Richard Bradley, Decision Theory with a Human Face

Nicolas Gravel

Comptes rendus / *Reviews*

Richard Bradley, *Decision Theory with a Human Face*

Cambridge: Cambridge University Press, 2017, 335 pages, ISBN 978-110700321-7

Nicolas Gravel*

The very title of this book, borrowed from Richard Jeffrey's "Bayesianism with a human face" (Jeffrey, 1983a) is a clear indication of its content. Just like its spiritual cousin *The Foundations of Causal Decision Theory* by James M. Joyce (2000), *Decision Theory with a Human Face* provides a thoughtful description of the current state of development of the Bolker-Jeffrey (BJ) approach to decision-making. While a full-fledged presentation of the BJ approach is beyond the scope of this review, it is difficult to appraise the content of *Decision Theory with a Human Face* without some acquaintance with both the basics of the BJ approach to decision-making and its fitting in the large corpus of "conventional" decision theories that have developed in economics, mathematics and psychology since at least the publication of Von von Neumann and Morgenstern (1947).

* Centre de Sciences Humaines, Delhi and Aix-Marseille University. nicolas.gravel@csh-delhi.com.

This review has benefited from Jean Bacelli's excellent remarks and suggestions. Of course the usual disclaimer applies fully.

Conventional decision theories analyze how agents (individuals, firms, governments, etc.) make decisions. The precise definition of what a decision is varies across theories and is still the object of significant discussions and debates. For the most part, the various categories of decisions differ in terms of what is assumed to be known to the decision-maker. In the simplest situations of *certainty*, the decision-maker is assumed to know the *unique consequence* of every decision. In situations of *risk*, analyzed for instance in Von von Neumann and Morgenstern (1947) and in most standard microeconomics textbooks such as Varian (1995), every decision is described as a *probability distribution*—sometimes called a lottery—over a set of consequences. In situations of *uncertainty,* formalized in Savage (1954), a decision, often referred to as an *act,* is a function from a set of *states of nature* to a set of consequences. In situations of *complete uncertainty* or *ignorance,* as studied in the literature surveyed by Barberà et al. (2004), a decision is described even more parsimoniously by the *set* of its foreseeable consequences, without reference to states of nature or to a process that maps those states into consequences.

All these descriptions of decisions are "consequentialist" in the sense that they postulate a set of ultimate consequences that matter to the decision-maker and for which he/she has preference. The "risk", "uncertainty", "ignorance", "ambiguity" or other aspect of decision-making is captured through the particular process relating the decision to the consequences that it entails. By contrast, the description of decisions provided by the BJ approach (the axiomatization being provided by Bolker, 1966; 1967) and the philosophical justification by Jeffrey, 1983b) is *not* consequentialist in this sense. It does not therefore posit a priori the existence of consequences that ultimately matter for the decision-maker and which could result from a decision. It does not exclude them either. Rather, in a somewhat abstract fashion, the BJ approach depicts a decision as a "proposition"—called "prospect" by Bradley—that is formally defined as a set of objects—"worlds" in Jeffrey's terminology—whose interpretational status is not explicitly spelled out. This interpretative vagueness of the decisions' elements is viewed by the defenders of the BJ approach as one of its main advantage. And they have some reasons for doing so. It is not indeed always clear, when describing day-to-day decision-making, what are the "ultimate consequences" of a decision. Consider for instance the example, discussed by Savage (1954), of the consequence "having a refreshing swim with a friend". As noticed by Broome (1990), it is not clear that this seemingly precise description of a "consequence" is really the end of the story: "If I have a refreshing swim with friends I might or might not get cramp, and my preferences about the swim will depend on my beliefs about the probabilities of these results. If I swim and get cramp, I might or might not drown, and my preferences about swimming and getting cramp will depend on my beliefs about the likelihood of these results. And so on" (Broome, 1990, 481). The advantage of the BJ description of decisions as "propositions"—such as "having a refreshing swim with a friend"—is to dispense one from the obligation of distinguishing consequences from the processes leading to them. Hence, the proposition "having a refreshing swim with a friend" describes all the "worlds" to which this verbal depiction can be associated: "having a refreshing swim with a friend and having a cramp and drawing", "having a refreshing swim with a friend and having a cramp without drawing", "having a refreshing swim with a friend without having a cramp", etc. There is no clear distinction, in the description of such propositions, between ultimate consequences and the event on the occurrence of which the consequence is contingent.

