Bet #1	Bet #2	Net gain
T	1 – (p + q)	– (1 – X)	X – (p + q)
F	– (p + q)	X	X– (p + q)

The Dutch Book Argument, however, only has force if these two initial assumptions are sound. Several authors have given reasons for thinking they are not, and I will present what seem to me the strongest of these reasons.

How the Dutch Book Fails for Thesis 1

The most straightforward version of the Dutch Book Argument simply *defines* degrees of belief in terms of your betting dispositions. Your degree of belief in *p* is equal to the value you place on a dollar bet on the truth of *p* – more precisely, the bet ($1 if *p* is true, $0 if *p* is false). If you take this bet to be as valuable as a gift of 43 cents, then you have 43% confidence in the truth of *p.* This definition does a good job of making sure that these degrees of belief will be normatively governed by probability axioms, but it does a terrible job of connecting up with our everyday notion of a credence. Finding counterexamples to this definition has been a great pastime for opponents of probabilism, probably in part because such counterexamples are so plentiful and easy to come by.

There are at least two simple formulas for concocting such a counterexample. The first way is by choosing someone with a strange attitude toward betting itself. Take a nun, for example. Surely her betting behavior does not reflect her degrees of belief, since she has none even hypothetically. Or how about a compulsive gambler? His disposition to take a bet doesn't reflect his belief state – it just shows how addicted he is. In both of these cases, it is simply implausible to call the agent's willingness to bet their "degree of belief." Or, if one insists that "degrees of belief" are just what we have defined them to be, then this new-fangled notion just isn't useful for what we wanted. The degrees of belief of the nun or the compulsive gambler will not actually tell us how these individuals take the world to be.

A second counterexample recipe is to separate monetary value and preference. If the betting amount is too small, then the gamble might not be worth your time. Rejecting such a bet would not be a result of your credences, but rather a result of not wanting to waste time betting on pennies. Or perhaps you are just filthy rich, and the possible gain of any bet is of little interest to you. As Hobart Brown said, "Money doesn't buy happiness – people with ten million dollars are no happier than people with nine million dollars." On the other hand, if the betting amount is very large in comparison to my bank account, then the gamble might be too risky for my taste. A million dollar bet on the Superbowl would put too much at stake to be of any value to me, despite my confidence in the Steelers and my opinion that the bet is a fair one. My rejection of this bet would be a result of my financial situation or my risk aversion rather than a result of my view of the world, so again my betting dispositions are not connected with "degrees of belief."

A less straightforward version of Dutch Book degrees of belief is given by Richard Jeffrey in *Subjective Probability – The Real Thing.* Instead of defining degrees of belief to be betting dispositions, Jeffrey asserts that betting is just one way of *measuring* the degrees which exist independently. Under the proper conditions, we can gauge someone's degrees of beliefs with wagers, but – no pun intended – all bets are off when the conditions fail to obtain. Jeffrey thinks we can interpret betting behavior into degrees of belief "the same way we interpret the relationship between the height of a column of mercury and temperature; the one is a reliable sign of the other within a certain range."[8] By restricting the connection between betting behavior and degrees of belief to proper contexts, Jeffrey avoids the problems involved with unusual quantities of money and people like the Puritan.

However, this approach is revealed as more problematic if we attempt to spell out what exactly is meant by "proper conditions." Certainly we must require that monetary values be moderate and that the person not have strong feelings about the practice of gambling, but clearly this is not sufficient. What if the person is superstitious about the number 13, and will avoid any bets involving that integer? Or what if the individual, seeking to be modest, wishes to win only small amounts of

---

[8] See Jeffrey (1977) p. 16-17.

money? We might stipulate that proper conditions exclude bets with numerologists or the overly considerate, but since any number of similar considerations can be generated the definition of ideal conditions would never be finished. What Jeffrey *really* wants, it seems, is conditions where our preferences regarding money are normal. That is, conditions where we value money linearly, and with a positive slope. But to talk about what we "prefer" or "value" we is to go beyond the apparatus of the Dutch Book Argument. If we cannot base degrees of belief exclusively on betting behavior, even under some stipulated conditions, then the Dutch Book machinery is unsupported. Since preferences seem to be an irreducible and important factor in determining degrees of belief, the Dutch Book faces a severe and fundamental problem. Later I will argue for a view that bases degrees of belief on preferences, which is in my opinion much more defensible.

How the Dutch Book Fails for Thesis 2

A second fundamental is that the Dutch Book Argument fails to establish the deep irrationality of violating the axioms of probability. While it has shown that disobeying the laws of probability can be very bad for your finances, this pragmatic point simply does not amount to a logical contradiction. An interesting point along these lines has been made in Hayek (2008), which points out that the Dutch Book Argument has a perfect mirror argument. The overall point of the Dutch Book Argument is that if your degrees of belief violate the laws of probability, then you are in a position to be exploited by a Dutch Bookie. Your degrees of belief are such that a series of bets, which you are disposed to accept, will lead you to certain loss. However, as Hayek points out, it is also true that if you are in a position that may lead to guaranteed loss you are also in a position that may lead to guaranteed gain. You degrees of belief would dispose you to accept a series of bets that lead inevitably to a guaranteed gain. Call such a guaranteed gain a Good Book. The Dutch Book Argument seems to be relying on an implicit Bad Neighborhood assumption, which makes the risk of Dutch Books more relevant than the opportunity for Good Books. But our epistemic justification does not seem like something that should depend on what kind of neighborhood we are in.

Another point, made by several authors, is that obeying the laws of probability does not seem to be the only way to avoid Dutch Books. Another way would be to not accept any bets at all, especially ones offered by bookies in wooden shoes. More modestly, we could even avoid Dutch Books if we simply knew that our degrees of belief were prone to getting us into financial trouble. The fact that our degrees of belief can cause problems for us does not imply that they are irrational – false beliefs can cause just as many problems. So it seems as though the gap between being Dutch Bookable and being irrational than defenders of this argument would like.

Depragmatized Dutch Books

Some authors have recently reformulated the Dutch Book Argument in response to concerns about its grounding in pragmatic considerations. These attempts seek to show that being Dutch Bookable reveals a flaw in your degrees of belief, separable entirely from whatever financial consequences that may result. The problem, they argue, is entirely epistemic and is merely dramatized by the talk of Dutch swindlers. I will present the three most prominent versions of this depragmatized Dutch Book Argument, and for each present the criticisms that I see as fairly decisive.

