ARGUMENTS FOR – OR AGAINST - PROBABILISM?

Alan Hájek

0. Introduction

On Mondays, Wednesdays, and Fridays, I call myself a *probabilist.*[1] In broad outline I agree with probabilism's key tenets: that

1) an agent's beliefs come in degrees, which we may call *credences;* and that

2) these credences are rationally required to conform to the probability calculus.

Probabilism is a simple, fecund theory. Indeed, it achieves such an elegant balance of simplicity and strength that, in the spirit of Ramsey's and Lewis's accounts of 'law of nature'[2], I am inclined to say that probabilism codifies the laws of epistemology. Or so I am inclined on those days of the week.

But on Tuesdays, Thursdays, and Saturdays, I am more critical of probabilism. A number of well-known arguments are offered in its support, but I find each of them wanting. I do not have the space here to spell out all of the arguments, and all of my sources of dissatisfaction with them. Instead, I will confine myself to four of the most important arguments—the Dutch Book, representation theorem, calibration, and gradational accuracy arguments—and I will concentrate on a particular source of dissatisfaction that I find in each of them.

(On Sundays I would like to rest, but often I find myself wondering on which other days of the week I am right.)

I think that it is underappreciated how structurally similar these four arguments for probabilism are. Each begins with a mathematical theorem that adverts to credences or degrees of belief, and that has the form of a *conditional* with an *existentially quantified consequent*. The antecedent speaks of the credences of some agent that violate the probability calculus. The consequent states the existence of *something putatively undesirable* that awaits such an agent, some way in which the agent's lot is *worse* than it could be by obeying the probability calculus, in a way that allegedly impugns her rationality. In each case, I will not question the theorem. But each argument purports to derive probabilism from the theorem. And it is *highly* underappreciated that in each case the argument, as it has been standardly or canonically presented, is invalid.

The trouble in each case is that there is a *mirror-image* theorem, equally beyond dispute, that undercuts probabilism; if we focus on *it,* we apparently have an argument *against* probabilism, of exactly equal strength to the original argument *for* probabilism. The original theorem provides good news for probabilism, but the mirror-image theorem provides bad news. Our final verdict can only be given after weighing the good and the bad news. Unfortunately for probabilism, the measures of good and bad news are exactly equal, so offhand it appears that the arguments for and against probabilism exactly counterbalance each other. The probabilist must then look elsewhere for more good news. In some cases it can be found, thus closing some of the loopholes that were left open in the original arguments.

---

[1] Much like Earman (1992), p. 1.

[2] The Ramsey/Lewis account has it that a law of nature is a theorem of the best theory of the universe—the true theory that best balances simplicity and strength.

1. The Dutch Book argument

The Dutch Book argument assumes that credences can be identified with corresponding betting prices. Your degree of belief in *X* is *p* iff you are prepared to buy or sell at $*p* a bet that pays $1 if *p*, and nothing otherwise. We may call *p* the price that you consider *fair* for the bet on *X*—at that price, you are indifferent between buying and selling the bet, and thus you see no advantage to either side. The betting interpretation, of course, involves a good deal of idealization, but I won't begrudge it here. (I begrudge it enough elsewhere.[3]) Instead, I will question the validity of the argument.

The centerpiece of the argument, as it has repeatedly been stated, is the following theorem, which I will not dispute:

<u>Dutch Book Theorem</u>
If you violate probability theory, there exists a set of bets, each of which you consider fair, which collectively guarantee your loss.

Call an agent who violates probability theory *incoherent.*[4] Call a set of bets, each of which you consider fair, and which collectively guarantee your loss, a *Dutch Book* against you. The Dutch Book theorem tells us that if you are incoherent, there exists a Dutch Book against you. Note the logical form: a conditional with an existentially quantified consequent. The antecedent speaks of a violation of probability theory; the consequent states the existence of something bad that follows from such a violation. We will see this form again and again.

So much for the theorem. What about the argument for probabilism? Prompted by Denis Robinson's and David Chalmers' invitations for presentations of famous philosophical arguments entirely in words of one syllable[5], I offer the following:

> You give some chance to *p:* it is the price that you would pay for a bet that pays a buck if *p* is true and nought if *p* is false. You give some chance to *q:* it is the price that you would pay for a bet that pays a buck if *q* is true and nought if *q* is false. And so on. Now, if you failed to live up to the laws of chance, then you would face a dire end. A guy—let's make him Dutch—could make a set of bets with you, each fair by your lights, yet at the end of the day you would lose, come what may. What a fool you would be! Do not tempt this fate. Be sure, then, to bet in line with the laws of chance!

This argument is *clearly* invalid. For all the Dutch Book theorem tells us, you may be just as susceptible to Dutch Books if you *obey* probability theory. Maybe the world is a tough place, and we're all suckers! This loophole is closed by the strangely neglected, yet equally important Converse Dutch Book theorem: if you obey probability theory, then there does *not* exist a Dutch Book against you. So far, so good for probabilism.

But nothing can rule out the following mirror-image theorem. In honor of my ethnic heritage, and with an eye to the financial gains that are in the offing, let's call it

[3] In Eriksson and Hájek (2007) and Hájek (2007c).
[4] De Finetti used the word "incoherent" to mean "Dutch bookable", while some other authors use it as I do. It will be handy for me to have this snappy word at my disposal even when I am not discussing Dutch books.
[5] See http://fragments.consc.net/djc/2005/02/phil_in_words_o.html.

the
<u>Czech Book Theorem</u>
If you *violate* probability theory, there exists a set of bets, each of which you consider fair, which collective guarantee your *gain*.

