In Memory of Richard Jeffrey:

Some Reminiscences, and Some Reflections on *The Logic of Decision**

Alan Hájek†

Philosophy of Science, PSA 2004 edition.

Abstract

This paper is partly a tribute to Richard Jeffrey, partly a reflection on some of his writings, *The Logic of Decision* in particular. I begin with a brief biography and some fond reminiscences of Dick. I turn to some of the key tenets of his version of Bayesianism. All of these tenets are deployed in my discussion of his response to the St. Petersburg paradox, a notorious problem for decision theory that involves a game of infinite expectation. Prompted by that paradox, I conclude with some suggestions of avenues for future research.

1. Introduction

Richard Jeffrey, known variously as “Dick”, “Diamond Jim”, “The Big Bopper”, and “Western Man” was a much beloved philosopher, teacher, and friend. Any number of people could appropriately be a part of this tribute – he had friends all over the world, “Comrades and Fellow Travellers in the Struggle for Bayesianism!” In any case, all of the participants in this symposium knew him well, valued his friendship, and were greatly influenced by his work.

Dick was born in Boston in 1926. He attended Boston University for two years and then went into the Navy towards the end of the war. He did his M.A. in Chicago, studying with Rudolf Carnap. He went to MIT to work on the logical design of computers in the early 50’s. During that time he met Edie Kelman (whom I thank for helping me put together these biographical notes). They married in 1955.

Dick went to Princeton in 1955 to do his Ph.D. with Carl Hempel. He finished in a startling two years. He went to Oxford for a year on a Fulbright fellowship. Then followed positions at MIT, Stanford, Princeton’s Institute of Advanced Study, City College of New York, and The University of Pennsylvania. He went to Princeton in 1974, and that remained his base for the rest of his life. More recently, he also spent part of his time as visiting distinguished professor of logic and philosophy of science at UC Irvine. He was elected to the American Academy of Arts and Sciences in 1994.

He wrote a number of classic books. How’s this for just part of a CV?: *The Logic of Decision, Formal Logic: Its Scope and Limits, Computability and Logic* (co-authored with George Boolos), *Probability and the Art of Judgment*, and *Subjective Probability: The Real Thing!*

The Logic of Decision was the third book that I ever read in philosophy, and it sure got me hooked. I attended every graduate seminar that Dick gave during my time at Princeton. When I received his comments on an early paper of mine, I almost thought he'd put them in the wrong mailbox, because they were addressed to "Wiggy Daddy". He famously did a guest spot singing with the Van Fraassens, a Princeton band in which country and western met decision theory. (He did a stirring rendition of "I'm Your Prisoner and You're My Dilemma".) He kept his sense of humor through his battle with cancer. Who else but Dick could speak during his arduous treatment of "the sensuous chemotherapy"?

He was the president of the PSA from 1999 to 2000. We'll never forget his presidential address, which was full of fascinating autobiographical and historical insights, and which was also side-splittingly funny—vintage Dick. Those who attended that talk may remember him saying with undue modesty that he had had "two good ideas" in his life. (That's a bit like Groucho Marx saying that he had had two good jokes in his life – Dick was a fan of the Marx Brothers, by the way.) Anyway, the two ideas he spoke about *were* indeed good: *the Logic of Decision*, which came to be known as 'evidential decision theory', and 'probability kinematics', which is now appropriately known as 'Jeffrey conditioning'. They were two keystones in a beautiful Bayesian edifice that Dick built, avowedly with help from friends—his "spiritual advisors"—and with the inspiration of "the sensuous Ramsey" and "the divine de Finetti".

The Bayesian House That Dick Built

Let me first say a bit more about the edifice. *The Logic of Decision* provides a framework for representing rational decision-making, in which an agent's values, represented as 'desirabilities' and opinions, represented as 'probabilities', work in tandem to determine what the agent should do. As Dick points out in a number of his works, this picture goes back to the *Port Royal Logic*. Noteworthy here, as he says on the very first page of *The Logic of Decision*, is that "[t]he numerical probabilities and desirabilities are meant to be subjective in the sense that they reflect the agent's actual beliefs and preferences, irrespective of factual or moral justification."