Howson and Urbach

The depragmatized Dutch Book Argument that Howson and Urbach offer is based on the notion of advantage, which they take to be pretheoretically intelligible. They argue that this concept is independent of our modern knowledge about probability theory, mainly citing a passage by Ian Hacking:

"We can actually 'see' the profits or losses of a persistent gamble. We naturally translate the total into average gain and thereby 'observe' the expectation even more readily than the probability… Certainly a gambler could notice that one strategy is in Galileo's words "more advantageous" than another."[9]

Howson and Urbach, of course, do not wish to use advantage in any frequentist sense (as this passage might suggest). They instead wish to take this notion as a subjective judgment, relative to the individual. We will denote the advantage for an individual r on a bet f by with the notation *adv(r,f)*. Out of advantage, Howson and Urbach construct the notion of a fair bet. A fair bet is defined

[9] Hacking (1975), p. 75.

as one that gives no advantage to either side, since "betting at your personal chance-based odds balances the risk (according to your own estimation) between the sides of the bet." More formally, this means:

> Definition: p is a fair betting quotient on A for me iff, for all possible stakes s,
>
> adv(r, bet(A,p,s)) = 0

Howson and Urbach argue that this notion of a fair betting quotient is important since, whenever we have a notion of how *likely* a proposition is, it must be "numerically identical" to the odds we give for that proposition. This does *not* imply that a person with a certain degree of belief would take any bet at all – simply that if she has that degree of belief she would *judge* such a bet as fair. The balance of advantage to each side of the bet would otherwise lopsided. Formally, this can be stated as:

> (U1): For all rational persons r and events A, there is a unique number p(A) which is a fair betting quotient on A for r.[10]

Howson and Urbach argue for an analogue of (DB 2), modified to use the notion of advantage. Instead of financial meltdown being the central point, here it is the judgment of fairness. A bet that is strictly negative or positive (for any way the world turns out) is simply not a fair bet. As the authors put it, "the assurance of a net gain

[10] Howson and Urbach later relax this idealized requirement to allow for a *range* of fair betting quotients. Since neither the positive argument here nor the rebuttal turn on this point, I will bypass the refinement in favor of simplicity.

or loss from finitely many simultaneous bets implies that they cannot all be fair." We can formally state this assumption as:

(U2): For all rational persons r and random variables f, if f > 0 or f < 0, then adv(r,f) ≠ 0.

The authors also assume, without argument, that "the sum of finitely (or even denumerably) many zeros is zero; hence the net advantage of a *set* of bets at fair odds is zero."[11] This premise can be stated as:

(U3): If each bet in a set of bets has adv(r,f) = 0, then the sum of the bets has adv(r,f) = 0.

Granting these premises, we can derive a similar depragmatized Dutch Book Theorem.

The Dutch Book Theorem for Howson and Urbach

Axiom 1: Take some p < 0. Then suppose you buy a bet on a proposition *a* paying out one dollar, for the price p. If *a* is true, you will gain 1 + |p| dollars. If *a* is false you will make |p|. So the bet will have an assured gain, and is a Dutch Book.

Axiom 2: Suppose that you buy a bet on a tautology *t* paying one dollar, at the price *p*. If *p* < 1, then you will make a certain gain of 1 – *p;* if *p* >1 then you will make a certain loss of *p* – 1. In either case, the bet will have an assured gain or loss and there will be a Dutch Book.

Axiom 3: Suppose that you buy bets on two mutually exclusive propositions *a* and *b,* each paying one dollar for the prices *p* and *q* respectively. Then your net gain is as below:

[11] 1993, p 79

a	*b*	Net gain
T	F	1 – p – q = 1 – (p + q)
F	T	–p + 1 – q = 1 – (p+q)
F	F	–p –q = – (p + q)

This is equivalent to:

a v *b*	Net gain
T	1 – (p + q)
F	– (p + q)

So your bets on *a* and *b* determine a bet on the disjunction *a* v *b* paying one dollar and with betting quotient *p+q*. If you were to, in addition, bet *against* this disjunction with a betting-quotient *r* not equal to *p+q*, where the stakes were also a dollar, then you would have a net gain of *r – (p+q)* whatever the truth values of *a* and *b* are:

a v *b*	Bet #1	Bet #2	Net gain
T	1 – (p + q)	– (1 – r)	r – (p + q)
F	– (p + q)	r	r – (p + q)

So, since r was assumed to be not equal to (p + q), the quantity r – (p + q) will be either strictly positive or negative, and thus will be a Dutch Book.

Problems with Howson and Urbach

The chief criticism of this depragmatized Dutch Book Argument is that our pretheoretical notion of "advantage" does not fulfill the role that Howson and Urbach expect it to. Specifically, it is dubious that "advantage" can be spelled out in terms of betting ratios. Consider the following two bets on H, the proposition that a coin toss will land heads-side up:

(i) bet(H, ½, 10) – a bet for $10 with 1:1 odds

(ii) bet(H, ½, 100000) – a bet for $100,000 with 1:1 odds

According to Howson and Urbach, both of these cases should have the same "advantage," since they have the same betting ratios. But does this accord with our intuitive notion of advantage? Most of us would judge the first bet as neither a favorable nor an unfavorable bet. We would ascribe to it zero advantage, since receiving this bet is neither a positive nor a negative result. But would we feel the same about the second bet? It seems that, in this case, receiving such a bet would be a huge liability. Since having such a bet is a burden, the advantage of this bet should be negative. This assessment could be rejected as not in accordance with the proper pretheoretical notion of advantage, but as Maher argues, "they are normal judgments of a kind that many people would make… The disadvantage of losing a large amount of money is greater than the advantage of gaining the same amount of money."[12]

---

[12] Maher (1993), p. 296

However now it appears that two bets, both with a 0.5 betting ratio for a single event, have different advantages. If this point is granted, Howson and Urbach's argument runs aground at several points:

First, betting advantage cannot be explicated as "expected value." If the expected value of the first bet is zero, then the second bet must also have expected value of zero. But despite this the two bets have different advantages.

Second, the assumption that every rational person has a single probability assignment for an event, defined as a fair betting quotient, is false. By definition, the first bet shows the betting quotient for our coin toss to be ½. But with the second, heftier bet the perceived advantage shifts downward and so the derivative notion of a "fair" betting quotient must shift as well. With so much at stake, better odds would have to be offered for us to take this as a neutral bet. Since our fair betting quotient in general depend on the stakes involved, there can be no single fair betting quotient for a bet on a coin toss.

Third, the package principle for fair bets is false. Take the $100,000 bet that we just considered. As we noted before, a neutral bet with these stakes would actually have to have better odds. Let's say that in my case having a betting quotient of 1/3 would render this bet fair. Since I think the coin equally likely to land heads or tails, a bet on tails would have the same advantage. Formally, this is represented as:

(1) adv(m, bet(H, 1/3, 100000) = 0

(2) adv(m, bet(~H,1/3, 100000) = 0

However if I were offered (1) and (2) together, my advantage for this bet would be very high. I would be assured of winning $333,333.33, no matter what happens. So combining bets with zero advantage need not result in a bet with zero advantage.

Of course, in each of these objections the central point is that our ordinary notion of "advantage" is not independent of preferences. For most of us, taking on a massive bet is a clear disadvantage – even when we are just as likely to win as to lose. What Howson and Urbach need is expected value, but since expected value is relative to a person's probability function this cannot be used without circularity. For, after all, your probability function is defined in terms of a bet with *zero advantage*.