The proof of the theorem is easy: just rewrite the proof of the original Dutch Book theorem, replacing 'buying' by 'selling' of bets, and vice versa, throughout. You thereby turn the original 'Dutch Bookie' who milks you into a 'Czech Bookie' whom you milk. Call a set of bets, each of which you consider fair, and which collectively guarantee your gain, a *Czech Book* for you. The Czech Book theorem tells us that if you are incoherent, there exists a Czech Book for you. It is a simple piece of mathematics, and there is no disputing it.

So much for the theorem. I now offer the following argument *against* probabilism, again in words of one syllable. It starts as before, then ends with a diabolical twist:

> … Now, if you failed to live up to the laws of chance, then you would face a sweet end. A guy—let's make him Czech—could make a set of bets with you, each fair by your lights, yet at the end of the day you would *win*, come what may. What a brain you would be! Seek this fate. Be sure, then, *not* to bet in line with the laws of chance!

This argument is clearly invalid. For all the Czech Book theorem tells us, you may be just as open to Czech Books if you obey probability theory. Maybe the world is a nice place, and we're all winners! This loophole is closed by the strangely neglected, yet equally important Converse Czech Book theorem: if you obey probability theory, then there does *not* exist a Czech Book for you.[6] So far, so bad for probabilism.

Let's take stock, putting the theorems side by side:

Iff you violate probability theory, there exists a *specific bad thing* (a Dutch Book against you).

Iff you violate probability theory, there exists a *specific good thing* (a Czech Book for you).

The Dutch Book argument sees the incoherent agent's glass as half empty, while the Czech Book argument sees it as half full. If we focus on the former, probabilism prima facie looks compelling; but if we focus on the latter, the denial of probabilism prima facie looks compelling.

<u>Saving the Dutch Book argument</u>

This sub-section mostly repeats the corresponding section of my (2005)—for this move in the dialectic I have nothing more, nor less, to say than I did there.

For some reason, most of the presenters, both sympathetic and unsympathetic, of the Dutch Book argument that I am aware of focus solely on bets bought or sold at exactly your *fair* prices, bets that you consider fair. The list, containing many a luminary in the philosophy of probability, includes: Adams (1962), Adams and Rosenkrantz (1980), Armendt (1992), Baillie (1973), Carnap (1950, 1955), Christensen (1991, 1996, 2001), de Finetti (1980), Döring (2000), Earman (1992),

[6] Proof: Suppose for reductio that you obey probability theory, and that there is a set of bets, each of which you consider fair, which collectively guarantee your gain. Then swapping sides of these bets, you would still consider each fair, yet collectively they would guarantee your loss. This contradicts the Converse Dutch Book theorem.

Gillies (2000), Howson and Urbach (1993), Jackson and Pargetter (1976), Jeffrey (1983, 1992), Kaplan (1996), Kemeny (1955), Kennedy and Chihara (1979), Lange (1999), Lehman (1955), Maher (1993), Mellor (1971), Milne (1990), Rosenkrantz (1981), Seidenfeld and Schervish (1983), Skyrms (1986), van Fraassen (1989), Weatherson (1999), Waidacher (1997), Williamson (1999), and my younger self in numerous undergraduate lectures. But bets that you consider fair are not the only ones that you accept; you also accept bets that you consider favorable—that is, better than fair. You are prepared to sell a given bet at higher prices, and to buy it at lower prices, than your fair price. This observation is just what we need to break the symmetry that deadlocked the Dutch Book argument and the Czech Book argument.

Let us rewrite the theorems, replacing ‘fair’ with ‘fair-or-favorable’ throughout, and see what happens:

Dutch Book theorem, revised:

If you violate probability theory, there exists a set of bets, each of which you consider fair-or-favorable, which collectively guarantee your loss.

Converse Dutch Book theorem, revised:

If you obey probability theory, there does not exist a set of bets, each of which you consider fair-or-favorable, which collectively guarantee your loss.

Czech Book theorem, revised:

If you violate probability theory, there exists a set of bets, each of which you consider fair-or-favorable, which collective guarantee your gain.

Converse Czech Book theorem, revised:

If you obey probability theory, there does not exist a set of bets, each of which you consider fair-or-favorable, which collectively guarantee your gain.

The first three of these revisions are true, obvious corollaries of the original theorems. Indeed, the revised versions of the Dutch Book theorem and the Czech Book theorem follow immediately, because any bet that you consider fair you ipso facto consider fair-or-favorable. The revised version of the Converse Dutch Book theorem also follows straightforwardly from the original version.[7]

But the revised version of the Converse Czech Book theorem is not true: if you obey probability theory, there *does* exist a set of bets, each of which you consider fair-or-favorable, that collectively guarantee your gain. The proof is trivial.[8] The revision

---

[7] Proof. Suppose you obey probability theory. Suppose for reductio that there does exist a set of bets, each of which you consider fair-or-favorable, that collectively guarantee a loss; let this loss be $L > 0$. Then you must regard at least one of these bets as favorable (for the Converse Dutch Book theorem assures us that if you regarded them all as fair, then there could not be such guaranteed loss). That is, at least one of these bets is sold at a higher price, or bought at a cheaper price, than your fair price for it. For each such bet, replacing its price by your fair price would increase your loss. Thus, making all such replacements, so that you regard all the bets as fair, your guaranteed loss is even greater than L, and thus greater than 0. This contradicts the Converse Dutch Book theorem. Hence, we must reject our initial supposition, completing the reductio. We have proved the revised version of the Converse Dutch Book theorem.