Regarding the probabilities, Dick favors an ecumenical, tolerant attitude that traces back to Ramsey and de Finetti: probability theory imposes consistency, of if you prefer, coherence constraints on opinions, but it does not legislate on which opinions you should have. (Compare: deductive logic imposes consistency constraints on sentences in a formal language, but it does not legislate on which sentences are true or false—and Dick often used this analogy to good effect. Indeed, he sometimes spoke of "probability logic".) Conformity to the probability axioms can be supported by Dutch Book arguments; so too, the diachronic updating constraints of conditioning and Jeffrey conditioning.

Regarding the desirabilities, Dick is also ecumenical: the Logic of Decision imposes certain constraints on preferences, such as transitivity and connectedness, but there is still a great deal of latitude. For example, it's not the job of decision theory to fault someone who prefers the destruction of the whole world to the scratching of their finger.

One of Dick's most profound ideas, which I would certainly have added to his far-too-short list of two, is his *non-foundationalist epistemology*. (It's related to Jeffrey conditioning, but it goes deeper.) Here, he opposes C. I. Lewis' idea that "[i]f anything is to be probable, then something must be certain" (1946, 186); Lewis found such certainties in the deliverances of the senses. For Dick, there need be no bedrock of certainties; epistemology need not embrace indubitable truths (apart from those found in logic or mathematics). This allows him to retain the spirit of the empiricists and logical positivists who inspired him, but without any commitment to dubious entities such as incorrigible sense data that can be encapsulated in observation sentences. As he memorably puts the point in his PSA Presidential Address (1998), it can be "probabilities all the way down to the roots". This is his "radical probabilism". It fits naturally with an idea promoted by his mentor Rudolf Carnap, that rational probability functions should be 'regular', or as I prefer to put it, 'open-minded'—they do not assign probability 0 to anything but logical contradictions, so that anything logically contingent gets assigned at least *some* probability. This adds a further constraint to the probability axioms. Still, the framework remains ecumenical, tolerant.

Dick's radical probabilism, in turn, goes hand-in-hand with Jeffrey conditioning, his proposal for how one's credences can be updated even when there is no proposition of which one becomes certain. Rather, one assigns probabilities across some partition which may be raised or dropped on the basis of sensory inputs, but not necessarily to 1 or 0 as strict conditioning would have it. Perhaps, as *David* Lewis insisted, an ideally rational agent can always formulate the contents of its experience as a single proposition that is fully learned. But this surely need not be the case for us humans—our episodes of learning need not be so decisive. So on this interpretation, Jeffrey conditioning is a step towards humanizing Bayesianism.

This brings me to another important strand in Dick's thought: his project of providing what he called "Bayesianism With a Human Face". Another respect in which we humans fall short of the orthodox Bayesian ideal is that our opinions are typically *imprecise*, whereas the orthodox Bayesian's opinions are always *sharp*, represented by a single probability function. Dick (and others, including Isaac Levi and Bas van Fraassen) did much to illuminate what imprecise opinion looks like, and how it should evolve. In "Bayesianism With a Human Face", Dick represents an imprecise state of opinion with a *set* of probability functions, what he calls a *probasition*.

There's a very nice completing of the circle here. For it turns out that the probability representation of a given agent that falls out of her preference ranking, as per the Logic of Decision, is unique iff her desirabilities are unbounded both above and below. If her desirabilities are bounded on either or both sides, then there are infinitely many equally good candidates for her probability function; in short, her probabilities according to that representation are indeterminate.

So we have several keystones of the edifice in place. On the side of opinion:

- radical probabilism (or "probabilities all the way down to the roots");
- rationality imposes the constraints of the probability calculus on credences, which can be thought of as a 'probability logic', and on the updating of credences, but it is otherwise tolerant;
- tolerance of imprecise probability assignments as part of Bayesianism with a Human Face.

On the side of desirability:

- the Logic of Decision, with
- its ecumenical stance on rational preferences, and
- its indeterminacy regarding probabilities when desirabilities are bounded on at least once side.

I turn now to Dick's treatment of the St. Petersburg paradox, where we will revisit all of these keystones.