The difficulty raised here could be avoided by using utility rather than money, but this added commitment to utility makes the Depragmatized Dutch Book redundant. To incorporate utilities into this argument, one needs to argue that rational people possess such utilities. This can be done with a representation theorem, but to use such a theorem would make Dutch Books superfluous. A representation theorem both establishes the existence of degrees of belief and the normative constraint of the probability axioms. Furthermore, Howson and Urbach intentionally distance themselves from utility approaches. A solid theory from the utility approach is "a task frought with difficultly, if not impossible" (77).

Since our pretheoretical notion of advantage cannot serve as the foundation of Howson and Urbach's depragmatized Dutch Book argument, their approach is unsound to begin with. Obvious fixes, such as swapping out “advantage” for “expected value” or incorporating utilities, have been shown to be not viable.

Hellman

Hellman endorses the Howson and Urbach account of the Dutch Book Argument, but seeks to generalize it. He takes judging a bet as fair to mean that “*anyone* – i.e. any *hypothetical* bettor – betting *on* or *against h* at those odds would confront zero expected advantage.” Hellman argues that this assessment has nothing to do with your personal preferences or your willingness to bet. In contrast to preference-based or behavioralist approaches, fulfilling the rational constraint of his depragmatized Dutch Book means that “a certain set of judgments –assessments of fairness of certain contractual arrangements –are consistent.” These general points aside, Hellman asserts that the Dutch Book Argument relies on two main premises:

> “What the Dutch Book arguments assume, in these terms, then is that (i) the net advantage of a finite (denumerable) set of fair bets is zero (countably infinite sets entering into the arguments for countable additivity); and (ii) that a set of bets, based on given odds, which is guaranteed to result in a net loss (or gain) in all possible circumstances cannot have zero net advantage, hence not all the odds can be fair.”

These premises can be formalized in exactly the same way as the respective premises from Howson and Urbach’s depragmatized Dutch Book Argument.

(U2): For all rational persons r and bets b, if $b > 0$ or $b < 0$, then $adv(r,b) \neq 0$.

(U3): If each bet b_i in a set of bets has adv(r,b_i) = 0, then the sum of the b_i's has adv(r,b) = 0.

Notably, Hellman does not assume anything like (U1), which defines a person's probability function in terms of her pretheoretical judgments of "advantage." He appears to take our degree of belief function as the starting point instead.

Hellman then makes a case for expanding the picture given by Howson and Urbach. The Dutch Book argument, he writes, is "just a *Gedanken-experiment* designed to help me pinpoint my degree of belief in the proposition in question," and that it should not be considered unique in this capacity. In fact he believes that there are countless scenarios that would make the same point, and which do not even involve bets. "What matters is that I estimate the 'expected flow' of some 'test quantity' either in the direction of truth or *h* or falsity of *h*," Hellman says. Thus bets can be entirely replaced by the more general "belief tests," whose definition does not involve money. "Literally these are just hypothetical set-ups in which a 'test-fluid' is said to display *positive, negative,* or *zero expected flow* according as the quantity xPr*(h)* is >, < or = yPr*(~h),"* Hellman says.

(H1): y/(y + x) is now called the *"fair* or *neutral belief-test quotient"* when the expected flow is zero, i.e. when

xPr(h) = yPr(~h)   (zero expected flow)

This generalized form of the “fair belief quotient” from Howson and Urbach, then, is the basis for our probabilities according to the Dutch Flow Argument. This notion of a fair belief test gives rise to two parallel assumptions to (U2) and (U3), which Hellman seems to assume.

(H2): For all rational persons r and *belief-tests b*, if $b > 0$ or $b < 0$ for all states of affairs, then $\mathrm{adv}(r,b) \neq 0$.

(H3): If each *belief test* b_i in a set of belief tests has $\mathrm{adv}(r,b_i) = 0$, then the sum of the b_i's has $\mathrm{adv}(r,f) = 0$.

Given these assumptions, the author is in a position to use a Dutch Book Theorem to prove that probability is a constraint on degrees of belief. However, since the Dutch Book Theorem for this view is not substantively different than that of Howson and Urbach, I’ll move directly to the criticism.

Problems for Hellman

Although Hellman’s version is meant to parallel the arguments from Howson and Urbach, it proceeds from different assumptions which make it unsuitable for defending probabilism. Howson and Urbach use their notion of “fair belief quotients” as a definition for degree of belief, and ultimately grounded this in the *pretheoretical* notion of advantage. Hellman, on the other hand, *defines advantage* in terms of *preexisting degrees of belief*, saying that “the *expected value* or (*expected* or *net*) *advantage* of the bet to the bettor is xPr(A) – y(Pr(~A), where Pr is the bettor's degree-of-belief function." Since his fair belief-test quotient notion is based on advantage, it cannot be used to define degrees of belief as Howson and Urbach did with their parallel notion.

Hellman could ground his belief-test quotient in terms of advantage, but he would face all of the difficulties that applied Howson and Urbach. Furthermore, his expanded Dutch Flow Argument would be much less plausible – it is difficult to see how our pretheoretical notion of advantage could apply to the general "test fluid" concept that Hellman advocates. Hellman explicitly states that "no value in the ordinary sense attached to the stakes at all," stating that sand or even manure could serve. If "advantage" applies to the flow of sand (in some test case), then it is certainly in a very contrived sense. Grounding degrees of belief in a pretheoretical notion of advantage will not work for Hellman's Dutch Flow Argument.

Hellman's notion of a fair belief-test quotient could be interpreted as merely providing a way of *discovering* your degrees of belief, rather than a way of *defining* them. However, in this case, Hellman's notion of a fair belief-test quotient involves an unsupported assumption. Given that a person judges the bet (*a* if *A*, -*b* if ~A) to have zero expected flow, Hellman concludes that their probability for A is b/(a + b). But this is to assume that Pr(~A) = 1 – Pr(A), which is part of what Dutch Book Arguments are meant to show. We cannot assume that our degrees of belief have properties of probabilities.

As it stands, Hellman's argument provides no support for probabilism. In addition, the most plausible fixes or re-interpretations do not seem to help the situation.

Christensen

In contrast to Howson and Urbach, Christensen takes "fair" to be his primitive notion. Setting aside a "behaviorist or functionalist account of partial belief, it is initially quite plausible that a degree of belief of, for example, 2/3 that of certainty *sanctions as fair* – in one relatively pretheoretic sense – a bet of 2:1 odds." Christensen acknowledges that several factors, such as risk aversion or the nonlinear utility of money, may override this judgment in the final decision on rational. However he still argues for a "normative ceteris paribus connection: other things being equal, an agent should evaluate such bets as fair." Christensen mentions situations involving unusual amounts of money as being such a cetaris paribus condition. This premise can be formalized as:

> (C1) For any event A and real number s, your probability function Pr sanctions judging as fair bet(A, Pr(A), s) unless cetaris paribus conditions obtain.

