SCOTCHING DUTCH BOOKS?

Alan Hájek
Australian National University

1. Introduction

The Dutch Book argument, like Route 66, is about to turn 80. It is arguably the most celebrated argument for subjective Bayesianism. Start by rejecting the Cartesian idea that doxastic attitudes are ‘all-or-nothing’; rather, they are far more nuanced degrees of belief, for short credences, susceptible to fine-grained numerical measurement. Add a coherentist assumption that the rationality of a doxastic state consists in its internal consistency. The remaining problem is to determine what consistency amounts to for credences. The Dutch Book argument, in a nutshell, says that if your credences do not obey the probability calculus, you are ‘incoherent’—susceptible to sure losses at the hands of a ‘Dutch Bookie’—and thus irrational. Conclusion: rationality requires your credences to obey the probability calculus.

And like Route 66, the fortunes of the Dutch Book argument have been mixed. Opinions on the argument are sharply divided. The list of its proponents is quite a ‘who’s who’ of philosophers of probability; they include de Finetti (1937, 1980), Carnap (1950, 1962, and more fully, 1955), Kemeny (1955), Lehman (1955), Shimony (1955), Adams (1962), Mellor (1971), Rosenkrantz (1981), van Fraassen (1989), Jeffrey (1983, 1992), and Gillies (2000). Ramsey (1926—hence the imminent 80th birthday) belongs at the head of the list, although we will see that the full story about him is more interesting than has been previously recognized. The argument has found some notable opponents too, including Kyburg (1978), Kennedy and Chihara (1979), Schick (1986), and Maher (1993, 1997); however, they are surely in the minority. In this paper I will begin by joining their ranks. The Dutch Book argument, in the form that has become a philosophical staple, is simply invalid, and to the extent that we have bought it for so many years we have, as it were, been Dutch Booked ourselves; indeed, in endorsing it we have been guilty of a certain form of incoherence. However, like Route 66, the argument can be reconstructed, or better still, restored. I will favor a version of the argument that avoids the defects of
the staple version—thus, I hope, a better argument for Bayesianism. I will conclude by arguing that it was implicit in Ramsey’s original presentation all along, and that it was subsequent expositors of the argument who got it wrong.

2. The Dutch Book argument

Let me make clear what I take the Dutch Book argument—the philosophi-
cal staple—to be, and thus, which argument it is that I claim is invalid. (In the final section I will consider a variant of the argument whose emphasis is rather different.) What follows should look completely familiar to any student of the argument. I begin with some background.

You are a rational agent who assigns credences to a non-empty set of sentences assumed to be closed under negation and disjunction—although nothing in the Dutch Book argument justifies these closure assumptions. We assume that your credences are sharp—as opposed to indeterminate, or vague—although again nothing in the argument justifies this assumption. (Levi 1974 and 1980 relaxes this assumption.) It is further assumed that your credences can be identified with your betting dispositions concerning the sentences. Your credence in \(X\) is \(p\) iff you are prepared to buy or to sell a bet that pays \(S\) units of utility (the ‘stake’) for \(pS\) (‘the price you consider fair’, for short ‘your fair price’ for the bet); equivalently, you are prepared to bet on or against \(X\) at odds \(p: 1/p\); at any stake. For example, if you assign a credence of 1/2 to this coin toss landing heads, and if for you utility is linear in money, then you are prepared to buy or to sell the bet

\[
\begin{align*}
&1 \text{ if heads} \\
&0 \text{ otherwise}
\end{align*}
\]

for 50 cents. (If for you utility is not linear in money, read the ‘$’ sign as representing one unit of utility, and ‘cents’ as hundredths of a unit.) So construed, a betting transaction involves four things: the proposition \(X\), the stake \(S\), your fair price \(pS\) for the bet—equivalently, your odds \(p: 1 - p\)—and the side of the bet that you take (buying or selling, betting on or against \(X\)). We can reduce this to three by allowing negative stakes, identifying positive stakes with your betting on the proposition, negative stakes with your betting against it.