While the refusal to distinguish a priori between consequences and, say, states of nature (as in the Savage approach) can be seen as an advantage of the BJ theory, its limitations are also worth pointing out, be it simply because they may explain why the BJ theory has not succeeded in imposing itself as a plausible theory of decision-making in economics or psychology. Beside its rather abstract character, *four* problems limit its usefulness as a guidance for understanding actual decision-making.

The *first* limitation of the BJ theory concerns its main criterion for comparing the various propositions (sets of worlds), taken from some algebra of sets, by means of a "conditional expected utility" criterion of the following kind: Proposition A ("having a refreshing swim with a friend") is considered better than Proposition B ("not having the swim and staying instead on the beach") if and only if

$$\frac{v(A)}{\rho(A)} \geq \frac{v(B)}{\rho(B)} \tag{1}$$

for two disjoint-additive set functions v and ρ, with ρ strictly positive. The expression $\frac{v(A)}{\rho(A)}$ can be viewed as the utility, for the decision-maker, of proposition A. Let us denote this by $U(A)$. We therefore have $U(A) = \frac{v(A)}{\rho(A)}$ or, equivalently:

$$v(A) = \rho(A)U(A).$$

Suppose that A_1 is the proposition "I have a refreshing swim with a friend and I don't get a cramp" and A_2 is the proposition "I have a refreshing swim with a friend and I get a cramp". The utility of proposition A can then be written as (exploiting disjoint set additivity):

$$\begin{aligned} U(A) &= \frac{v(A_1) + v(A_2)}{\rho(A_1) + \rho(A_2)} && (2) \\ &= \frac{\rho(A_1)U(A_1) + \rho(A_2)U(A_2)}{\rho(A_1) + \rho(A_2)} && (3) \end{aligned}$$

That is, the utility of "having a refreshing swim with a friend" *can be* interpreted, in the BJ approach, as the expected utility of having such a swim without cramp (A_1) and that of having the very same swim with a cramp (A_2), with expectations taken with respect to the probabilities $\rho(A_1)/(\rho(A_1) + \rho(A_2))$ and $\rho(A_2)/(\rho(A_1) + \rho(A_2))$ of not having and having (respectively) a cramp conditional upon the decision to have a refreshing swim. The emphasis put on the "can be" is particularly important here because the nice theorem, proved by Bolker (1966), that provides an axiomatic justification for evaluating decisions by the utility function U of Expression (3) does not characterize the disjoint additive set functions v and ρ *uniquely*. As shown by Bolker (1966), and discussed by many authors, there are other ways of writing the utility function U that do not take the expected utility form described in (3) that enables the identification of both the utility U of a proposition and its probability ρ. This lack of clear identification contrasts with what is obtained by the Savage approach in which the probabilities and the utility that are expectationnaly combined in the numerical representation are uniquely identified up to a linear transformation.

The *second* problem with the BJ theory lies in its description of decisions as *atomless* propositions. In the BJ approach, a proposition such as "having a refreshing swim with a friend" can always be arbitrarily subdivided into finer propositions without limit. While one could consider that "having a swim" can be subdivided into "having a swim with a cramp" and "having a swim with no cramp", and that "having a swim with a cramp" can itself be subdivided between "having a swim and a cramp and drowning" and "having a swim and a cramp and not drowning", there must presumably a point where such subdivision is not any more possible (for example having a swim, and a cramp and drowning up to death). However, the axiomatic characterization by Bolker (1966) of the class of preferences that can be numerically represented as per inequality (1) rides crucially on the fact that decisions are atomless propositions. Richard Bradley also endorses this "atomlessness" assumption in many parts of his book. One really wonders why, especially given the existence of axiomatic characterizations of rankings of sets representable as per (1) for some additive set functions v and ρ that do not ride on atomlessness (see e.g., Gravel et al., 2018). There are for sure other axiomatizations of criteria in conventional decision theory that ride on similar looking requirements of uncountability of the compared objects. The classical axiomatization of expected utility by Savage (1954) is one of them. Yet, the atomlessness of the propositions in BJ approach strikes me as a significantly less intuitive requirement than that of having a non-countable set of events on which the Savagian acts are defined. In effect, a proposition in the BJ setting is interpreted as an object of choice that is both an event and a consequence. A proposition is literally a sentence that describes the world (e.g., "having a refreshing swim with a friend"). It is somewhat counterintuitive that the only such sentences to which the BJ approach applies are those which are atomless. Sentences, it seems to me, are genuinely finite objects, be it simply because they are finite collection of words. And the descriptions of the world that they provide are also finite, at least when they are as completely specified as possible (e.g., "having a refreshing swim with a friend, having a cramp and dying"). I therefore think that having a theory of decision-making over propositions that rides so heavily upon the atomlessness of these propositions is somewhat limitative.