The remaining assumptions are standard Dutch Book machinery, comparable to the respective assumptions in the other authors' proposals. Based on plausibility, Christensen assumes "if a set of betting odds allows someone to devise a way of exploiting those odds to inflict a sure loss, then there is something amiss with those betting odds.” Formally, this is:

(C2) For all random variables f, if $f < 0$ then the judgment that f is fair is defective.

Christensen also assumes that misjudgments about fairness imply the defectiveness of the one’s belief set. As Christensen puts it, “if a single set of beliefs sanction as fair each of a set of betting odds, and that set of odds is defective, then there is something amiss with the beliefs themselves.” This is, formally:

(C3) If Pr sanctions judgments about fairness that are defective, then Pr is defective.

Although Christensen does not explicitly assume the additivity of fair bets, it is necessary for what he argues. This premise:

(C4) If Pr sanctions judging f fair and Pr sanctions judging g fair then Pr sanctions judging f+g fair.

Given these assumptions, the author is in a position to use a Dutch Book Theorem to prove that probability is a constraint on degrees of belief. However, since the Dutch Book Theorem for this view is not substantively different than that of Howson and Urbach, I’ll move directly to the criticism.

Problems for Christensen

The first major issue in Christensen's argument is the ceteris parabis clause from (C1). What are these conditions? The only example he mentions is the nonlinear utility of money. Unless the ceteris parabis clause is defined to include *anything* that would make one unjustified in a judgment of fairness, this is not a trivial premise. Nevertheless, it has not received any argument besides a pronouncement of plausibility. Indeed, considering its ambiguous cetaris paribus condition, it has not even been spelled out what *would* need to be argued for. So this premise is substantive and unsupported.

The second major issue with Christensen's account is (C3). A plausible assumption would be that if Pr *justifies* a bet as fair, then Pr is defective. But (C3) is quite different – it assumes that if Pr *sanctions* a bet as fair then Pr is defective. Since the notion of "sanction" contains the mysterious ceteris parabis clause, the typical set of degrees of belief might mistakenly sanction any number of things in non-ideal circumstances. Take a judgment of fairness based on the nonlinear utility of money – Christensen's example of a cetaris paribus condition. If the stakes involved in a gamble are too high, the ceteris parabis justification provided by my beliefs will fail. I may judge a bet as fair when it in fact it heavily favors me. But does this reveal anything inherently wrong with my degrees of belief? These same degrees of belief would be perfectly consistent if the stakes were in a moderate range. So (C3) seems implausible at best, and like (C1) it is only supported by an appeal to intuition.

Diachronic Dutch Books and Thesis 3

The Dutch Book Argument we have been considering has been *synchronic,* regarding a static set of beliefs. The argument aims to show that probability places certain internal constraints on sets like these. However, there is also a *diachronic* Dutch Book Argument, originally attributed to David Lewis. This argument instead aims to show how probability constrains the updating of a set of beliefs given new evidence.

The argument starts from the idea that, given degrees of belief for an algebra of propositions, we can deduce how much one believes a proposition *assuming another is true*. For instance if we know how likely both P and Q are, we can construct a likelihood for P *given* Q. This likelihood is denoted (P|Q) and is defined to be:

(Definition): Pr(P|Q) = Pr(P&Q)/Pr(Q) when Pr(Q)>0.

This definition is fairly intuitive. Let's say I believe there's a 0.2 chance of rain and lightning, but a 0.5 chance of rain generally. So *given* that there is rain, the possibility of rain *and lightning* is not as improbable. To be precise, the likelihood will be exactly 0.2/0.5 = 0.4. So your degree of belief in the possibility of lightning *given* rain is 0.4. The point of the diachronic Dutch Book is that these conditional beliefs determine how you should update your degrees of belief. The proposed requirement – called the Simple Principle of Conditionalization – is that your belief in P upon learning that Q should be equal to the value you *beforehand* assigned to

Pr(P|Q). The technical argument for the diachronic Dutch Book resembles the synchronic version, and I will not dedicate here the space required for a full explanation. The result is that unless you update your beliefs in accordance with the principle of conditionalization, you are subject to Dutch Books.

However, the diachronic Dutch Book Argument is open to numerous objections beyond those facing the synchronic version. The first is that an updating *policy* is perhaps not as legitimate an idea as its name suggests. The notion that we commit to a rule for updating (as Quine would put it) "come what may" is actually a rather hefty claim. Do we actually have such an enshrined update rule? As Mark Kaplan has pointed out, this would be to "treat all your conditional degree of confidence assignments as absolutely incorrigible."[13] One could not legitimately revise simply because of a moment of reflection, for example. Perhaps I used to think walking on water was good evidence for someone being a prophet. Couldn't I simply realize that it wasn't such great evidence after all? That perhaps it could be explained by foam sandals instead? Secondly, as Hajek has pointed out, there are added problems due to the fact that the bets are made on different occasions.[14] Why couldn't the bettor change her prices when she is betting at a different time with different information? There is a "lag time" between the two bets, as Hajek points out, and the new bets are made in a changed world. The diachronic Dutch Book seems to assume that I value money uniformly across time, and that the way that I gamble will not have changed. Considerations like these make Bayesian conditionalization a heavy

[13] See Kaplan (1996), p. 65.

[14] See Hajek (2009), p. 173-195.

thesis to advance. I do not wish to try and refute the thesis here – I only hope to show my reasons for remaining agnostic on this point. For the rest of the paper, it can be assumed that I am referring to synchronic constraints. We can now proceed to what I see as the strongest case for probabilism.

Part 2: The Representation Theorem and Preference-Based Probabilism

Interpretivism

In the arguments so far considered, degrees of belief have been defined in terms of *dispositions toward* or *judgments about* proposed bets. This way of fulfilling the first thesis of probabilism has been shown to face serious difficulties. I will follow (Maher 1993) in advocating the view that degrees of belief are instead attributed to an agent *as part of an interpretation* of a person's preferences. According to this view, an attribution of probabilities and utilities is correct just in case it is part of "an overall interpretation of the person's preferences that makes sufficiently good sense of them and better sense than any competing interpretation does." [15]

However, as we will see, when your preferences can be interpreted as a probability-utility function, they will in a way make *perfectly* good sense. You preferences will be fully rational because they will reflect what you expect to be the best option, the overall "the best bet."

Preference

[15] See Maher(1993), p.9.

Of course, this account of beliefs rests upon a notion of preference which has not yet been explained. For some authors, preference is given a behavioral definition. On such a view, you prefer *a* to *b* just in case you are disposed to choose *a* to *b.* However there are numerous objections to this definition. Savage (1954), for instance, argues that we may be indifferent between the choices, but simply be forced to make a choice. Alternatively, someone's pattern of choices could be intentionally random. When comes right out and states, "I wanted to choose as chaotically as possible," it is hard to argue that their choices reflect anything about how they actually value the items involved.