Infinite Desirability: the St. Petersburg Paradox

Dick is very clear in his writings that his decision theory could not accommodate propositions of infinite desirability. Such propositions would yield some bad headaches—infinitely bad for the Logic of Decision, you might say, for they would apparently *destroy* rational decision-making. As he points out, if someone attributes infinite desirability to some prospect, then a probability of 0.99 of that prospect has the same desirability as a probability of 0.01 of that prospect, namely, infinite. Consider the notorious St. Petersburg game, to which Dick devotes a section of *The Logic of Decision*. We toss a fair coin until it lands heads for the first time. The longer it takes, the better for you. You receive exponentially escalating pay-offs according to the following schedule:

Coin lands heads for the first time on:	Probability	Pay-off
1st toss	1/2	\$2
2nd toss	1/4	\$4
3rd toss	1/8	\$8
...	...	...
nth toss	$1/2^n$	$\$2^n$

... ...

$$\text{Your expectation (in dollars)} = (1/2 \times 2) + (1/4 \times 4) + (1/8 \times 8) + \ldots$$

$$= 1 + 1 + 1 + \ldots$$

$$= \infty.$$

Decision theory would apparently tell you that you should be prepared to pay any finite amount to play this game once. This seems absurd—and thus we have the St. Petersburg paradox.

Once we let St. Petersburg games in the door, all hell breaks loose. Consider the Moscow game: all payoffs are raised by a dollar—or a ruble, if you prefer. (They're slightly richer in Moscow!) This is surely preferable to the St. Petersburg game, but decision theory adjudicates them the same. Surely doubling all the payoffs in the St. Petersburg game should improve it; but decision theory cannot say that. Surely playing the St. Petersburg game at no entry fee *twice* is better than playing it once; but decision theory cannot say that either.

Dick's answer to the St. Petersburg paradox begins unforgettably:

> Put briefly and crudely, our rebuttal of the St. Petersburg paradox consists in the remark that anyone who offers to let the agent play the St. Petersburg game is a liar, for he is pretending to have an indefinitely large bank... there is not that much money in the world. (154-5)

He is avowedly putting the objection quickly here, but it is surely a natural objection, pithily stated. He goes on to develop the point further. He imagines the Treasury department specially printing bills to honor the St. Petersburg contract, and he concludes:

> Due to the resulting inflation, the marginal desirabilities of such high payoffs would presumably be low enough to make the prospect of playing the game have finite expected desirability. (155)

Economists sometimes put the point in terms of ‘risk aversion’: your utility function is concave, indeed so much so that the expected desirability of the game converges. The upshot is that:

> the prospect of playing the St. Petersburg game ... is either illusory... or has finite expected desirability … Moreover, similar remarks apply to versions of the game in which the payoffs are things that money cannot buy. (155)

And he goes on to consider, for example, the agent remaining alive and in good health for 2^n years, for each n = 1, 2, … His conclusion is that *whatever the prizes might be*, St. Petersburg-like games with infinite desirability cannot arise.

Dick’s solution is subtle, and I would like to develop it further. Let’s proceed in three stages—the solution is really in the final stage, but the first two stages are instructive hints as to how to get there. The Logic of Decision, remember, involves considerations both on the side of opinion and on the side of desirability. Let’s take each in turn, and then put them together at the end.

Dick’s Response to the St. Petersburg Paradox in Three Steps

1. Probability

Of course, it’s a fact that nobody can make good on the St. Petersburg offer: “anyone who offers to let the agent play the St. Petersburg game is a liar”. But it is built into Dick’s system that what matters to rational decisions are not the facts, but rather a given agent's opinions about the facts, and if the paradox is to be genuinely resolved, it has to be resolved for all possible rational agents. So even if *in fact* nobody has an indefinitely large bank, the paradox still has teeth if some rational agent believes that somebody does have such a bank.

Indeed, they don't even need to believe it. For the paradox remains even if any rational agent gives a *tiny probability* to the prospect of playing the St. Petersburg game. Suppose I offer you the St. Petersburg game. You don't believe me; in fact you assign probability one-in-a-trillion to the offer being genuine. Still, the paradox has a hold on you: for now the expectation of the game is a trillionth of infinity, which is still infinity.

