The Foundations of Causal Decision Theory, by James Joyce, Cambridge: Cambridge University Press, 1999, Pp. xii + 268. H/b £45.00.

This book is a welcome addition to many distinct branches of the recent philosophical literature, including rational decision making and probabilistic theories of evidence and confirmation. The book is well organized and well written and should be accessible both to students and researchers with interests in these issues. Joyce begins with an entertaining but trenchant historical introduction to the topic of decision theory (including an incisive critical discussion of Leonard Savage's theory). Within just a few brief chapters, Joyce has the reader well prepared for a thorough treatment of the modern 'evidential' theory of decision making due to Richard Jeffrey. This is followed by a chapter on causal decision theory, and a chapter on conditional beliefs and probabilistic accounts of evidence and confirmation (including a novel approach to the 'problem of old evidence', which will be discussed critically below). The final chapter contains the book's most important contribution to the literature: a unified representation theorem that simultaneously provides a firm foundation for both evidential and causal decision theory.

The accounts of rational decision discussed by the author all presuppose that a rational agent should act so as to maximize some sort of 'expected utility', which is taken to be some sort of weighted average of the utilities (or values) of the outcomes of a decision or an act. What's at issue in the foundational disputes discussed in the book is the matter of which kind of expected utility should be maximized, and this boils down to a dispute about which weights should be used in the weighted average of the values of the outcomes of the decision or act being considered (nobody in this field seems to argue about whether the values or utilities are 'correct'). Moreover, all parties seem to agree that the weights should be set according to 'the probabilities of the outcomes given that the act is performed'. The disagreement concerns how to best unpack this subtle conditional-like expression for the purpose at hand.

Evidential decision theorists (e.g., Jeffrey and Eells) unpack 'the probability of the outcome O given that the act A is performed' as P(O/A), which is an agent's degree of belief (or degree of credence) in O on the supposition that A is true (i.e., on the supposition that A is performed — here, acts, outcomes, and states are all interpreted as propositions). The 'supposition that A' involved in P(O/A) is to be understood as a matter-of-fact supposition, as opposed to a subjunctive supposition (see below). The evidential decision theorist then defines the expected utility of an act A as the following weighted average: EU(A) = $\sum$ P(O/A) $\cdot$ u(O $\cap$ A), where u is the agent's utility (or value) function. This quantity is sometimes called the 'news value' of the act A, and it is sometimes said that evidential decision theory recommends performing that act which provides the 'best evidence for the good outcomes (on average)'. On the other hand, causal decision theorists (e.g., Stalnaker and Lewis) propose a different way of unpacking 'the probability of the outcome O given that the act A is performed'. They suggest that we unpack this (roughly) as the 'degree to which (the agent judges it to be the case that) A causally promotes O'. In Joyce's notation, this 'causal (subjective) conditional probability' – which Joyce understands as the degree of credence in O on the subjunctive supposition that A (as opposed to the matter-of-fact supposition that A in P(O/A) above) – is written as P(O\A). The formula for 'causal expected utility' is the same as before, except that we toggle the '/' to a '\' in the probability weights, so as to indicate this switch from evidential to causal (subjective) conditional probability. It is sometimes said that causal decision theory recommends performing that act which is judged (by the agent in question) to be 'most likely to causally promote the outcome O (on average)'. Many interpretations of P(O\A) have been offered in the literature,

and the connections between P(O\A) and various kinds of conditionals have been studied extensively in recent decades. In chapters 5 and 6, Joyce provides a very nice discussion of many of these proposals, and he offers some general and insightful accounts of his own along the way. Some of the book's deepest lessons appear in these chapters.

Obviously, when P(O/A) = P(O\A), as will sometimes be the case, the evidential and causal theories of decision will generate exactly the same prescriptions. However, it is well known that evidential and causal conditional probability judgments can differ in important ways. For instance, if O and A are joint effects of a common cause, then A can have very high evidential 'news value' regarding O (i.e., P(O/A) can be high, and O and A can be strongly correlated under P), even though it is known that A has little or no causal efficacy for bringing O about (i.e., P(O\A) is low). In such cases (most famously, in Newcomb-type problems), evidential and causal decision theory can (at least prima facie) come apart and offer different advice about which actions should be performed. I will not discuss further these most interesting and controversial cases and how they bear on Joyce's discussion concerning causal vs. evidential decision theory. This aspect of Joyce's treatise has been reviewed expertly elsewhere (see Ellery Eells' review in the British Journal for the Philosophy of Science **51** (2000), 893–900).