The Dutch Book argument purports to show that rationality requires your credences to obey the probability calculus. That is, it requires that your credences can be represented as a probability function, \(P\), over a non-empty set of sentences \(S\) closed under negation and disjunction, where \(P\) satisfies the usual axioms:

1. **Non-negativity**: \(P(X) \geq 0\) for all \(X\) in \(S\).
2. **Normalization**: \(P(T) = 1\) for any tautology \(T\) in \(S\).
3. **Additivity**: \(P(X \lor Y) = P(X) + P(Y)\) for all \(X, Y\) in \(S\) such that \(X\) is incompatible with \(Y\).
It is important now to distinguish three things: the definition of a Dutch Book, the Dutch Book theorem (and its converse), and the Dutch Book argument. First, the definition. A Dutch Book is a set of bets, each of which you consider fair, that collectively guarantee your loss.2

Next, the theorem. If your credences violate the probability calculus, then there exists a Dutch Book against you. That is, if your credences disobey any of the above axioms, then there exists a set of bets, each of which you consider fair, that will result in a net loss for you however the world turns out. Said another way, if your credences violate the probability calculus, then there exists a set of propositions and stakes (positive or negative) such that betting at your odds at those stakes on those propositions, you will lose, whatever transpires.

I will not present the full proof of the theorem here, but I can easily convey how it works. We really have three sub-theorems, one for each axiom, with a separate proof of each (see Skyrms 1986). The most interesting one concerns the additivity axiom. Suppose, for example, that you violate additivity by assigning $P(A \lor B) > P(A) + P(B)$ for incompatible $A$ and $B$. Then you effectively overprice a bet on $A \lor B$ relative to bets on $A$ and on $B$, and you will suffer a sure loss if you buy the first bet (at a comparatively high price) and sell the second and third bets (at comparatively low prices). Specifically, you are prepared to buy—from the dreaded ‘Dutch Bookie’—a $1 bet on $A \lor B$ for $P(A \lor B)$ and to sell bets on $A$ and on $B$ for $P(A)$ and $P(B)$ respectively, for an initial loss of

$$[P(A \lor B) - (P(A) + P(B))] > 0.$$ 

But from then on your net wealth remains the same, whatever happens. If neither $A$ nor $B$ occurs, all bets lose; if exactly one of $A$ and $B$ occurs, the $1 that you pay out on that bet is offset by the $1 that you gain on the $A \lor B$ bet. This exhausts the cases (since $A$ and $B$ cannot both occur). And so it goes for all of the other ways in which you could violate the probability calculus: in each case you are susceptible to a Dutch Book.

Moreover, the converse theorem also holds: if your credences do not violate the probability calculus, then there does not exist a Dutch Book against you (Kemeny 1955, Lehman 1955). Thus, what I will call the complete Dutch Book theorem, which combines the Dutch Book theorem and its converse, asserts a biconditional between your violating the probability calculus and your receptivity to a Dutch Book. This is simply a piece of mathematics, and there is no disputing it.

Finally, the argument. It is thought by many to follow from the Dutch Book theorem that rationality requires you to obey the probability calculus. After all, it is said, your state of opinion would be an unhappy one if, in virtue of its purely intrinsic characteristics, it exposed you to sure losses at the hands of a Dutch Bookie. This inference is clearly too quick: for all the Dutch Book theorem says, you may be exposed to the sure losses of a Dutch Book whether or not you obey the probability calculus; indeed, for all the theorem says, your
losses may be even greater if you obey it. But the Converse Dutch Book theorem rules out these possibilities: adherence to the probability calculus shields you from the losses inherent in a Dutch Book. So it is supposed to follow from the complete Dutch Book theorem that you would be irrational to violate the probability calculus.

It is supposed to follow, but it does not.

3. The Good Book argument

I will show that the Dutch Book argument is invalid first by parodying it, then by identifying the flaw in both the parody and the argument itself. I call my parody the Good Book argument. It is rather younger than Route 66; in fact, I have not seen it before.

It is important now to distinguish three things: the definition of a Good Book, the Good Book theorem (and its converse), and the Good Book argument. First, the definition. A Good Book is a set of bets, each of which you consider fair, that collectively guarantee your gain.

Next, the theorem. If your credences violate the probability calculus, then there exists a Good Book for you. That is, if your credences disobey any of the above axioms, then there exists a set of bets, each of which you consider fair, that will result in a net gain for you however the world turns out. Said another way, if your credences violate the probability calculus, then there exists a set of propositions and stakes (positive or negative) such that betting at your odds at those stakes on those propositions, you will win, whatever transpires.