[8] Suppose you obey the probability calculus; then if T is a tautology, you assign $P(T) = 1$. You consider fair-or-favorable paying less than $1—e.g, 80 cents—for a bet on T at a $1 stake, simply because you regard it as favorable; and this bet guarantees your

from 'fair' to 'fair-or-favorable' makes all the difference. And with the failure of the revised version of the Converse Czech Book theorem, the corresponding revised version of the Czech Book argument is invalid. There were no Czech Books for a coherent agent, because Czech Books were defined in terms of *fair* bets. But there are other profitable books besides Czech Books, and incoherence is not required in order to enjoy those. Opening the door to fair-or-favorable bets opens the door to sure profits for the coherent agent. So my parody no longer goes through when the Dutch Book argument is cast in terms of fair-or-favorable bets, as it always should have been.

I began this section by observing that most of the presenters of the Dutch Book argument formulate it in terms of your fair prices. You may have noticed that I left Ramsey off the list of authors.[9] His relevant remarks are confined to "Truth and Probability", and what he says is somewhat telegraphic:

> If anyone's mental condition violated these laws [of rational preference, leading to the axioms of probability], his choice would depend on the precise form in which the options were offered him, which would be absurd. He could have a book made against him by a cunning bettor and would then stand to lose in any event … Having degrees of belief obeying the laws of probability implies a further measure of consistency, namely such a consistency between the odds acceptable on different propositions as shall prevent a book being made against you. (1980, 42)

Note that Ramsey does not say that all of the bets in the book are individually considered fair by the agent. He leaves open the possibility that some or all of them are considered better than fair; indeed "acceptable" odds is synonymous with "fair-or-favorable" odds. After all, one would accept bets not only at one's fair odds, but also at better odds. Ramsey again:

> By proposing a bet on p we give the subject a possible course of action from which so much extra good will result to him if *p* is true and so much extra bad if *p* is false. Supposing the bet to be in goods and bads instead of in money, he will take a bet at any better odds than those corresponding to his state of belief; in fact his state of belief is measured by the odds he will just take;... (1980, 37).

It was the subsequent authors who restricted the Dutch Book argument solely to fair odds. They did Ramsey, and the Dutch Book argument itself, a disservice.

(This ends the sub-section that was mostly lifted from my (2005); the remainder of this paper is again new.)

---

gain.

[9] Skyrms (1986) was on the list, but not Skyrms (1980, 1984, or 1987). For example, in his (1987) he notes that an agent will buy or sell contracts "at what he considers the fair price or better" (p. 225), and in his (1980), he explicitly states the Dutch Book theorem in terms of "fair or favorable" bets (p. 118). Shimony, (1955), Levi (1974), Kyburg (1978), Armendt (1993), Douven (1999), and Vineberg (2001) also leave open that the bets concerned are regarded as favorable. It is hard to tell whether certain other writers on the Dutch Book argument belong on the list or not (e.g., Ryder 1981, Moore 1983).

‘The Dutch Book argument merely dramatizes an inconsistency in the valuations of an agent whose credences violate probability theory’

Before leaving the Dutch Book argument, there is one other important interpretation of it, also deriving from Ramsey, which requires some discussion. Recall his famous line: “If anyone’s mental condition violated these laws [of rational preference, leading to the axioms of probability], his choice would depend on the precise form in which the options were offered him, which would be absurd.” Authors such as Skyrms (1984) and Armendt (1993) regard this is as the real insight of the Dutch Book argument: an agent who violates probability theory would be guilty of a kind of double-think, “divided-mind inconsistency” in Armendt’s phrase. Such authors downplay the stories of mercenary Dutch guys and sure monetary losses; these are said merely to dramatize that underlying state of inconsistency. Skyrms describes the Dutch Book theorem as “a striking corollary” of an underlying inconsistency inherent in violating the probability axioms (1984, 22). The inconsistency is apparently one of regarding a particular set of bets both as fair (since they are regarded individually as fair) and as unfair (since they collectively yield a sure loss).

Notice that put this way, there is no need to replace talk of ‘fair’ bets with ‘fair-or-favorable’ bets, the way there was before. But we could do so: the inconsistency equally lies in regarding the same set of bets both as fair-or-favorable and as not fair-or-favorable. Moreover, there is nothing essentially *Dutch* about the argument, interpreted this way. The Czech Book theorem is an equally striking corollary of the same underlying inconsistency: regarding another set of bets both as fair (since they are regarded individually as fair) and as *better-than-fair* (since they collectively yield a sure gain). To be sure, guaranteed losses may be more *dramatic* than guaranteed gains, but the associated double-think is equally bad.

So now the real argument for probabilism seems not to stem from the Dutch Book theorem (which is merely a “corollary”), but from another putative theorem, apparently more fundamental. I take it to be this: *If you violate probability theory, there exists a set of propositions (involving bets) to which you have inconsistent attitudes*. Either the Dutch Book bets or the Czech Book bets could be used to establish the existence claim. This again is a conditional with an existentially quantified consequent. Now I don’t have a mirror-image theorem to place alongside it, in order to undercut it.

However, nor have I seen the *converse* of this more fundamental putative theorem, still less am I aware of anyone claiming to have proved it. It seems to be a live possibility that if you obey probability theory, then there also exists a set of propositions to which you have inconsistent attitudes—not inconsistent in the sense of being Dutch-bookable (the converse Dutch Book theorem assures us of this), but inconsistent nonetheless. That is, I have not seen any argument that in virtue of avoiding the inconsistency of Dutch-bookability, the probabilistic coherent agent is guaranteed to avoid all inconsistency. Without a proof of this further claim, it seems an open question whether probabilistically coherent agents might *also* have inconsistent evaluations (somewhere or other).

Moreover, I have called the conditional a ‘*putative* theorem’ because its status as a theorem is less clear than before—this status *is* disputed by various authors. Schick (1986) and Maher (1993) question the inconsistency of the attitudes at issue regarding the additivity axiom. They reject the ‘package principle’, which requires one to value a set of bets at the sum of the values of the bets taken individually, or less specifically, to regard a set of bets as fair if one regards each bet individually as fair.