Well, maybe a trillionth is still too much credence to give the offer—you really must give it probability 0! *But this would no longer be a completely non-foundationalist epistemology.* We would have found a bedrock of certainty: 'there are no St. Petersburg games'. And there would be other bedrocks: 'there are no Moscow games', 'there are no doubled St. Petersburg games', and so on. The demand for such bedrocks also goes beyond the ecumenical Bayesianism that recognizes only the probability axioms as constraints on rational opinions—no probability axioms say, or imply, "Zero out the St. Petersburg game and its kin!". The indubitable truths may not be the deliverances of the senses, but indubitable they would be. On this view, C. I. Lewis was on the right track after all; he just chose the wrong examples.

This purely probabilistic banishing of St. Petersburg games also conflicts with regularity, or 'open-mindedness', since apparently rationality requires you to zero out various logically contingent possibilities—for instance, that an offer of the St. Petersburg game is genuine. Carnap would not be pleased.

But we are not done yet with Dick's solution. These considerations have come from the side of probability. How about desirability? Here Dick's point about inflation, and more generally risk aversion regarding any good, is apposite. So maybe considerations of desirability can by themselves rule out the St. Petersburg game? Let's see.

2. Desirability

A popular response to the St. Petersburg paradox is to insist that all utility functions are *bounded*. (See Martin 2004 for discussion.) Bernoulli noted that risk aversion could bring the St. Petersburg game under control—for example, if the utility of money goes not by its face value, but rather by the logarithm of that value. But Menger pointed out that as long as the utility function is unbounded, one can still generate the St. Petersburg game by appropriately stepping up the prizes—say, by having them grow not just exponentially, but super-exponentially, so that even after taking logs, the expectation is infinite.

This problem would not arise if all desirability functions were bounded—not just in practice, but in principle. And Dick does say at one point regarding the case of the Treasury department backing the St. Petersburg game: “in that case there would clearly be a finite upper bound on the desirabilities of the possible payoffs” (155). That way, desirability considerations alone would block St. Petersburg games. We need not worry about the probabilities, for whatever they might be, the expectation will converge to a value no greater than the bound.

But recall the extent to which preferences determine probabilities in the *Logic of Decision’s* representation: whenever an agent’s desirabilities are bounded, her corresponding probabilities are *indeterminate*. So if *all* utility functions are bounded, *all* agents come out as having indeterminate probabilities according to the representation. Then it seems that the theory of indeterminate opinion is not especially Bayesianism with a *human* face. Of necessity, all agents are indeterminate at all times. There *could not be* a sharp Bayesian agent, even in principle, according to the representation! We surely don't want this result, either. And it would be implausible to insist that rationality *requires* indeterminacy in one's opinions.

3. Probability and Desirability

Perhaps, then, we should make probability *and* desirability work in tandem to ensure that all desirabilities come out finite—that's surely the right way to go, and I think that it's really what Dick has in mind. They could work to ensure, for example, that necessarily inflation puts a rein on the desirabilities of ever-increasing monetary amounts (risk aversion), *and* that the corresponding probabilities decay sufficiently quickly to offset the corresponding growth in desirabilities. Similar remarks would presumably apply to things that money cannot buy.

Now you might worry that this is not so ecumenical. You are *constrained* not to assign certain probability distributions over propositions displaying St. Petersburg-like growth in utilities. This constraint did not come from the probability calculus, for the outlawed distributions may very well be non-negative, additive, and sum to 1 in the required way. And it seems to be a rather odd constraint given that the *very same distribution* across *the very same outcomes* would be quite permissible if the corresponding utilities did not grow so fast—if, for example, the house were less generous in its payoff scheme, and the payoffs escalated only linearly rather than exponentially. It is also odd that your opinions about *coin-tossing* should be sensitive to something as extraneous as the prizes that you might or might not get, as if you always have to keep an eye on what payoffs are in the offing before you can assign probabilities to the outcomes of a coin-tossing experiment. Moreover, there is the threat that in doing so you will violate an epistemic norm on opinion. Lewis's *Principal Principle* states roughly that your credences should coincide with your expectation of the corresponding *objective chances*. Yet in the coin-tossing experiment, the chances presumably decay exponentially, and it is surely permissible and perhaps even obligatory for you to believe this. Your credences, then, must follow suit. Your problem appears to be *overconstrained:* the worry is that you cannot obey both the epistemic norm enshrined in the Principal Principle and the practical norms enshrined in decision theory when St. Petersburg payoff schedules are lurking.