The main contribution of the book occurs in the final chapter. There, Joyce presents an elegant, unified (simultaneous) representation theorem for both evidential and causal decision theory. A representation theorem is a mathematical result which shows that a rational agent's preferences (assumed on pain of irrationality to satisfy certain conditions like transitivity and normality) can be faithfully captured by inequalities among the agent's expected utilities (according to the formula for expected utility presupposed by the decision theory in question). To be more precise, a representation theorem establishes existence and uniqueness of probability and utility (i.e., value, desirability) functions such that a rational agent's preferences match-up with their expected utility comparisons (again, according to the formula for expected utility presupposed by the decision theory in question). I will not get into the technical details of Joyce's representation theorem. What's important about it is that it generalizes Ethan Bolker's representation theorem for Richard Jeffrey's evidential decision theory in several crucial ways. Most importantly, Joyce's theorem provides a (simultaneous) representation of both causal and evidential decision theory. No truly general and satisfactory representation theorem (along the lines of Bolker's evidential representation theorem) had been proven for causal decision theory. Not only does Joyce achieve this goal, but he does so in a way which clarifies the precise relationships between causal and evidential decision theory and their theoretical foundations. Joyce succeeds in constructing a very general analytical framework within which both the evidential and causal theories can be couched, and this is the key to his successful presentation of a unified theoretical and philosophical foundation for decision theory.

For the remainder of this brief review, I would like to make some critical remarks about Joyce's novel and interesting analysis (in chapter 6) of the so-called 'problem of old evidence' in Bayesian confirmation theory. According to Bayesian confirmation theory, evidence E confirms (or supports) a hypothesis H (for a rational Bayesian agent with an evidential probability function P(•/•)) just in case P(H/E) > P(H/□E). That is, E confirms H iff E and H are positively correlated under the agent's evidential probability function P. On the standard (Kolmogorov) definition of probability functions, there are many logically equivalent ways of saying that E and H are positively correlated under P. Here are three such ways ('**T**' denotes a logical truth):

(1) P(H/E) > P(H/□E)

(2) $P(H/E) > P(H/\mathbf{T})$
(3) $P(E/H) > P(E/\neg H)$

According to 'orthodox Bayesianism,' once an observation of E's truth is made by an agent, E becomes certain for that agent. That is, evidence is assumed to have probability 1 once it is learned. However, as Clark Glymour was the first to clearly point out, this has a strange consequence for Bayesian confirmation theory: once *E* is learned, it can no longer provide evidence for any hypothesis. This is because, if $P(E) = P(E/\mathbf{T}) = 1$, then none of the inequalities (1)–(3) is true. In particular, if $P(E) = P(E/\mathbf{T}) = 1$, then $P(H/E) = P(H/\mathbf{T}) = P(H)$, and E and H are not positively correlated under P. This problem has been called 'the problem of old evidence'. Many proposals have been made in recent years concerning possible solutions to this problem (see chapter 5 of John Earman's (1992) book Bayes or Bust? Cambridge, MA: MIT Press for a nice survey). In chapter 6, Joyce offers a novel approach to this problem. While it is both clever and new, I think Joyce's treatment of this matter in chapter 6 has several potential flaws and shortcomings. Getting clear on this issue is of fundamental importance for the foundations of Bayesian confirmation theory.