In the case of violations of the normalization axiom, it is even less clear that there is an inconsistency of attitudes. (Compare Vineberg 2001.) Suppose that you assign probability 0.8 to some very complicated sentence, which unbeknownst to you is in fact a tautology. You have a *single* attitude to this sentence: you consistently assign it 0.8, come what may. To be sure, you are guilty of some failure of ideal rationality, namely lack of logical omniscience. But it seems to be a lesser sin than inconsistency —it's an omissive error, rather than a commissive error, and not clearly a case of double-think at all. For further discussion, see Hájek (2007c).

2. Representation theorem-based arguments

The centerpiece of the argument for probabilism from representation theorems is some version of the following theorem, which I will not dispute:

<u>Representation theorem</u>

> *If all your preferences satisfy certain 'rationality' conditions, then there exists a representation of you as an expected utility maximizer, relative to some probability and utility function.*

(The 'rationality' constraints on preferences are transitivity, connectedness, independence, and so on.) The contrapositive gets us closer to the template that I detect in all the arguments for probabilism:

> *If there does not exist a representation of you as an expected utility maximizer, relative to some probability and utility function, then there exist preferences of yours that fail to satisfy certain 'rationality' conditions.*

Focusing on the probabilistic aspect of the antecedent, we have a corollary that fits the conditional-with-an-existentially-quantified-consequent form:

> *If your credences cannot be represented with a probability function, then there exist preferences of yours that fail to satisfy certain 'rationality' conditions.*

The antecedent involves a violation of the probability calculus; the consequent states the existence of a putatively undesirable thing that follows: some violation of the 'rationality' conditions on preferences. In short, *if your credences cannot be represented with a probability function, then you are irrational*.

I will dispute that probabilism follows from the original theorem, and *a fortiori* that it follows from the corollary. For note that probabilism is, in part, the stronger thesis that *if your credences violate probability theory, then you are irrational* (a restatement of what I called tenet 2) at the outset). It is clearly a stronger thesis than the corollary, because its antecedent is weaker: while 'your credences cannot be represented with a probability function' entails 'your credences violate probability theory', the converse entailment does not hold. For it is possible that your credences violate probability theory, and that nonetheless they can be *represented* with a probability function. Merely being *representable* some way or other is cheap, as we will see; it's more demanding actually to *be* that way. Said another way: it's one thing to act *as if* you have credences that obey probability theory, another thing to actually *have* credences that obey probability theory. Indeed, probabilism does not even follow from the theorem coupled with the premises that Maher adds in his meticulous presentation of his argument for probabilism, as we will also see.

The concern is that for all we know, the mere *possibility* of representing you one way or another might have less force than we want; your acting *as if* the representation is true of you does not make it true of you. To make this concern vivid, suppose that I represent your preferences with *Voodooism*. My voodoo theory says that there are warring voodoo spirits inside you. When you prefer *A* to *B*, then there are more *A*-favoring spirits inside you than *B*-favoring spirits. I interpret all of the usual rationality axioms in voodoo terms. Transitivity: if you have more *A*-favoring spirits than *B*-favoring spirits, and more *B*-favoring spirits that *C*-favoring spirits, then you have more *A*-favoring spirits than *C*-favoring spirits. Connectedness: any two options can be compared in the number of their favoring spirits. And so on. I then 'prove' Voodooism: if your preferences obey the usual rationality axioms, then there exists a Voodoo representation of you. That is, you act *as if* there are warring voodoo spirits inside you in conformity with Voodooism. Conclusion: rationality requires you to have warring Voodoo spirits in you. Not a happy result.

Hence the need to bridge the gap between the possibility of representing a rational agent a particular way, and this representation somehow being *correct*. Maher, among others, attempts to bridge this gap. I will focus on his presentation, because he gives one of the most careful formulations of the argument. But I suspect my objections will carry over to any version of the argument that infers the rational *obligation* of having credences that are probabilities from the mere represent*ability* of an agent with preferences obeying certain axioms.

Maher claims that the expected utility representation is *privileged*, superior to rival representations. First, he assumes what I will call *interpretivism:*

> 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. (1993, 9).

Then he maintains that, when available, an expected utility interpretation is a *perfect* interpretation:

> if a person's preferences all maximize expected utility relative to some *p* and *u,* then it provides a perfect interpretation of the person's preferences to say that *p* and *u* are the person's probability and utility functions.

He goes on to give the argument from the representation theorems:

> … we can show that rational persons have probability and utility functions if we can show that rational persons have preferences that maximize expected utility relative to some such functions. An argument to this effect is provided by representation theorems for Bayesian decision theory.

He then states the core of these theorems:

> These theorems show that if a person's preferences satisfy certain putatively reasonable qualitative conditions, then those preferences are indeed representable as maximizing expected utility relative to some probability and utility functions. (1993, 9).

We may summarize this argument as follows:

<u>Representation theorem argument</u>

1. (Interpretivism) You have a particular probability and utility 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.

Thus, (generalizing what has been established about 'you' to 'all rational persons')

Conclusion: "[All] rational persons have probability and utility functions" (9).

The conclusion is probabilism, and a bit more, what we might call *utilitism.*

Elsewhere (Hájek 2007a), I question every premise except 3. Here I will confine myself to:

The invalid inference to probabilism

According to Premise 1, a necessary condition for you to have a particular probability and utility function is their providing an interpretation of you that is *better* than any competing interpretation. Suppose we grant, as I elsewhere do not, that the expected utility representation is a perfect interpretation when it is available. To validly infer probabilism, we need also to show that *no other interpretation is as good.* Perhaps this can be done, but nothing in Maher's argument does it. For all that he has said, there are other perfect interpretations out there (whatever that means).