The view that I support is that preferences are instead attributed to a person as part of our interpreting that person rationally.[16] So although a disposition to choose is not sufficient for preference, it is a strong prima facie evidence for preference. In the opposite direction, while disposition to choose is not a sufficient condition for preference, it is a necessary one. That is, if you prefer *a* to *b,* you are disposed to choose *a* and *b.* This is a strict requirement for interpretation, meaning that attributing a preference of *a* over *b* to a person is never correct if that person is not disposed to choose *a* over *b*. If this is fulfilled, the best interpretation for preferences is the one which maximizes the following virtues:[17]

(i) When the person is disposed to choose *g* over *f,* the person strictly prefers *g* to *f.*
(ii) The person's preferences are normal ones for people to have in the circumstances that the person is in.

[16] See Maher (1993), p. 11.

[17] These criteria are taken from Maher (1993).

(iii) The person's preferences are what the person says they are, except where we have reason to think the person is mistaken or insincere.

(iv) The person's preferences are rational.

Now that we've established what degrees of belief and preferences *are,* we can move on to arguing for why the axioms of probability constrain rationality. Since there are a lot of details involved, it will be helpful to have an overall picture of how the pieces fit together. The full argument for probabilism can be summarized as follows:[18]

1. (Interpretivism)You have a particular probability function iff attributing them to you provides an interpretation that makes.

    (i) sufficiently good sense of your preferences and

    (ii) better sense than any competing interpretation

2. (Perfect interpretation) Any maximizing-expected-utility interpretation is a perfect interpretation (when it fits your preferences).

3. (Representation theorem) If you satisfy certain constraints on preferences (transitivity, connectedness, etc.) then you can be interpreted as maximizing expected utility.

4. The constraints on preferences assumed in the representation theorem of 3 are rationality constraints.

Conclusion: Rational persons have probability and utility functions.

[18] This way of formulating the interpretivist argument for probabilism is taken from Hayek "Arguments for – or against – probabilism?"

Ramsey's Representation Theorem

In mathematics, a "representation theorem" proves that every abstract structure with certain properties is isomorphic to a specific concrete structure. In other words, anything with the given properties can be viewed as *the same thing as* one archetypical case, for all intents and purposes. In the specific case of representation theorems for probabilism, the idea is to prove that when your preferences fit certain rationality constraints, they are isomorphic to a probability-utility function. Put another way, this means that we can view your preferences as the product of (1) a utility function which assigns values to the possible results and (2) a probability function which assigns likelihoods to the possible results. The result is that your preferences can be interpreted as a rationally based on expected utility. This isn't a purely mathematical undertaking – we must argue that the rationality constraints for your preferences *are* rational and that being interpretable in this way is a significant fact. We can break this into a three-part argument for probabilism:[19]

1. <u>Rationality</u>: Rational preferences must obey the axioms of expected utility theory.

2. <u>Representability</u>: If a person's preferences obey the axioms of expected utility theory, then she can be represented as having degrees of belief that obey the probability calculus. This component is entirely mathematical – the actual theorem within the "representation theorem" approach.

[19] This way of dividing things up is adapted from Zylda (1998).

3. Reality: If a person can be represented as having degrees of belief that obey the probability calculus, then that person *really does* have such degrees of belief.

In "Truth and Probability," Ramsey gave the original representation theorem which proved the second premise of the above argument for probabilism. His paper, however, did little to establish the other two premises. I'll present the full argument, discussing each premise in the order given above.[20][21]

Rationality

The following three axioms of the Representation Theorem are constraints on your preferences. The arguments that accompany them aim to show that, for each axiom, a violation of that axiom would imply inconsistent preferences. Here "inconsistent preferences" is understood as valuing one option in two different ways depending on how it is presented.

In what follows, notation such as "(A if P, B if ~P)" will denote a conditional prospect – namely the prospect of having A if P is true and having B if P is false. This plays an analogous role to bets or gambles in the Dutch Book Argument, but without the complications due to gambling being a social practice with associations and a

[20] Ramsey's proof was presented in much different form, with several more technical premises. Several of these were technical requirements rather than rational cones, simply included to guarantee a real-valued, continuous function as a result. Following Skyrms (1984), we will weaken the demands of our theorem to only demand that one's preferences be *embeddable* in a continuous set of preferences.

[21] The following axioms and proofs are adapted from Kaplan (1996).

history. The Puritan, for example, will not fail to have degrees of belief since even such an ascetic can have preferences about what sorts of uncertain situations she would rather be in. Use of conditional prospects was not available on the Dutch Book approach precisely because preferences were not incorporated into their argument. Degrees of belief had to be explained in terms of dispositions or judgments about *actual* gambles in the world.

<u>(1) Identity: If P is an ethically neutral proposition in which you have degree of belief ½ , and A is any situation, then you are indifferent between A and (A if P, A if ~P).</u>

First we must define what it means for a proposition to be "ethically neutral." Intuitively, an ethically neutral proposition is just one that you do not particularly care about. More formally, we can say that an ethically neutral proposition N is one such that, for any other proposition P, you are indifferent between (P & N) and (P & ~N).

With this in hand, the result falls out by definition. Given that the hypothesis is *actually* ethically neutral for you, you should have no reason to prefer a state of affairs differently whether it is the case that P or ~P. Since either bet will certainly deliver the same situation A, there is otherwise no difference between the two choices. To violate Identity, you would have to prefer one choice or the other simply based on how the choice was presented.

<u>(2) Dominance: If P is an ethically neutral proposition in which you have degree of belief ½, and A and B are set of conditional prospects on P, then:</u>

<u>i. if each conditional prospect in A can be uniquely paired with a less preferable conditional prospect from B, then you prefer A to B</u>

ii. if each conditional prospect in A can be uniquely paired with an equally preferable conditional prospect from B, then you are indifferent between A and B

This axiom states that the preferences you assign to individual conditional propects should constrain the preferences you have for the combinations. Consider the individual who does not obey this axiom. He might, for example, prefer X to Y and then prefer A to B, but when taken in conjunction prefer the package B&Y over the package A&X. However there is nothing substantively different between the sequence of the two choices and the single package choice – there is no logically possible world where they return different values. Again, to violate Dominance, you would have to prefer one choice or the other simply based on how the choice was presented.