Joyce suggests that we should use what he calls 'Réyni-Popper measures,' rather than the more traditional Kolmogorov probability functions, as our formal theory of evidential conditional probabilities $P(X/Y)$. According to the Réyni-Popper theory, $P(X/Y)$ is well-defined even when $P(Y/\mathbf{T}) = 0$. But, according to Kolmogorov's definition, $P(X/Y)$ is undefined when $P(Y/\mathbf{T}) = 0$. (Note: If $P(Y/\mathbf{T}) > 0$, then the Kolmogorov and the Réyni-Popper theory of $P(X/Y)$ are in complete agreement. See Roeper and Leblanc's (1999) book Probability Theory and Probability Logic, Toronto: University of Toronto Press for an encyclopedic discussion of alternative formal theories of conditional probability.) So, for instance, if $P(E/\mathbf{T}) = 1$, as in the orthodox Bayesian theory of evidential learning, then $P(H/\neg E)$ is undefined on the Kolmogorov definition, but it remains well-defined on the Réyni-Popper definition. Next, Joyce suggests that Bayesian confirmation theorists interested in escaping the problem of old evidence should consider the following quantitative measure: $s(H, E) = P(H/E) - P(H/\neg E)$. This measure has also been recently proposed for similar purposes by David Christensen (see his paper 'Measuring Confirmation,' Journal of Philosophy (1999) **XCVI**: 437–461). Like Christensen, Joyce contrasts s with the more traditional Bayesian measure of confirmation: $d(H, E) = P(H/E) - P(H/\mathbf{T})$. However, unlike Christensen, Joyce does not recommend replacing d with s. Joyce (p. 206) thinks that d and s are both adequate measures; but they are measures of different quantities. Joyce's view seems to be that s measures a 'component' of the total (or 'resultant') incremental confirmation (which he thinks may be adequately measured using d) that is invariant as $P(E/\mathbf{T})$ changes via learning. His suggestion is that by focusing on the 'support' quantity measured by s, we can gain insight into (and some relief from) the problem of old evidence. Joyce's claim that d and s are both adequate measures (of different, confirmationally salient quantities) is somewhat controversial in the context of Kolmogorov-based confirmation theory. Most authors in the Kolmogorov-Bayesian confirmation theoretic tradition seem to suppose that there is 'one true measure of confirmation', and that s and d (which are not ordinally equivalent) cannot both be 'adequate measures of confirmationally salient quantities' (see, for instance, Milne, P. 'log[p(h/eb)/p(h/b)] is the One True Measure of Confirmation', Philosophy of Science (1996) **63**: 21–26, and Fitelson, B. 'The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity,' Philosophy of Science (1999) **66**: S362–S378). But, as we will see below, in the context of the Réyni-Popper theory of conditional probability, it is inevitable

(and uncontroversial) that d and s will measure not only different, confirmationally salient quantities, but different kinds of (qualitative) Bayesian confirmation altogether.

Joyce correctly points out that, while the traditional measure of confirmation d says that d(H,E) = 0, if P(E/**T**) = 1 (on either the Kolmogorov or the Réyni-Popper theory of P(X/Y), see below), the measure s does not have this consequence. On the Kolmogorov theory of P(X/Y), s(H,E) is undefined if P(E/**T**) = 1. But, on a Réyni-Popper theory of P(X/Y), s(H,E) will generally be well-defined even if P(E/**T**) = 1. Moreover, s(H,E) need not be zero in such cases, and it can in fact be arbitrarily large (i.e., s(H,E) can be arbitrarily close to its maximum value of 1; Christensen op cit makes similar points). So, Joyce shows that the combination of (a) adopting a Réyni-Popper theory of P(X/Y), and (b) focusing on the quantity measured by s(H, E) = P(H/E) – P(H/□E) allows us to explain how evidence which has been learned with certainty can still provide (non-zero, or even strong) evidential support (in some sense). This is how Joyce finds his way out of Glymour's conundrum.

There are several problems with Joyce's discussion of this problem in chapter 6. First, Joyce seems to suggest (p. 206, note 39) that the measures d and s will only disagree on quantitative judgments of degree confirmation, but that they will agree on qualitative judgments as to whether or not E confirms, disconfirms, or is irrelevant to H. This is incorrect. In fact, if one adopts a Réyni-Popper theory of P(X/Y), then one will be faced with many, non-equivalent qualitative accounts of Bayesian confirmation. This is easily seen in the case at hand. Consider a case in which P(E/**T**) = 1, as in the old evidence problem. In such a case, we could have s(H, E) > 0. Indeed, this very possibility is what makes Joyce's proposal tick. But, in such a case, we will always have d(H, E) = 0. So, according to d, E will be confirmationally irrelevant to H, but according to s, E will confirm (or support) H (these considerations also make it unclear how s could possibly measure a 'component' of the quantity measured by d). This is a qualitative disagreement about 'what confirms what,' not merely a quantitative disagreement about 'what confirms what more strongly than what.' On the other hand, from the Kolmogorov perspective, d and s are guaranteed to agree on qualitative judgments (although, they will have some rather serious quantitative and comparative disagreements — see Eells and Fitelson (2000) 'Measuring Confirmation and Evidence,' Journal of Philosophy **XCVII**: 663-672 and Eells and Fitelson (2002) 'Symmetries and Asymmetries in Evidential Support' Philosophical Studies **107**: 129-142 for discussion). Thus, from the point of view of Kolmogorov-based Bayesian confirmation theory, this move to Réyni-Popper theory has a serious ambiguating and/or disunifying effect. The fact that moving to a Réyni-Popper theory of P(X/Y) leads to a non-trivial ambiguation/disunification of the (traditional, Kolmogorovian) qualitative Bayesian concept of confirmation has not been widely noted in the literature (this ambiguation/disunification is implicitly noted in Roberto Festa's (1999) paper 'Bayesian Confirmation' which appears in the volume Experience, Reality, and Scientific Explanation, M. Galavotti and A. Pagnini (eds.), Dordrecht: Kluwer Academic Publishers, pp. 55–87). Bayesian confirmation theorists in the market for a resolution to the old evidence problem should consider the cost of the qualitative disunification associated with part (a) of Joyce's proposal.