Probabilism would arguably follow from the representation theorem if *all* representations of the preference-axiom-abiding agent were probabilistic representations.[10] Alas, this is not the case, for the following 'mirror-image' theorem is equally true:

If all your preferences satisfy the same 'rationality' conditions, then you can be interpreted as maximizing non-expected utility, some rival to expected utility, and in particular as having credences that violate probability theory.

How can this be? The idea is that the rival representation compensates for your credences' violation of probability theory with some non-standard rule for combining your credences with your utilities. Zynda (2000) proves this mirror-image theorem. As he shows, if you obey the usual preference axioms, you can be represented with a sub-additive belief function, and a corresponding combination rule. For all that Maher's argument shows, this rival interpretation may also be "perfect".

Zynda's rival interpretation is just the beginning. Abstracting away from the details of this or that expected utility theory (Ramsey's, Savage's, von Neumann and Morgenstern's, Jeffrey's, Joyce's, …), any such theory requires the rational agent to maximize a weighted average of utilities, with weights provided by corresponding probabilities:

$$EU(A_i) = \sum_j u(S_j \,\&\, A_i).P(S_j),$$

[10] Only arguably. I argue (2007a) that it does not follow, because the preference axioms are not all rationality constraints.

where $EU(A_i)$ is the expected utility of action A_i. In short, an expectation is a sum of products of utilities and probabilities. But we can recover the same preference ordering with a sum of products of *schmutilities* and *schmobabilities:*[11]

$$EU(A_i) = \sum_j [f(A_i,S_j)) u(S_j \,\&\, A_i)] \left[\frac{P(S_j)}{f(A_i,S_j)}\right]$$

Here f can be any non-zero function of A_i and S_j that you like. The first square-brackets contain the schmutility term; the second contain the schmobability term. Expected utility theory is the special case in which f is identically 1. The rational agent is *representable* in terms of schmutilities and schmobabilities. But merely being *representable* some way or other is cheap; it's more demanding actually to *be* that way. Said another way: it's one thing to act *as if* you have credences that obey schmobability theory, another thing to actually *have* credences that obey schmobability theory. In particular, it doesn't follow this rival representation that credences are rationally required to obey schmobability theory.

According to probabilism, rationality requires an agent's *credences* to obey the probability calculus. We have many rival ways of representing an agent whose preferences obey the preference axioms; which of these representations correspond to her *credences?* In particular, why should we privilege the probabilistic representation? Well, there may be reasons. Perhaps it is favored by considerations of simplicity, fertility, consilience, or some other theoretical virtue or combination thereof—although good luck trying to clinch the case for probabilism by invoking these rather vague and ill-understood notions. And it is not clear that these considerations settle the issue of what rational credences *are*, as opposed to how they can be fruitfully modeled.[12] It seems to be a further step, and a dubious one at that, to reify the theoretical entities in our favorite model of credences. In any case, the point remains that the representation theorem argument is invalid as it stands.

3. The calibration argument

The centerpiece of the argument is the following theorem—another conditional with an existentially quantified consequent—which I will not dispute:

Calibration theorem

If c violates the laws of probability then there is a probability function c^+ that is better calibrated than c under every logically consistent assignment of truth-values to propositions.

Calibration is a measure of how well credences match corresponding relative frequencies. Suppose that you assign probabilities to some sequence of propositions – for example, each night you assign a probability to it raining the following day, over a period of a year. Your assignments are *(perfectly) calibrated* if proportion p of the propositions to which you assigned probability p are true, for all p. In the example, you are perfectly calibrated if it rained on 0.1 of the days to which you assigned probability 0.1, on 0.75 of the days to which you assigned probability 0.75, and so on. More generally, we can measure how well calibrated your assignments are, even if they fall short of perfection.

---

[11] Jim Joyce pointed out a version of this trick to me.

[12] See Eriksson and Hájek (2007) for further discussion.

The clincher for probabilism is supposed to be the calibration theorem. If you are incoherent, then you can figure out *a priori* that you could be better calibrated by being coherent instead. Perfect calibration, moreover, is supposed to be A Good Thing, and a credence function that is better calibrated than another one is thereby supposed to be superior in at least one important respect. Thus, the argument concludes, you should be coherent. See Joyce (2004) for a good exposition of this style of argument for probabilism (although he does not endorse it himself).

I argue elsewhere (Hájek 2007b) that perfect calibration may be A Rather Bad Thing, as does Seidenfeld (1985) and Joyce (1998). More tellingly, the argument, so presented, is invalid.

<u>The calibration argument is invalid</u>

I will not quarrel with the calibration theorem. But I draw to your attention what I take to be a 'mirror-image' theorem:

> *If c violates the laws of probability then there is a NON-probability function c^+ that is better calibrated than c under every logically consistent assignment of truth-values to propositions.*

Think of c^+ as being more coherent that *c*, but not entirely coherent. If *c* assigns, say, 0.2 to rain and 0.7 to not-rain, then an example of such a c^+ is a function that assigns 0.2 to rain and 0.75 to not-rain. If you are incoherent, then you know *a priori* that you could be better calibrated by *staying in*coherent, but in some other way. To be sure, the mirror-image theorem gives you no advice as to which non-probability function you should move to. But nor did the calibration theorem give you advice as to which probability function you should move to. Moreover, for all the theorem tells us, you can *worsen* your calibration index, come what may, by moving from a non-probability function to a 'wrong' probability function. Here's an analogy (adapted from Aaron Bronfman and Jim Joyce). Suppose that you want to live in the best city that you can, and you currently live in an American city. I tell you that for each American city, there is a better Australian city. (I happen to believe this.) It does not follow that you should move to Australia. If you do not know *which* Australian city or cities are better than yours, moving to Australia might be a backward step. You might choose Coober Pedy.