A *third* limitation of the BJ approach lies in its difficulty in handling dominance reasoning. The well-known Newcomb's paradox, beautifully exposed in Nozick (1969), provides a good example of the issue at stake. Imagine that a (male) decision-maker is told that a good (female) predictor of his behavior has decided to put—or not—$1 000 000 in a closed box that stands nearby an open box in which $1 000 is already present. The decision-maker is asked to choose between:

1) taking *only* what is in the closed box, and leaving the $1 000 in the open one.

2) taking *both* the $1 000 *and* whatever there is in the closed box (that is, zero or $1 000 000).

Before deciding, the decision-maker is told that the predictor has put the $1 000 000 in the closed box only if she anticipates that the decision-maker will choose option 1. If, on the other hand, the predictor expects the decision-maker to choose option 2, then the predictor has put nothing in the closed box. A BJ decision-maker who would attach high probability to the predictor's ability to anticipate his behavior would choose decision 1. In effect, let A_1 and A_2 be the propositions associated to decisions 1 and 2 respectively, and let A_{11}, A_{12}, A_{21} and A_{22} be the following propositions:

A_{11} = "take decision 1 under the assumption that the predictor has correctly anticipated the decision taken",

A_{12} = "take decision 1 under the assumption that the predictor has not anticipated the decision taken",

A_{21} = "take decision 2 under the assumption that the predictor has correctly anticipated the decision taken",

A_{22} = "take decision 2 under the assumption that the predictor has not anticipated the decision taken".

Hence, applying Formula (3) to this decision problem leads to the conclusion that:

$$\begin{aligned} U(A_1) &= \frac{\rho(A_{11})U(\$1000000) + \rho(A_{12})U(0)}{\rho(A_{11}) + \rho(A_{12})} \\ &> \frac{\rho(A_{21})U(\$1000) + \rho(A_{22})U(\$1001000)}{\rho(A_{21}) + \rho(A_{22})} \\ &= U(A_2) \end{aligned}$$

under the assumption that $\frac{\rho(A_{11})}{\rho(A_{11})+\rho(A_{12})} \approx \frac{\rho(A_{21})}{\rho(A_{21})+\rho(A_{22})} \approx 1$ (the decision-maker is fairly confident in the predictor's ability). Yet, the choice of decision 1 seems absurd here. In effect, the predictor's decision to put, or not, \$1 000 000 in the closed box has been taken at the time the decision-maker is asked to make his choice. Hence the decision-maker choice will not affect whether or not the 1 million is there. Either it is there, or it is not. If the million is in the box, then choosing option 1 provides the decision-maker with \$1 000 000, while option 2 gives \$1 001 000. If on the other hand there is nothing in the closed box, then option 1 gives nothing while option 2 gives \$1000. Hence, no matter what has been decided by the predictor, the decision-maker is always better off with option 2 than with option 1. The inability of the expected utility criterion most closely associated to the BJ theory to handle dominance reasoning of this sort is a clear limitation of this criterion.