One might think that there is still a loophole. After all, strictly speaking it's not the propositions about *coin-tossing* whose probabilities are constrained to decay so fast. Rather, it's the propositions about your *payoffs* whose probabilities are so constrained. As the number of tosses before the first heads increases, you may become increasingly skeptical that the bank will really pay according to the St. Petersburg contract. You may, for example, assign probability 2^{-100} to the first heads occurring on toss 100, but a far smaller probability to this occurring *and the bank paying $* 2^{100}.

Well, you may. Or you may not. After all, nothing in the probability calculus forbids you from assigning probability 1 to the conjunction of biconditionals that bind the coin-tossing outcomes to the payoffs in the St. Petersburg contract. Furthermore, trying to pursue this loophole puts strain on Dick's analogizing of probability theory to deductive logic, for *payoffs* play no role in determining whether an argument is valid or whether a set of sentences is consistent.

Of course, one could look at the constraint that reins in the St. Petersburg expectation in the other way, namely, as one on the desirabilities: supposing the probability distribution to be given, the desirabilities must not grow too fast. But then the worry just goes the other way from before: now it seems odd that the very same desirability profile across the very same outcomes would be quite permissible if the probabilities decayed faster. Now you have to keep an eye on what probabilities are in the offing before you can decide how much you value various payoffs. That seems odd in itself, and in any case it is hardly ecumenical.

Really, it is best to think of it as a *joint* constraint on desirabilities and probabilities.[1] You have preferences over propositions. There can be no proposition that corresponds to the St. Petersburg game *and* that you assign positive probability. If the payoffs of the game described by the proposition grow St. Petersburg-style, then you assign the proposition probability 0; if you assign the proposition positive probability, then the payoffs don't really grow St. Petersburg-style, but rather in such a way that the expectation is finite.

I take this to be the resolution of the St. Petersburg game that Dick has in mind, albeit spelled out a bit more. Still, you might protest: we have gone beyond the rationality, or 'coherence' constraints imposed by the probability calculus. Of course we have. But we have not gone beyond the constraints imposed by the Logic of Decision's preference axioms.

The Preference Axioms, and the Representation Theorem

It is a consequence of those axioms that we never get an infinite utility assigned to any prospect of positive probability. Insofar as the axioms are taken to be requirements of rationality, they guarantee that St. Petersburg problems do not arise for rational agents. A proper discussion of those axioms must await another occasion; for now, I will confine myself to some quick remarks.

Some of the preference axioms are pretty clearly demands of rationality—transitivity, for example. But completeness? Atomlessness? Continuity? These are apparently not as compelling as 'rationality' conditions as are the axioms of probability. Still less, it would seem, are they 'consistency' conditions as we find in logic. Rather, they seem to be structural axioms of 'richness' that ensure that the resulting representation is real-valued and well-behaved in various ways.

But one might insist that they are more than that. Perhaps the violation of these axioms does lead to an incoherence of sorts—indeed, that's exactly what the St. Petersburg paradox shows! For it certainly seems irrational, and perhaps even incoherent, for you to be *indifferent* between a prize that you value as highly as the St. Petersburg game, and a one-in-a-trillion probability of that prize.[2] The house could repeatedly lower the probability of your getting the prize, without any complaint from you. This is not quite a Dutch Book argument, but it is somewhat in the same spirit, and one might insist that it displays an incoherence in your valuations, much as Dutch Book arguments are supposed to. Spelled out a little more formally: while the house cannot *guarantee* your loss, as is the case with a Dutch Book, it can drive the probability of your loss, by your own lights, above any threshold—without any expected return for you. You will readily pay a million dollars to play the game, *and* just as readily pay a million dollars for an ε probability of playing the game, and a 1– ε probability of losing that money, for any positive ε, however small. We might call this a *1– ε Dutch Book*.