Moreover, given that you can improve your calibration situation *either* by moving to some probability function *or* by moving to some other non-probability function, why do you have an incentive to move to a probability function? The answer, I suppose, is this. If you moved to a non-probability function, you would only recreate your original predicament: you would know a priori that you could do better by moving to a probability function. Now again, you could *also* do better by moving to yet another non-probability function. But the idea is that moving to a non-probability function will give you no rest; it can never be a stable stopping point. Still, the argument for probabilism is invalid as it stands. To shore it up, we had better be convinced that probability functions *are* stable stopping points.

We thus need the following theorem:

> If *c obeys* the laws of probability then there is *not* another function c^+ that is better calibrated than *c* under every logically consistent assignment of truth-values to propositions.

I offer the following near-trivial proof: Let *P* be a probability function. *P* can be perfectly calibrated—just consider a world where the relative frequencies are exactly as *P* predicts, as required by calibration. (If *P* assigns some irrational probabilities, then the world will have to provide infinite sequences of the relevant trials, and calibration will involve agreement with limiting relative frequencies.) At that world, no other function can be better calibrated than P. Thus, *P* cannot be beaten by some other function, come what may, in its calibration index.

Putting this result together with the calibration theorem, we have the result that *probability functions are exactly the functions that are not beaten come what may by any other function in their calibration index.*

The original calibration argument for probabilism, as stated above, was invalid, but I think it can be made valid by the addition of this theorem. But this is not yet a happy ending for calibrationists. If you are a fan of calibration, surely what matters is being well calibrated in the *actual* world, and being coherent does not guarantee that.[13] A coherent weather forecaster who is wildly out of step with the actual relative frequencies can hardly plead that at least he is perfectly *in* step with the relative frequencies *in some other possible world!* We will see a version of this problem in the next, and final, argument for probabilism.

4. The gradational accuracy argument

Joyce (1998) rightly laments the fact that "probabilists have tended to pay little heed to the one aspect of partial beliefs that would be of most interest to epistemologists: namely, their role in representing the world's state" (576). And he goes on to say: "I mean to alter this situation by first giving an account of what it means for a system of partial beliefs to accurately represent the world, and then explaining why having beliefs that obey the laws of probability contributes to the basic epistemic goal of accuracy."

The centerpiece of his ingenious (1998) argument is the following theorem—yet another conditional with an existentially quantified consequent—which I will not dispute:

Gradational accuracy theorem

> "if *c* violates the laws of probability then there is a probability function c^+ that is strictly more accurate than *c under every logically consistent assignment of truth-values to propositions."* (Joyce 2004, 143)

Joyce gives the following account of the argument. It

> relates probabilistic consistency to the *accuracy* of graded beliefs. The strategy here involves laying down a set of axiomatic constraints that any reasonable gauge of accuracy for confidence measures should satisfy, and then showing that probabilistically inconsistent measures are always less accurate

[13] Seidenfeld (1985) has a useful discussion of a theorem due to Pratt and rediscovered by Dawid that may seem to yield the desired result. Its upshot is that if an agent is probabilistically coherent, and updates by conditionalization after each trial on feedback information about the result of that trial, then in the limit calibration is achieved almost surely (according to her own credences). This is an important result, but it does not speak to the case of an agent who is coherent but who has not updated on such an infinite sequence of feedback information, and indeed who may never do so (e.g., because she never gets such information).

than they need to be. (2004, 142)

With the constraints on reasonable gauges of accuracy in place, Joyce (1998) proves the gradational accuracy theorem. He concludes: "To the extent that one accepts the axioms, this shows that the demand for probabilistic consistency follows from the purely epistemic requirement to hold beliefs that accurately represent the world" (2004, 143).

Let us agree for now that the axioms are acceptable. (Maher 2003 doesn't.) I have already agreed that the theorem is correct. But I do not agree that the demand for probabilistic consistency follows.

Again, we have a 'mirror-image' theorem: if c violates the laws of probability then there is a NON-probability function c^+ that is strictly more accurate than c *under every logically consistent assignment of truth-values to propositions.* (As with the corresponding calibration theorem, the trick here is to make c^+ *less* incoherent than c, but incoherent nonetheless.) If you are incoherent, then your beliefs could be made more accurate by moving to another incoherent function. Why, then, are you under any rational obligation to move instead to a coherent function? The reasoning, I gather, will be much as it was in our discussion of the calibration theorem. Stopping at a non-probability function will give you no rest, because by another application of the gradational accuracy theorem, you will again be able to do better by moving to a probability function. To be sure, you will *also* be able to do better by moving to yet another non-probability function. But a non-probability function can never be a stable stopping point: it will always be accuracy-dominated by some other function.[14]

So the (1998) argument is invalid as it stands. As before, to shore it up we had better prove that probability functions *are* stable stopping points. We need the following converse theorem:

> If c *obeys* the laws of probability then there is *not* another function c^+ that is strictly more accurate than c under every logically consistent assignment of truth-values to propositions.

And sure enough, I see that Joyce (this volume) proves the needed converse theorem, assuming a condition that he calls "propriety".[15] The subsequent debate will no doubt turn on the status of this condition. The upshot is that (assuming propriety), *probability functions are exactly the functions that are not accuracy-dominated.*[16]

But even that may not be enough. It is not much use to the incoherent agent to know that *some* probability functions would be better than her current state of opinion, if she doesn't know *which* probability functions they are. Recall the plight of the American who doesn't know which Australian city to move to. There may be 'Coober Pedy' probability functions out there.