A last limitation of the BJ approach, which has some bearing for the normative use of the theory, especially to handle collective decisions, lies in its inability to distinguish between the *ex ante* and the *ex post* situation of the decision-maker. Consider for example the case, discussed by Diamond (1967), of the collective decision of giving a unique kidney to two *a priori* equally deserving ill patients: Anna and Bob. It seems plausible that the decision-maker in charge of making the decision be indifferent, preference-wise, between giving the kidney for sure to Ann and giving it for sure to Bob. But Diamond (1967) argues that the decision-maker should then prefer flipping a coin before deciding who should get the kidney rather than giving it for sure to either Ann or Bob. While we may disagree with Diamond on this, it is clear that the distinction between the *ex ante* perspective of putting oneself *before* the resolution of the uncertainty and the *ex post* one of considering the situation after its resolution is an important one, especially for social ethics. Yet, such distinction is very difficult to make in a BJ setting, in which there does not seem to be a "resolution of the uncertainty". As argued by Broome, "the *ex post* approach is not possible within the Bolker-Jeffrey theory" (Broome, 1990, 482).

Richard Bradley's book discusses some of these limitations, albeit often in an allusive way. He does not mention those pertaining to the *ex ante* and *ex post* approaches. As for the Newscomb's paradox, it is only mentioned once in the book (39), and the article by Nozick (1969) is not even quoted. However, the book does spend some time discussing possibilities of allowing for causal (or dominance-based) reasoning in the BJ framework. The book does also mention some of the difficulties associated to the non-unique identification of the probabilities and utilities in Bolker (1966) representation theorem, but does not mention those associated with the atomlessness depictions of decisions.

The book takes for granted that the BJ modeling of decisions is the right way to proceed and develops itself on this premise. Its 304 pages (excluding the appendix and the bibliography) are organized into 14 chapters grouped into four main parts. In the first part, consisting of four chapters, the author discusses the basics of decision theory. The first chapter defends the Bayesian idea of maximizing an expected benefit criterion as an appropriate way to decide. At no place in the book does the author consider decision criteria—such as Maximin or Leximin—that do *not* ride on beliefs about the likelihood of alternative propositions. Accepting this "expectationnal" ideal, the chapter then goes on to oppose the Savage setting to the preferred BJ one. Chapter 2 provides a short discussion of Bayesianism in relation to the postulate, attributed to Hume, that there should be no debate about the rationality or plausibility of individual preferences. Chapter 3 discusses several forms of uncertainty, within the frame of BJ framework. One interesting form of uncertainty that is considered is the evaluative one according to which the decision-maker is uncertain about how valuable prospects are. This form of uncertainty is akin to a form of preference incompleteness. Chapter 4 is devoted to a fairly complete presentation of the Savagian setting, including the well-known critique of the Sure thing principles in the form of Allais paradox. The second part of the book discusses the main BJ theory, and the particular conditional-based approach developed by the author (see, e.g., Bradley, 1999) to handle some of its aforementioned limitations. Indeed, the notion of conditional plays a key role in BJ theory, because the value of a particular decision in this model is nothing else than its conditional expected utility. For example, if one wants to address the aforementioned problem of the Newscomb's paradox in a BJ setting, it is important that one admits, in the set of possible prospects, the prospect in which the predictor has put one million in the closed box and the decision-maker has taken action 1, the prospect in which the predictor has put the million and the decision-maker has taken action 2, etc. The abstract character of the prospects in the BJ approach makes the analysis of conditional somewhat complex, especially when compared with more traditional approaches to conditionals such as that of Luce and Krantz (1971). The basic objects considered in this theory are conditional propositions, which are depicted as functions from a set of states of the world (that will serve as a condition) to a set of consequences. From a formal point of view, the approach of Luce and Krantz (1971), discussed also in (Krantz et al., 1971, ch. 8), contains the BJ approach as a particular case. Richard Bradley only alludes to the Luce-Krantz model by dismissing it as shedding "little light on these quantitative concepts" (Chapter 6). Yet it remains unclear to me that the light shed by the BJ approach is significantly brighter. The third part of the book discusses connections between the BJ approach and formal logic (Chapter 8), as well as with conventional decision theory (Chapter 9). In particular, Chapter 9 discusses the possibilities of integrating both the classical von Neumann and Morgenstern (1947) approach and the Savagian one in the abstract BJ setting. For VNM, this is done by first postulating, from the start, that some of the prospects in the collection are chance prospects, defined to be prospects "to which it is meaningful to ascribe chances" (113). Bradley then goes on in considering all probability distributions that can be assigned to these chance prospects, and by constructing, with these distributions, chance prospects of the type "the chance of prospect X is x, and the chance of prospect $\neg X$ is $1 - x$". For the Savagian setting, one similarly defines, again exogenously, consequences as "prospects that are maximally specific with regard to all that matters to the agent" and states of nature as "prospects which are maximally specific with regard to all possible features of the environment relevant to the determination of the consequence" (167). Once this exogenous identification of the consequences and the states of nature is done, one can define the Savagian acts as specific conditional statements connecting the second kind of prospects (states of nature) to the first kind (consequences). The author is not precise as to what the criterion of "maximal specificity" (for either the consequence or the state of nature) could be, especially if the prospects are atomless. Is the prospect "taking a refreshing swim with a friend without a cramp" maximally specific, for example? Chapter 10 discusses a proposal to introduce learning in a BJ setting (Chapter 10). The last part of the book discusses some issues that arise when the BJ decision-makers are assumed to be boundedly rational. Various kinds of "bounds" on the agents' rationality are considered, among which incomplete preferences, unawareness, and ambiguity.