(3) Ordering: If A, B, and C are all states of affairs that are not pairwise equally prefered, then:[22]

i. (reflexivity) you do not prefer A to A

ii. (transitivity) if you do not prefer A to B and you do not prefer B to C then you do not prefer A to C

This axiom prevents the possibility of preference loops. A preference loop is an indefinitely long sequence of trades that seems to rise ever higher in value, but which is actually just the recirculation of a finite number of items. For the simplest example, consider the agent who violates (i). Holding a conditional prospect A, that agent prefers to "trade up" and acquire the conditional prospect A. And when we

[22] The less natural "negated" form of this ordering axiom is necessary to allow for cases of equal preferences.

repeat this, the circulation of a single bet seems to create a series of ever-preferable items that the agent is acquiring. This point holds for loops that use any finite number of items and any agent whose preference form such a loop possess the same kind of irrationality. He or she has preferences that change based on how their options are presented

The next three axioms of the Representation Theorem are ontological assumptions. They are assumptions about you as an agent and also about the nature of degrees of belief. For each, I will try to argue why these are modest or at least reasonable premises.

<u>1) There is at least one proposition P that is ethically neutral for you.</u>

As we defined it earlier, an ethically neutral proposition P is one such that for any proposition A, (A & P) and (A & ~P) would be equally preferable. Traditional examples of such a proposition would be "The coin will land heads" or "The next card will be a jack of hearts." However, as Bradley (2004) has pointed out, this does not strictly fulfill the definition. After all, the proposition that the next card is a jack will affect the appeal of every other conditional prospect on that deck of cards. This is certainly true, but I will follow Bradley in arguing that since we are only making the assumption of ethical neutrality for purposes of *measuring* degrees of belief it does not matter. We only need a proposition that is ethically neutral *relative* to the other propositions we will deal with, and even then only neutral enough for our purposes.

So, in the end, our assumption is only that there is a proposition P such that for all relevant propositions Q, PQ ≈ Q≈ ~PQ.

<u>2) Where P is ethically neutral, your degree of belief in P is ½ iff for any A, B where A > B, you are indifferent between (A if P, B if ~P) and (B if P, A if ~P).</u>

This assumption bridges the gap between preferences and degrees of belief. If a conditional prospect on an ethically neutral proposition P is just as preferable as one on its negation, then you can know that P has degree of belief 0.5. Since your intrinsic interest in P is nonexistent, the only way in which P is affecting the conditional prospect is in its likelihood. And since the likelihood of P and ~P is equal, the likelihood must be 0.5.

<u>3) For any conditional prospect you can construct, there is a state of affairs such that you are indifferent between the conditional prospect and the state of affairs. Conversely, for any state of affairs there is a conditional prospect you can construct such that you are indifferent between the two.</u>

This axiom is necessary for technical reasons. The complete range of utility values is defined inductively through repeatedly halving the certainty of the existing conditional prospect with a new one. However, we assumed only one ethically neural proposition to start and such a "halved" conditional prospect cannot be constructed from the elements of the old one. The value of (A if P, (A if P, B if ~P) if ~P) is just the same as the value of (A if P, B if ~P).

Aside from the motivations for the axiom, it is a plausible premise to grant. Once we've established that a gradient of utility exists for some conditional prospect, it does not seem particularly important what that level of utility is about. We don't even need to make claims about the power of imagination in conceiving of possible situations – we know that the position upon the utility scale is available and logically possible to fill.

These three axioms give us a recipe for measuring the "intensity" of your preferences. What results is then called a utility function, but this name should not be misinterpreted. The function is only grounded in binary relations of preference. There is no quantity of utility that is being measured here. As should be apparent, the utility measure only captures how you prefer an item in relation to a continuum of such choices.

1. Where 'A' is the best case scenario, let A=1. And let 'Z' be the worst case scenario, with Z = 0. Then the situation (A if *h,* Z if ~*h*) should have a utility of ½, since it is more preferable than Z but not as preferable as A.

2. Take any situation that you find to be just as preferable as this last gamble, and call it X. Then the utility of (X if *h,* Z if ~*h*) should be ¼, again since this conditional prospect is not as preferable as X but still more preferable than Z. We can likewise consider the utility of (X if *h,* A if ~*h*). The utility measure of this conditional prospect should be ¾, since it is preferable to X but not preferable to A.

3. Continuing with this method in both directions, we can construct conditional prospects with values of 1/8, 3/8, 1/16, 7/16, or any other rational value between 0 and 1. Now, for any proposition P, you can find your degree of utility for P by comparing it with the values we've defined. If you are indifferent between P and a known value, then they must be equally preferred, and thus have the same utility.

Now that we have degrees of utility, we can argue for the two strongest premises. In these premises, *b*(x) will denote your degree of belief function and *u*(x) will denote your utility function.

<u>(General Degrees of Belief): Suppose that X ≠ Y and C = (X if P)(Y if ~P). Then: *b*(P) = *u*(C) – *u*(Y) / *u*(X) – *u*(Y).</u>

This generalizes the derivation of degrees of belief from utilities, allowing for the full range of credences. Since we have defined all our utility values, this works out fairly easily for ethically neutral propositions. The numerator captures the difference in value between the conditional prospect C and the proposition Y. This value depends both on the values of X, Y and the likelihood of P. But the denominator then scales this value by the difference in value of X and Y, leaving the likelihood of P as the only contributing factor.

However, this definition is also meant to generalize to *any* proposition, not just ethically neutral ones. How is this possible? It seems, after all, that the truth of P could affect the values of X and Y. The fact that it will rain will affect how much I value a picnic or a bike ride. Ramsey's answer to this problem is to say that, while this might be true, "there is a world with any assigned value in which P is true, and one in which it is false."[23] So, for the purposes of calculation we can substitute an equally valuable but logically independent proposition.

(Gamble Utility): For any set of mutually exclusive and jointly exhaustive hypotheses {P1,..,Pn} and any set of states of affairs $\{A_1,\ldots,A_n\}$, $u(A_1 \text{ if } P_1,\ldots, A_n \text{ if } P_n) = u(A_1 \text{ and } P_1)b(P1) + \ldots + u(A_n \text{ and } P_n)b(P_n)$.

Gamble utility is the generalization of our earlier premise relating degrees of belief and utilities. This premise assumes that the same relationship holds for sets of conditional prospects, and not just for ethically neutral propositions of degree 0.5. To see the plausibility of this premise, consider if it were false. An agent would have to have preferences based on mathematical expectation in regards to conditional prospects involving ethically neutral propositions, but a completely different way of relating utility and degrees of belief otherwise. To the extent that we consider degrees of belief to be a single epistemological concept, involving the same justification, we should be committed to this premise.

Representability

We can now show that probabilism follows from the rationality axioms we argued for earlier. Readers with less technical inclinations can skip this section if they

[23] See Ramsey (1990), p. 75.

are willing to grant that a violation of our rational constraints for preference is sufficient and necessary for a violation of the probability axioms by our degrees of belief.

Probabilism – for any hypotheses P and Q:
i. b(P) ≥ 0;
ii. if P is a tautology, then b(P) = 1; and
iii. if P and Q are mutually exclusive, then b(P v Q) = b(P) + b(Q).