Moreover, when one realizes that there are many qualitative senses of 'confirmation' on a Réyni-Popper-Bayes theory, one sees that Joyce has — at best — provided a resolution to only one of many distinct, qualitative problems of old evidence that arise in this new framework. On the Réyni-Popper-Bayes approach, there will now be d-confirmation, s-confirmation, and many other kinds of qualitative confirmation as well. Joyce only addresses the problem of old evidence in connection with s-confirmation. But, the problem is alive and well in the context of

d-confirmation, since d(H, E) = 0 whenever E is known with certainty (on any theory of conditional probability). So, on Joyce's proposal, many new problems of old evidence — problems created by the very move to Réyni-Popper functions — will remain unsolved (e.g., l-confirmation or likelihood-type qualitative confirmation based on inequality (3) will also still suffer from its own problem of old evidence, untouched by Joyce's proposed resolution). Perhaps the most accurate way to characterize the situation is as follows. In the orthodox, Kolmogorov-Bayes framework, there is only one kind of (qualitative) confirmation, and only one problem of old evidence to go with it. In the Réyni-Popper-Bayes framework, there are many kinds of (qualitative) confirmation, and many (potential) problems of old evidence. What Joyce has correctly pointed out is that, in the Réyni-Popper-Bayes framework, s-confirmation does not suffer from a problem of old evidence. But, this leaves the problems of old d-evidence, old l-evidence (and other kinds of old evidence besides) unresolved in the Réyni-Popper-Bayes framework. It appears that we will need a further story about how to resolve these persistent problems of old evidence in the Réyni-Popper-Bayes framework (see Eells and Fitelson (2000) op cit for one possible line of attack here).

Part (b) of Joyce's proposed resolution of the old evidence problem also has some potential drawbacks. The measure s Joyce (and Christensen op cit) considers has many properties that Bayesian confirmation theorists (and others interested in quantitative evidential support) will find undesirable. Due to space limitations, I will not rehearse these odd properties of Joyce's measure s here. See Eells and Fitelson op cit for a discussion concerning several potentially problematic properties of s in the context of the Kolmogorov theory of probability (many of the points about s made by Eells and Fitelson will carry over to the Réyni-Popper theory of probability as well). The considerations raised by Eells and Fitelson seem to cast some doubt even on Joyce's relatively weak claim that s is an adequate measure of some 'confirmationally salient quantity' or some 'kind of confirmation' (especially if the quantity that s measures is thought to be a 'component' of 'Bayesian incremental confirmation' as the term is normally used, or as a 'component' of 'the quantity adequately measured by d').

Although I have made some brief critical remarks concerning Joyce's novel proposed resolution of the problem of old evidence in Bayesian confirmation theory, this should not lead the reader toward any negative conclusions about this book. On the contrary, I think Joyce's book is an edifying must read for anyone interested in rational decision making, conditional belief (or conditional logic), probability, evidence, causality, or a myriad of other related issues.

I would like to thank James Joyce for providing useful feedback on an previous version of this review, and Alan Hájek for numerous edifying discussions about probability over the past several years.

Department of Philosophy
San José State University
San José, CA 95192-0096
USA

BRANDEN FITELSON