The book is undoubtedly an excellent synthesis of the BJ theory. It is comprehensive and rigorous, even thought the formalism of atomless boolean algebra on which the BJ theory is built is sometimes introduced in a slightly "casual" fashion which makes the reading difficult. The book is a bit idiosyncratic—by presenting R. Bradley's view of the issue—and is clearly written for a "converted" (to the BJ approach) audience, who is likely to be interested by the original proposals developed by the author to address some of the limitations of the BJ framework. On the other hand, people like myself who are somewhat doubtful about the appeal of the BJ approach for analyzing decision-making may not be totally inclined to "revise their belief" after reading the book. But Richard Bradley should certainly not be blamed for this. Convincing the skeptics has always been more difficult than preaching to the converted.

References

Barberà, Salvador, Walter Bossert, and Prasanta K. Pattanaik. 2004. Ranking Sets of Objects. In Salvador Barberà, Peter J. Hammond, and Christian Seidl (eds), *Handbook of Utility Theory, vol. 2: Extensions*, Dordrecht: Kluwer. 893–977.

Bolker, Ethan D. 1966. Functions Resembling Quotient of Measures. *Transaction of the American Mathematical Society*, *124*(2): 292–312.

Bolker, Ethan D. 1967. A Simultaneous Axiomatization of Utility and Subjective Probabilities. *Philosophy of Sciences*, *34*(4): 333–340.

Bradley, Richard. 1999. A Representation Theorem for a Decision Theory with Conditionals. *Synthese*, *116*(2): 187–229.

Broome, John. 1990. Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism. *Review of Economic Studies*, *57*(3): 477–502.

Diamond, Peter A. 1967. Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility: a Comment. *Journal of Political Economy*, *75*(5): 765–766.

Gravel, Nicolas, Thierry Marchant, and Arunava Sen. 2018. Conditional Expected Utility Criteria for Decision Making under Ignorance or Objective Ambiguity. *Journal of Mathematical Economics*, *2018*(78): 79–95.

Jeffrey, Richard. 1983a. Bayesianism with a Human Face. In John Earman (ed.), *Testing Scientific Theories*, Midwest Studies in the Philosophy of Science vol X, Minneapolis: University of Minnesota Press. 133–156.

Jeffrey, Richard. 1983b. *The Logic of Decision*. Chicago: University of Chicago Press. 2nd edition.

Joyce, James M. 2000. Why We Still Need the Logic of Decision. *Philosophy of Sciences, Proceedings*, *67*: S1–S13.

Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky. 1971. *Foundations of Measurement*. New York & London: Academic Press.

Luce, R. Duncan and David H. Krantz. 1971. Conditional Expected Utility. *Econometrica*, *39*(2): 253–271.

von Neumann, John and Oskar Morgenstern. 1947. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Nozick, Robert. 1969. Newcomb's Problem and Two Principles of Choice. In Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel*, Dordrecht: Reidel. 114–146.

Savage, John L. 1954. *The Foundation of Statistics*. New York: Wiley.

Varian, Hal R. 1995. *Microeconomic Analysis, 3rd edition*. New York: Norton.