Interestingly, Dick himself does not place much emphasis on the preference axioms, nor on the resulting representation theorem. The axioms do not appear until rather late in *The Logic of Decision* (Ch. 9), and he doesn't spend much time motivating or defending them. In fact, if you look carefully, he never actually *states* the representation theorem (although, to be sure, it's easy enough to figure out what it must be from the things that he *does* state). There is not the sort of fanfare that you find in, say, Savage's *Foundations of Statistics* regarding *his* representation theorem.

I think that Dick is comparatively low-key about the preference axioms and the representation theorem because for him, desirabilities and probabilities are more basic. Again, the picture goes back to the *Port Royal Logic*. As he says in the first sentence of his final book, "probability is a mode of judgment". Various economists might regard facts about an agent's preferences to be basic, and the probabilities that fall out of the representation theorem to be epiphenomenal. But for Dick, after all, *Subjective Probability is The Real Thing!*

Future Developments?

Let me end by just gesturing at two other approaches that I think are worth exploring further.

The first approach is to adopt a non-Archimedean representation of desirability. We allow infinite desirabilities, but in a more nuanced manner, so that, for example, we can tell the difference between a game of infinite desirability, and a one-in-a-trillionth probability of such a game. (Compare how set theorists have developed formal devices for distinguishing the cardinalities of various infinite sets.[3]) Elsewhere (2003) I have suggested that non-standard analysis, and Conway's 'surreal numbers', may provide frameworks for representing such desirabilities.

The second approach is one that I hope to pursue with Michael Quinn. I can only sketch it here, and it is admittedly only a first step. It is inspired by a remark of Feynman's regarding the handling of infinities in physics: while certain quantities may explode when taken individually, their *differences* may behave more tractably. Now at the end of the day, what matters to decision theory is a ranking of the alternatives, and perhaps a measure of the differences in their desirabilities. So the idea is to replace matrices of desirabilities, à la the Logic of Decision, with matrices of *differences* in desirabilities. Multiply the difference matrices by the corresponding probability matrix, to yield an expected difference matrix. Positive entries in this matrix correspond to expected gains in switching from one action to another; negative entries correspond to expected losses. You should then perform an action for which there are no expected losses compared to other actions. For example, in every state of the world, the *difference* in utilities between the Moscow game and the St. Petersburg game is $1. The expected difference is thus $1, which is positive. This requires you to prefer the Moscow game, as intuition demands. We hope in this way to unify decision theory with dominance reasoning more generally.

We believe that expected difference theory also delivers the right verdicts in some cases where both decision theory and dominance reasoning go silent, but intuition speaks, and even shouts. Such a case was our preference for playing the St. Petersburg game twice, rather than once (at no charge). Now, I do not pretend that expected difference theory will cure all of the headaches associated with infinite utilities. But it may be good medicine until a more definitive cure is found. I invite you to join the search!

So, much as we should be thankful to Dick for carrying his research program as far as he did, I think that it is also fertile ground for future developments. I think that that's exactly as he would have wanted it.[4]

Research School of the Social Sciences,

Australian National University

alanh@coombs.anu.edu.au

REFERENCES

Hájek, Alan (2003), “Waging War on Pascal’s Wager”, *Philosophical Review* 113 (January): 27–56.

Jeffrey, Richard C. (1966), *The Logic of Decision*. Chicago: University of Chicago Press; 2nd ed. 1983.

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

Jeffrey, Richard (1998), “I Was a Teenage Logical Positivist (Now a Septuagenarian Radical Probabilist)”, PSA Presidential Address, notes available at: http://66.102.7.104/search?q=cache:5Fc13EjFsHMJ:www.princeton.edu/~bayesway/KC.tex.pdf+%22probabilities+all+the+way+down%22+Jeffrey&hl=en.

Lewis, C. I. (1946), *An Analysis of Knowledge and Valuation*. La Salle, Illinois: Open Court.

Martin, Robert (2004), "The St. Petersburg Paradox", *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2004/entries/paradox-stpetersburg/>

Savage, Leonard J., (1954), *The Foundations of Statistics*. New York: John Wiley.

FOOTNOTES

[1] I thank Jim Joyce for putting the point this way.

[2] I thank Daniel Nolan for suggesting a version of this point to me.

[3] Peter Vranas suggested this analogy.

[4] I thank Adam Elga, Matthias Hild, Peter Vranas, and especially Jim Joyce and Daniel Nolan for very helpful feedback.