Moreover, the converse theorem also holds: if your credences do not violate the probability calculus, then there does not exist a Good Book for you. The proofs are straightforward: simply replace ‘buy’ by ‘sell’ and vice versa throughout the proofs of the Dutch Book theorem and its converse. You just swap sides with the Dutch Bookie in the original proofs, turning him into a Good Bookie. Thus, what I will call the complete Good Book theorem, which combines the Good Book theorem and its converse, asserts a biconditional between a violation of the probability calculus and receptivity to a Good Book. This is simply a piece of mathematics, and there is no disputing it.

Finally, the argument. It might be said to follow from the Good Book theorem that rationality requires you to violate the probability calculus. After all, it might be said, your state of opinion would be a happy one if, in virtue of its purely intrinsic characteristics, it exposed you to sure gains at the hands of a Good Bookie. This inference is clearly too quick: for all the Good Book theorem says, you may be exposed to the sure gains of a Good Book whether or not you disobey the probability calculus; indeed, for all the theorem says, your gains may be even greater if you obey it. But the Converse Good Book theorem rules out these possibilities: adherence to the probability calculus shields you from the gains inherent in a Good Book. So it follows from the complete Good Book theorem that you would be irrational to obey the probability calculus.
4. Diagnosis

What is wrong with the Good Book argument, and with it, the Dutch Book argument? We have before us a pair of biconditionals with existentially quantified consequents:

If you violate the probability calculus, there exists a SPECIFIC BAD THING (a Dutch Book against you).

If you violate the probability calculus, there exists a SPECIFIC GOOD THING (a Good Book for you).

Each argument focuses on one of the conditionals while ignoring the other. But we apparently have perfect symmetry here, the yin of the one mirrored by the yang of the other. The Dutch Book argument sees the incoherent agent’s glass as half empty, while the Good Book argument sees it as half full. Advocates of the Dutch Book argument seem to regard the possibility of sure losses as more compelling than the possibility of sure gains. It’s as if they are employing some sort of maximin reasoning, more concerned as they are by the prospect of loss than by the prospect of gain. But this should seem quite out of place coming from the mouth, or pen, of a *Bayesian*—someone who seeks to supplant a theory of decision under uncertainty, in which probabilities for outcomes are not assigned (thus requiring decision rules such as maximin), with a theory of decision under risk, in which they are assigned. An advocate of the Good Book argument, in turn, could regard the possibility of sure gains as more compelling than the possibility of sure losses, perhaps employing some sort of maximax reasoning. Like a particle and its antiparticle, the two arguments annihilate each other.

It is said that the Dutch Book argument exposes the incoherent agent as guilty of an evaluative inconsistency, valuing the same state of affairs in two different ways depending on how it is presented.¹ Someone who propounds the Dutch Book argument while baulking at the Good Book argument is not quite guilty of that sin, since they are of course not the same argument. Still, given the symmetry between them, these opposed evaluations seem to evince a kind of double-think, if not quite double-Dutch.

You might try to break the symmetry with an assumed *asymmetry* in the world. You might contend that it is easier to find Dutch Bookies than it is to find Good Bookies: the incoherent agents will have no trouble finding Dutch Bookies eager to exploit their incoherence, but they will not find Good Bookies for their incoherence to exploit. I will call this the *bad neighborhood defence*, because it assumes that agents are more likely to run into bookies of the ‘bad’ kind than of the ‘good’ kind.

I have often heard this line of argument but I find it puzzling, several times over. Firstly, the Dutch Book theorem is a mathematical result that concerns the existence of abstract objects of a certain sort: sequences of bets with certain properties. Susceptibility to a Dutch Book is a dispositional property of an agent, one that he or she has independently of what other people are out there, and what they are like—in fact there need not be other people out there.
at all. (Compare: solubility is a property that salt has independently of whether or not water is out there.) No surprise, then, that the theorem makes no mention of bookies, Dutch or Chinese, bad or otherwise. Yet now the argument brings on such dramatis personae, and moreover attributes to them self-interested or nefarious motives, getting their kicks at the agent’s expense. No piece of mathematics can derive the existence of such characters, let alone their preponderance in this neighborhood or that.