At this point it might be tempting to say: "Moving to *any* probability function would improve her situation. For *wherever* she moves, she will no longer have to

---

14 I thank Selim Berker and Justin Fisher for independently making this point. It inspired the similar point in the previous section, too.

15 I had completed a first draft of this paper before I had a chance to see Joyce's new article, with its converse theorem. I thank him for sending it to me.

16 Thanks to Kenny Easwaran for putting the point this way. It inspired the wording of the similar point in the previous section, too.

worry about being accuracy-dominated by some other function." But is being accuracy-dominated really such a concern? Consider an agent who is just slightly incoherent, but almost perfectly in step with the truth-values of the actual world. For example, incoherent Nola assigns probability 0.99 to not being a brain in a vat, and 0.001 to being one. Compare her to coherent Brianna, who assigns 0 to not being a brain in a vat, and 1 to being one. To be sure, Brianna can't be beaten 'come what may'—she can't be beaten in a world where she is a brain in a vat. But whose credences would you rather have?

This is particularly telling for Joyce, who finds motivation for his 'non-pragmatic vindication of probabilism' in traditional epistemology's concern with the role of beliefs in representing the *actual* world's state, and thus its concern with the *truth* of beliefs. The basic epistemic goal is accuracy in the *actual* world. Traditional epistemology is rather less concerned with a belief's truth in *other* possible worlds. To the extent that Joyce wants to assimilate probabilistic epistemology to traditional epistemology, I think he needs to put extra weight on divergence from the truth in the *actual* world. The analogue of truth for credences that Joyce seeks is plausibly *accuracy*—not impossibility-of-accuracy-domination, which is the analogue of *consistency*.[17] I say that Nola is more accurate than Brianna. Impossibility-of-accuracy-domination scores them the other way round. Which verdict is more faithful to traditional epistemology?

But faithful or not, Joyce's approach surely *ought* to be of great interest to the epistemologist. After all, consistency is, or ought to be. And I suspect that wherever we find a Dutch Book argument for some thesis or other, it can be replaced by a Joyce-style gradational accuracy argument. Now there's a fertile research program!

5. Conclusion

I began by confessing my schizophrenic attitude to probabilism. I have argued that the canonical statements of the major arguments for it have needed some repairing. Why, then, am I sympathetic to it at the end of the day, or at least at the end of some days? Partly because I think that to some extent the arguments *can* be repaired, and I have canvassed some ways in which this can be done (although to be sure, I think that some other problems remain—see my papers listed in the References). Once repaired, they provide a kind of triangulation to probabilism. And once we get to probabilism, it provides us with many fruits. Above all, it forms the basis of a unified theory of decision and confirmation—it combines seamlessly with utility theory to provide a fully general theory of rational action, and it illuminates or even resolves various hoary paradoxes in confirmation theory.[18] I consider that to be the best argument for probabilism. Sometimes, though, I wonder whether it is good enough.

So on Mondays, Wednesdays and Fridays, I consider myself to be a probabilist. But as I write these words, today is Saturday.[19]

---

[17] In (2007b) I argue for another analogue of truth for credences: agreement with corresponding objective chances.

[18] See Earman (1992), Jeffrey (1992), and Howson and Urbach (1993), among others.

[19] For very helpful comments I am grateful to: Selim Berker, David Chalmers, James Chase, Andy Egan, Lina Eriksson, Jim Joyce, Christian List, Teddy Seidenfeld, Katie Steele, and especially Kenny Easwaran (who also suggested the name "Czech Book"), Justin Fisher, Franz Hüber, and Stephan Leuenberger.

Research School of Social Sciences
Australian National University
Canberra, ACT 0200
Australia

REFERENCES

Adams, Ernest (1962), "On Rational Betting Systems", *Archive für Mathematische Logik und Grundlagenforschung* 6, 7-29, 112-128.

Adams, Ernest W. and Roger D. Rosenkrantz (1980), "Applying the Jeffrey Decision Model to Rational Betting and Information Acquisition", *Theory and Decision* 12, 1-20.

Armendt, Brad (1992), "Dutch Strategies for Diachronic Rules: When Believers See the Sure Loss Coming", *PSA 1992*, vol. 1, eds. D. Hull, M. Forbes, K. Okruhlik, East Lansing: Philosophy of Science Association, 217-229.

Armendt, Brad (1993), "Dutch Books, Additivity and Utility Theory", *Philosophical Topics* 21, No. 1, 1-20.

Baillie, Patricia (1973), "Confirmation and the Dutch Book Argument", *The British Journal for the Philosophy of Science* 24, 393-397.

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

Carnap, Rudolf (1955), *Notes on Probability and Induction* (UCLA, Philosophy 249, Fall Semester 1955), Unpublished Manuscript, Carnap Archive, University of Pittsburgh, RC-107-23-01.

Christensen, David (1991), "Clever Bookies and Coherent Beliefs." *The Philosophical Review* C, No. 2, 229-247.

Christensen, David (1996), "Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers", *The Journal of Philosophy*, 450-479.

Christensen, David (2001), "Preference-Based Arguments for Probabilism." *Philosophy of Science* 68 (3): 356-376.

de Finetti, Bruno (1980), "Probability: Beware of Falsifications", in *Studies in Subjective Probability*, eds. H. E. Kyburg, Jr. and H. E. Smokler, Robert E. Krieger Publishing Company.

Döring, Frank (2000), "Conditional Probability and Dutch Books", *Philosophy of Science* 67 (September), 39,1-409.

Douven, Igor (1999), "Inference to the Best Explanation Made Coherent", *Philosophy of Science* 66 (Proceedings), S424-S435.

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

Eriksson, Lina and Alan Hájek (2007), "What is a Degree of Belief?", forthcoming in *Studia Logica*, special edited on epistemology edited by Branden Fitelson.