(Proof of i) Let P be any hypothesis and let A1, A2, B1 and B2 be four states of affairs such that u(A1 and P) = 1, u(A2 and ~P) = 0 and u(B1 and P) = u(B2 and ~P) = b(P). By Gamble Utility, and the fact that u(A1 and P) = 1 and u(A2 and ~P) = 0,

b(P) = u(A1 if P, A2 if ~P).

By Identity, and the fact that b(P) = u(B1 and P) = u(B2 and ~P),

b(P) = u(B1 if P, B2 if ~P).

But if 0 > b(P), the foregoing is not possible. For, if 0 > b(P), then u(A1 and P) > u(B1 and P) and u(A2 and ~P) > u(B2 and ~P) and thus by Dominance(i),

u(A1 if P, A2 if ~P) > u(B1 if P, B2 if ~P),

and thus,

b(P) > b(P).

which violates Ordering. So b(P) ≥ 0.

(Proof of ii) Suppose that P is a tautology and that A and B are two states of affairs such that *u*(A and P) = 1 and *u*(B and ~P) = 0. By Gamble Utility, and the fact that *u*(A and P)=1 and *u*(B and ~P) =0,

b(P)=u(A if P, B if ~P).

By Identity and the fact that of the two hypotheses, P and ~P, only the first is logically possible,

u(A if P, B if ~P) = u(A and P).

So

b(P) = u(A and P).

That is (given that *u*(A and P) = 1),

b(P) = 1.

(Proof of iii)

Let P and Q be any mutually exclusive hypotheses and A, B, C, D and E five states of affairs such that *u*(A) and *u*(B and P) = *u*(C and Q) = *u*(D and ~P & ~Q)) = *u*(E and (P v Q)) = 1. By Identity, and these equalities:

u(B if P, C if Q, D if ~P &~Q) = u(A)

and

u(E if P v Q, D if ~P &~Q) = u(A).

So by Gamble Utility,

u(B and P)b(P) + u(C and Q)b(Q) + u(D and ~P&~Q)b(~P &~Q) = u(E and (P v Q) + u(D and ~P & ~Q)b(~P&~Q).

We divide both sides by *u*(A) and by our original stipulation for A:

b(P) + b(Q) + b(~P &~Q) = b(P v Q) + b(~P &~Q).

Subtracting b(~P&~Q) from both sides:

b(P) +b(Q) = b(P vQ).

Concluding remarks

This argument for probabilism has proceeded from several axioms that were independently argued for as requirements for rationality. Note that unlike the Dutch Book Argument, this theorem does not require that degrees of belief be directly cashed out in terms of dispositions to bet or judgments of fairness regarding bets.

Against the Zen Monk

A fundamental tenet of the Representation Theorem – and nearly any attempt to relate epistemology to decision theory – is that your credences are related to your preferences. This connection is not an obvious one, and some have challenged whether it exists at all. Lina Ericksson and Alan Hajek provide a clear statement of this doubt:

> "…every account of 'degrees of belief' that we will discuss seeks to connect it to *preferences,* in the service of an eventual argument for probabilism. The arguments supposedly prove that some preference-based thing obeys the probability calculus. The question remains, however: does that thing deserve the name 'degrees of belief'? In fact, every account fails to establish even a

> contingent connection between the two, still less a necessary connection, still less a sound argument for probabilism."[24]

To cast doubt on the existence of any link between preferences and credences, Ericksson and Hajek have conjured up a thought experiment involving a Zen Buddhist monk "who has credences but no preferences." The ideal monk they describe is sitting high atop a Himmalayan mountaintop, contemplating his place in the universe. He has many credences about the world – he assigns a high probability, for instance, to the notion that he is atop a Himmalayan mountain and that there is a fundamental Oneness to the universe – but being an ideal Zen monk he has no preferences about any of these things. Any event that might pass before his eyes would be greeted with the same total equanimity and acceptance.

The upshot of this thought experiment, according to Ericksson and Hajek, is that preferences and credences are conceptually independent and, in principle, entirely separable. We would still want to say that the monk in question has credences, but (granting the idealization) it doesn't seem that he has preferences. So, the argument goes, we should not assume that preferences are at all related to credences.

However, I will argue that this thought experiment is stranger than it initially appears, and that it is not really so relevant to probabilism. One way of getting at this strangeness is to consider how the monk would make a decision. If we take the thought experiment seriously, the monk would see any choice – say the choice of

[24] See Erickson, Lina and Hayek, Alan (2007), p 3.

rubbing his chin or tumbling down the mountain to his demise – as being a tossup between two results that are for all purposes indistinguishable. Note that it is not that he has *equal* preference for either option, and that after calculating he finds both choices to be suitable. Rather, he has no preferences at all, and calculation cannot even begin. This is an important difference since a monk with *equal* preference toward any state of affairs would not show that it is possible to separate preferences and credences, as Ericksson and Hajek intend. The preference-less monk wouldn't just have as much reason to rub his chin as fall to his death – he wouldn't have any reasons at all.

This is not to say that he would necessarily be just as likely to do one as to do the other, for it could plausibly be claimed that he has more of a disposition toward chin rubbing than toward suicide. But this – and any similar reason – would simply be a physical or neurological matter of fact, and we would not really be treating the outcome as a *decision* on the part of the monk. The eeriness of the scenario is that, if preferences really are lacking, then the monk has no *reasons* for his choices and, in the proper sense of the word, is no longer really making *decisions*.

Here is the point again, in different clothes. Imagine that a young philosopher risks her hide to climb up and seek the monk's counsel. When she poses a question to the monk – perhaps about the nature of wisdom or the stock market – why expect him to accurately report his credences? Given that he has no preferences, there would be just as much value in responding truthfully as there would be in starting to yodel or foam at the mouth. My point, perhaps overwrought, is to show that the familiar image of the monk is misleading in this scenario. The person we are presented with is not just an extreme case of being "laid back" but a non-agent who cannot properly be said to make decisions.

The underlying problem is that rationality is (at least largely) instrumental and furthermore if there is no *telos* for the instrument, it ceases to be one. It seems fair to grant that the monk has things which, *if only for a lack of preferences,* would be called degrees of belief. But this is an important difference. Eriksson and Hajek themselves, when giving a Ramsey sentence for degree of belief, in part describe it as that thing "that guides decision." But if we look at the monk, it seems as though the monk isn't really making decisions at all. Can we still consider those idle psychological devices to be "degrees of belief"? In the same way that a boat in the desert isn't really a boat, these things don't deserve the name "degrees of belief."

Against probabilities as content

For a conservatively-minded objector, it might seem that we could capture the distinctions between degrees of belief within the *contents* of categorical beliefs. On this view, having 70% confidence in the proposition that it will rain is *actually* to have a categorical belief in the proposition "There is a 70% chance that it will rain tomorrow." However, there is a major problem with this picture. If the notion of degree is considered as content, then we cannot attribute degrees of belief to anyone who does not have this concept. Children, for example, would have no degrees of belief, and neither would people living in the 14th century. But this is a wrong result, since it seems like these individuals are just as capable of being good probabilist agents as we are. So it seems as though placing probabilities in the *contents* of our beliefs is simply too implausible.