On the other hand, the argument surely does not turn on the empirical claim that our world has more would-be Dutch Bookies than would-be Good Bookies. If you want to conduct a survey of just how mean the streets are out there, be my guest. For what it’s worth, the findings of experimental economists on the surprisingly ‘fair’ behavior of actual subjects in ‘dictator’ and ‘ultimatum’ games suggest to me that they may not be so mean after all; and recent work in comparative anthropology suggests that how mean they are will vary from one part of the world to another. But it’s worth little here, because of course the Dutch Book argument is intended as a piece of philosophy, not sociology. Indeed, it is premised on the highly idealized assumption of perfect rationality of the agent concerned. (No respect for the empirical facts there!) The imaginary Dutch Bookies, in their turn, are supposed to be perfectly rational, logically omniscient, and more besides: they must know the agent’s betting odds in order to exploit them, and they must be willing to do so. All the more dubious, then, is the apparently empirical claim that such characters predominate. But if this claim is another idealization (which is to say, a fiction), why should we accept it? In particular, why should we prefer it to the idealization (fiction) that Good Bookies predominate? Again, we seem to have a perfect symmetry here.

However, let us grant for the sake of the argument that Dutch Bookies prowl the agent’s ‘hood, and that Good Bookies are nowhere to be found. Still, it does not follow that the incoherent agent cannot enjoy the benefits of a Good Book. All he or she needs to do is spread the bets around: as it might be, selling a bet on \( A \lor B \) to Bookie 1, while buying bets on \( A \) and on \( B \) to Bookies 2 and 3. No single bookie, then, need be incoherent; no single bookie need be milked. The Dutch Book theorem did not assume that there actually are bookies at the other ends of the imagined transactions; still less did it assume that there is exactly one, involved with them all.

However, all this concedes far too much to the bad neighborhood defence; indeed, the whole tenor of the defence is quite misguided. It turns on certain contingent putative ‘facts’—rather dubious ones, as I have noted. But according to the Bayesian account of rationality, what matters are not the facts, but rather what the agent takes the facts to be. More precisely, what matters in determining what the agent should do are the agent’s own probability assignments. Your neighborhood could be bad without you realizing it—for example, because you have no evidence that it is. So the bad neighborhood defence needs to make an assumption to the effect that the agent assigns sufficiently high probability to his or her neighborhood being bad. When fully spelled out, the assumption will be
rather more complicated than this, involving the agent’s probability assignments to various hypotheses about just how bad the neighborhood is, and what the expected utilities will look like in each case. But let us not waste time trying to articulate this assumption, because it is all beside the point. Rationality does not require the agent to have particular opinions about any contingent matter, and that includes the local demographics. Coherence is not answerable to the world (the way that truth is), and coherence is all that the Dutch Book argument claimed to address.

I conclude that the standard Dutch Book argument fails. Like Route 66, it needs some repairing.5

5. Reconstructing/restoring the argument


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

Dutch Book theorem, revised:

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

Converse Dutch Book theorem, revised:

If you obey the probability calculus, there does not exist a set of bets, each of which you consider fair-or-favorable, that collectively guarantee your loss.
Good Book theorem, revised:

If you violate the probability calculus, there exists a set of bets, each of which you consider fair-or-favorable, that collectively guarantee your gain.

Converse Good Book theorem, revised:

If you obey the probability calculus, there does not exist a set of bets, each of which you consider fair-or-favorable, that 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 Good 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: Suppose you obey the probability calculus. 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.

But the revised version of the Converse Good Book theorem is not true: if you obey the probability calculus, there does exist a set of bets, each of which you consider fair-or-favorable, that collectively guarantee your gain. The proof is trivial. 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 gain. The revision from ‘fair’ to ‘fair-or-favorable’ makes all the difference. And with the failure of the revised version of the Converse Good Book theorem, there can be no revised version of the complete Good Book theorem, and thus no revised version of the Good Book argument.

There were no Good Books for a coherent agent, because Good Books were defined in terms of fair bets. But there are other profitable books besides Good 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. 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 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” (1980, 41). “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 sounds 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.

Ramsey’s formulation of the argument in terms of consistency has proved to be very influential. De Finetti also speaks of invulnerability to Dutch Books as a “condition of consistency” (1980, 213). The view has been forcefully advocated more recently by Skyrms, who writes: “… what is basic is the consistency condition that you evaluate a betting arrangement independently of how it is described (e.g., as a bet on \( p \) or \( q \) or as a system of bets consisting of a bet on \( p \) and a bet on \( q \))” (1984, 22). On this view, Dutch Books merely dramatize a deeper defect of an incoherent agent. Armendt (1993) endorses this interpretation and calls giving two or more different evaluations to the same option “divided mind inconsistency”.