Gillies, Donald (2000), *Philosophical Theories of Probability,* Routledge.

Hájek, Alan (2005), "Scotching Dutch Books", *Philosophical Perspectives* (issue on epistemology), eds. John Hawthorne and Dean Zimmerman.

Hájek, Alan (2007a), "What Representation Theorems Do Not Show", MS.

Hájek, Alan (2007b), "A Puzzle About Degree of Belief", MS.

Hájek, Alan (2007c), "Dutch Book Arguments", in *The Oxford Handbook of Corporate Social Responsibility*. ed. Paul Anand, Prasanta Pattanaik, and

Clemens Puppe, forthcoming 2007.

Howson, Colin and Peter Urbach (1993), *Scientific Reasoning: The Bayesian Approach*. La Salle, Illinois, Open Court.

Jackson, Frank and Robert Pargetter (1976), "A Modified Dutch Book Argument", *Philosophical Studies* 29, 403-407.

Jeffrey, Richard (1983), *The Logic of Decision*, University of Chicago Press, 2nd ed.

Jeffrey, Richard (1992), *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Joyce, James M. (1998), "A Non-Pragmatic Vindication of Probabilism", *Philosophy of Science* 65, 575-603.

Joyce, James M. (2004), "Bayesianism", in Alfred R. Mele and Piers Rawlings, *The Oxford Handbook of Rationality,* Oxford: Oxford University Press.

Joyce, James M. (this volume), "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief".

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

Kennedy, Ralph and Charles Chihara (1979), "The Dutch Book Argument: Its Logical Flaws, Its Subjective Sources", *Philosophical Studies* 36, 19-33.

Kemeny, John (1955), "Fair Bets and Inductive Probabilities", *Journal of Symbolic Logic,* 20: 263-273

Kyburg, Henry (1978), "Subjective Probability: Criticisms, Reflections and Problems", *Journal of Philosophical Logic* 7, 157-180.

Lange, Marc (1999), "Calibration and the Epistemological Role of Bayesian Conditionalization", *The Journal of Philosophy*, 294-324.

Lehman, R. (1955), "On Confirmation and Rational Betting", *Journal of Symbolic Logic* 20, 251-262.

Levi, Isaac (1974), "On Indeterminate Probabilities", *Journal of Philosophy*71, 391-418.

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

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

Maher, Patrick (2003), "Joyce's Argument for Probabilism", *Philosophy of Science* 69, 73-81

Mellor, D. H. (1971), *The Matter of Chance*, Cambridge: Cambridge University Press.

Milne, Peter (1990), "Scotching the Dutch Book Argument", *Erkenntnis* 32, 105-126.

Moore, P. G. (1983), "A Dutch Book and Subjective Probabilities", *The British Journal for the Philosophy of Science* 34, 263-266.

Ramsey, Frank P., (1980/1931), "Truth and Probability", in *Foundations of Mathematics and other Essays,* R. B. Braithwaite (ed.), Routledge & P. Kegan 1931, 156-198; reprinted in *Studies in Subjective Probability*, H. E. Kyburg, Jr. and H. E. Smokler (eds.), 2nd ed., R. E. Krieger Publishing Company, 1980, 23-52; reprinted in *Philosophical Papers*, D. H. Mellor (ed.) Cambridge: University Press, Cambridge, 1990.

Rosenkrantz, Roger D. (1981), *Foundations and Applications of Inductive Probability*, Ridgeview Publishing Company.

Ryder, J. M. (1981), "Consequences of a Simple Extension of the Dutch Book Argument", *The British Journal for the Philosophy of Science* 32, 164-167.

Schick, Frederic (1986), "Dutch Bookies and Money Pumps", *Journal of Philosophy*

83, 112-119.
Seidenfeld, Teddy (1985), "Calibration, Coherence, and Scoring Rules", *Philosophy of Science,* Vol. 52, No. 2 (June), 274-294.
Seidenfeld, Teddy and Mark J. Schervish (1983), "A Conflict Between Finite Additivity and Avoiding Dutch Book", *Philosophy of Science* 50, 398-412.
Shimony, Abner (1955), "Coherence and the Axioms of Confirmation", *Journal of Symbolic Logic* 20, 1-28.
Skyrms, Brian (1980), "Higher Order Degrees of Belief", in *Prospects for Pragmatism*, ed. D.H. Mellor, Cambridge: Cambridge University Press, 109-37.
Skyrms, Brian (1984), *Pragmatics and Empiricism*, Yale University.
Skyrms, Brian (1986), *Choice and Chance*, Wadsworth, 3rd ed.
Skyrms, Brian (1987), "Coherence", in N. Rescher (ed.), *Scientific Inquiry in Philosophical Perspective*, Pittsburgh: University of Pittsburgh Press, 225-242.
van Fraassen, Bas (1984), "Belief and the Will", *Journal of Philosophy* 81, 235-256.
van Fraassen, Bas (1989), Laws and Symmetry, Oxford: Clarendon Press.
Vineberg, Susan (2001), "The Notion of Consistency for Partial Belief", *Philosophical Studies* 102 (February), 281-296.
Waidacher, C. (1997), "Hidden Assumptions in the Dutch Book Argument", *Theory and Decision* 43 (November), 293-312.
Weatherson, Brian (1999), "Begging the Question and Bayesians", *Studies in History and Philosophy of Science* 30A, 687-697.
Williamson, Jon (1999), "Countable Additivity and Subjective Probability", *The British Journal for the Philosophy of Science* 50, No. 3, 401-416.
Zynda, Lyle (2000), "Representation Theorems and Realism About Degrees of Belief", *Philosophy of Science* 67, 45-69.