Against the charge of circularity

One objection, raised in Ericksson and Hajek (2007), is that interpretation seems to involve a notion of rationality. If probabilism culminates in a set of rational constraints on credences, how can rational interpretatation be part of how we attribute beliefs? As the authors put it, "the application of the principles of charity and rationalization presupposes that we already know what a charitable and rationalizing interpretation is. The interpretivist presupposes that the appropriate normative standard is probabilism."[25] There seems to be some circularity in proposing *rational constraints* on credences originally attributed to be *rational*. I argue, however, that this is a misunderstanding of the whole probabilist undertaking. The probabilist does not presuppose that probability constrains rationality – nowhere in the proof of the Representation Theorem was there an appeal to the probability axioms. Instead, we began with *pretheoretical* intuitions about what it meant to be rational, and then showed that obeying these constraints required one to be a probabilist. While it may be true that interpretivism assumes that we know what a charitable and rationalizing interpretation amounts to, these notions do not in any way rely on probabilism as a

[25] See Ericksson and Hajek (2007), p. 22.

premise. These are pretheoretical notions which are required to even be capable of interpreting others at all.

Furthermore, there is no contradiction in holding both our pretheoretical notion of rationality and the axioms of probability to both be normative over our degrees of belief. Why couldn't it turn out that being ideally rational, in our intuitive sense, turns out to be a form of probabilism? In fact, it seems like one *needs* a paradigm case for rationality to be able to explain irrationality in other cases. It is only when we have an ideal case that we can explain *why* some other choice is not rational.

Conclusion

I've argued that Dutch Book Arguments do not provide legitimate support for either of the two theses of probabilism. Defining degrees of belief in terms of *dispositions* or *judgments* toward bets is simply not realistic. Puritans, for instance, should have degrees of belief even though they accept no bets and judge any gamble as unfair (since it risks hellfire). Also, it is implausible that safety from Dutch Books is the reason that our degrees of belief are constrained by the probability axioms. Isn't losing money a separate issue from epistemology? Even the depragmatized versions of the Dutch Book Argument were shown to be unsuccessful in getting away entirely from this notion.

As an alternative, I've argued that we should instead think of degrees of belief as *interpretations of preference*. Once we've established a person's preferences, we take their degrees of belief to be whatever makes these preferences maximally rational. By a result from Frank Ramsey, we can then say that interpreting a person's preferences as the product of a probability and utility function will provide a perfectly rational interpretation of those preferences whenever such an interpretation is possible. The final result: If your degrees of belief violate the axioms of probability, there must be an inconsistency in your preferences.

Among the objections to this view are the Zen Monk and the charge of circularity, which I saw as particularly deep questions. The Zen Monk – who has credences but no preferences – seems to show that interpreting preferences out of belief is misguided. But, as I've argued, such a monk could not be considered as actually making decisions and so his "degrees of belief" are no longer playing a normative role which is constitutive for their being degrees of belief at all. Secondly, the interpretivist view has been charged with circularity. If our degrees of belief are attributed to be maximally rationality, isn't a notion of rationality being presupposed which isn't probabilism? I argue that there isn't any circularity here at all – *of course* we must have an intuitive notion of rationality before we begin. In fact, the constraints on rational preference which are assumed by the Representation Theorem are based on our pretheoretical intuitions about what is rational. However, who is to say that probabilism isn't the ideal case of this pretheoretical notion? In fact, it seems like one *needs* a paradigm case for rationality to even begin attributing irrationality to persons who make bizarre choices.

With all this in mind, I take preference-based probabilism to have significant support. While there remain reasoned objections, no in-principle argument against this view seems to be decisive or “knock-down.” Probabilism is a legitimate contender in epistemology, not just for its technical flourishes but also for its philosophical depth and substance. Since John Locke first declared that “assent ought to be regulated by the grounds of probability,” probabilism has flourished into a full-blown approach to explaining rationality and justification.[26] Looking forward, there is hope that probability will grow into being an even fuller guide to the examined life.

---

[26] Locke, John. An Essay Concerning Human Understanding, London: Baynes and Son, 1823, p. 601.

Works Cited

Ariew, Roger and Watkins, Eric, eds (2009). Modern Philosophy: An Anthology of Primary Sources, 2nd Ed. Hackett Publishing Co.

Bradley, Richard (2004). "Ramsey's Representation Theorem." *Dialectica* Vol. 58, No 4, p. 483-498.

Christensen, David (1996). "Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers." The Journal of Philosophy, Vol. 93, No. 9 (Sept.) p. 450-479.

Davidson, Donald (1990). "The Structure and Content of Truth." *The Journal of Philosophy,* Vol. 87, No. 6. (June). p.279-328.

Earman, John (1992). Bayes or Bust? Cambridge, MA: MIT Press.

Ericksonn, Lina and Hayek, Alan (2007). "What are Degrees of Belief?", *Studia Logica* 86, Special Issue Formal Epistemology I, Branden Fitelson (ed.), p. 183-213.

Hayek, Alan (2005). "Scotching Dutch Books?", Philosophical Perspectives 19, John Hawthorne (ed.), p. 139-51.

Hajek, Alan (2009). "Dutch Book Arguments", in Paul Anand, Prasanta K. Pattanaik, Clemens Puppe (ed.), *The Handbook of Rational and Social Choice: an overview of new foundations and applications*, Oxford University Press, Oxford, p. 173-195.

Hellman, Geoffrey (1997). "Bayes and Beyond." Philosophy of Science, Vol. 64, No. 2, (June), p. 191-221.

Howson, Colin and Urbach, Peter (1993). Scientific Reasoning: The Bayesian Approach. Second Edition, Chicago: Open Court.

Howson, Colin and Urbach, Peter (2005). Scientific Reasoning: The Bayesian Approach. Third Edition, Chicago: Open Court, 2005.

Hunter, Daniel (1996). "On the Relation Between Categorical and Probabilistic Belief." Nous 30(1), p. 1.

Jeffrey, Richard (1977). Subjective Probability – The Real Thing. Cambridge University Press.

Kaplan, Mark (1996). Decision Theory as Philosophy. New York: Cambridge University Press.

Maher, Patrick (1993). Betting on Theories. Cambridge, MA: Cambridge University Press.

Maher, Patrick (1997). "Depragmatized Dutch Book Arguments." Philosophy of Science, 64 (June), p. 291-305.

Ramsey, Frank P (1990). "Truth and Probability," in Frank Ramsey, *Philosophical Papers,* edited by D. H. Mellor. Cambridge: Cambridge University Press, p. 52-109.

Skyrms, B. (1984), Pragmatics and Empiricism. New Haven: Yale University Press.

Zynda, Lyle (2000). "Representation Theorems and Realism about Degrees of Belief." *Philosophy of Science* 67: p. 45-69.