Put this way, the argument has nothing especially to do with sure losses per se. The sure gains inherent in Good Books presumably dramatize the same defect. The only difference is that your sure losses are more dramatic than your sure gains; an incoherent state of opinion somehow seems more defective when it results in sure losses rather than in sure gains. But as far as the point about inconsistency goes, the Dutch Book and the Good Book are on a par. That point is free-standing, and it is no more ‘Dutch’ than ‘Good’.

Skyrms describes the Dutch Book theorem as “a striking corollary” of an underlying inconsistency inherent in violating the probability axioms (1984, 22).
But either the striking corollary does no real work, or it does real work: either the prospect of sure losses is not taken seriously, or it is. In the former case, the *argument* does not especially concern Dutch Books at all, and ‘Dutch Book argument’ does not fully capture the essence of the argument; indeed, ‘Good Book argument’ would be just as good a name for it (although for historical reasons I doubt that it will catch on!). In the latter case, the symmetric prospect of sure gains should be taken *exactly as seriously*. The Good Book theorem is another corollary—equally striking—of an underlying inconsistency. But then the prospect of sure losses loses its sting; it is neutralized by the symmetric prospect of sure gains. In any case, it is the original corollary that has received all the limelight in the literature, so much so that one could easily get the impression that it was the heart of the argument itself.

I am suggesting that there is a simple way to restore the full force of the corollary, and that Ramsey saw it all along. He did not live to finish “Truth and Probability”, and his version of the Dutch Book argument will forever remain telegraphic. But there is enough in what he *did* say to convince me that subsequent expositors did him a disservice in recasting the argument in terms of ‘fair’ bets. Once we return to Ramsey’s notion of “acceptable” bets, replacing ‘fair’ bets by ‘fair-or-favorable’ bets throughout the Dutch Book argument, there is no Good Book-style neutralization. An important route to Bayesianism is clear again.9

Notes

1. On a liberal understanding of ‘Dutch Book arguments’, the list would also include Skyrms (1980, 1984, 1986, 1987), Armendt (1992), and Lewis (1999). They down-play pragmatic talk of sure losses, bookies and so on, taking all of this to be a dramatization of a deeper defect of an incoherent agent: evaluating a single betting situation in conflicting ways, depending on how it is described. We will return to these authors in the final section, where I will be less liberal in my understanding.

2. Note that we should really make explicit the relativization *to you*: a Dutch Book is really a two-place *relation* that a set of bets bears (or not) to an agent, so there is no such thing as a Dutch Book *per se*. Nevertheless, I will sometimes follow the usual practice of gliding over this relativization, since the agent will be *you* in what follows.

3. Note that we should really make explicit the relativization *to you*: a Good Book is really a two-place *relation* that a set of bets bears (or not) to an agent, so there is no such thing as a Good Book *per se*. Nevertheless, I will sometimes glide over this relativization, since the agent will be *you* in what follows.

4. See footnote 1, and Section 5.

5. We could similarly parody, and offer similar diagnoses for, other Dutch Book arguments: for countable additivity (Adams 1962); for updating by conditionalization (Lewis 1999); for updating by Jeffrey conditionalization (Armendt 1992); and for the ‘Reflection Principle’ (van Fraassen 1984). We could even parody the
so-called ‘semi Dutch Book’ argument for strict coherence (Shimony 1955) with a ‘semi Good Book’ against it, with a similar diagnosis again. In this paper I am happy to confine myself to that part of Bayesianism that is most entrenched: the demand for synchronic conformity to the probability calculus, as presented in §2.

6. In commenting on an earlier version of this paper, Tom Cunningham and Brad Monton independently came up with versions of this formulation of the argument.

7. 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).

8. This interpretation also has its share of critics, including Maher (1993) and Vineberg (2001).

9. I am grateful to Erik Angner, Fiona Cowie, Tom Cunningham, Peter Gerdes, Matthias Hild, Andrea Scarantino, Neil Thomason, Brian Weatherson, Jim Woodward, Lyle Zynda, and especially Brad Armendt, Mark Colyvan, Adam Elga, Chris Hitchcock, Brad Monton, Paul Updike, Susan Vineberg, and James Worcester for very helpful discussion. Thanks also to audiences at Universidade Federale de Santa Caterina the Catholic University of Rio de Janeiro, Monash University, and Melbourne University. Some of this paper was written at the Center for Philosophy of Science at the University of Pittsburgh, which I thank for its financial and intellectual support. Special thanks to Branden Fitelson, who gave me very incisive feedback on several time-slices of this paper.

References


Skyrms, Brian (1984), Pragmatics and Empiricism, Yale University.


