Economics and Philosophy, **32** (2016) 409–433 © Cambridge University Press
doi:10.1017/S0266267115000346 First published online 12 August 2015
journals.cambridge.org/eap

ON THE INTERPRETATION OF DECISION THEORY

SAMIR OKASHA*

Abstract: This paper explores the contrast between mentalistic and behaviouristic interpretations of decision theory. The former regards credences and utilities as psychologically real, while the latter regards them as mere representations of an agent's preferences. Philosophers typically adopt the former interpretation, economists the latter. It is argued that the mentalistic interpretation is preferable if our aim is to use decision theory for descriptive purposes, but if our aim is normative then the behaviouristic interpretation cannot be dispensed with.

Keywords: Decision theory, expected utility, behaviourism, mentalism

1. INTRODUCTION

Modern decision theory is a cross-disciplinary enterprise, spanning economics, statistics, philosophy and psychology. This is reflected in the recent history of the subject; key contributors in the 20th century include the philosopher Frank Ramsey (1931), the mathematician/economist pair John von Neumann and Oskar Morgenstern (1944), and the statistician Leonard Savage (1954). Despite this fact, the standard *interpretation* of decision theory appears to differ widely across these disciplines. In particular, there is a striking mismatch between how economists and philosophers typically understand decision theory, which has impeded communication between them. The aim of this paper is to discuss this mismatch, diagnose its roots, and offer a tentative adjudication.

My focus will be mostly on expected utility (EU) theory, in both its 'objective' and 'subjective' versions. EU is the classical theory of

* Department of Philosophy, University of Bristol, Cotham House, Bristol BS6 6JL, UK. Email: Samir.Okasha@bristol.ac.uk. URL: http://www.bristol.ac.uk/school-of-arts/people/samir-okasha.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

decision under risk/uncertainty, and typically the only one discussed in the philosophical literature. Indeed many philosophers appear to use 'decision theory' simply to mean EU theory. From the economist's point of view this may seem odd, given the numerous alternatives to EU developed in the economics literature of the last thirty years,[1] but its explanation lies in the fact that philosophers are usually interested in decision theory construed normatively rather than descriptively, i.e. as a theory of ideally rational, rather than actual, choice. Economists by contrast are typically interested in the descriptive construal; the point of developing non-EU theories is precisely to account for observed behaviour.

The normative/descriptive dichotomy will be discussed below; but it is the dichotomy between *behaviouristic* and *mentalistic* interpretations of decision theory that will occupy centre stage.[2] The former regards preferences or choices as primary and utilities and credences as derivative; maximization of expected utility is a strictly 'as if' story, on this view. The mentalistic view, by contrast, treats an agent's utility function and credence function as psychologically real, and capable of causing/explaining their preferences and choices. Economists typically endorse the behaviouristic view, and often explicitly reject the mentalistic view as wrong-headed; but among philosophers the mentalistic view is widespread. Indeed many philosophers who discuss decision theory simply assume the mentalistic interpretation without argument; while others explicitly argue against the reigning behaviouristic orthodoxy of the economists.

This situation prompts an important question. Given that behaviourism as a general view of the mind is widely discredited among philosophers and psychologists, does this speak against the behaviouristic interpretation of decision theory and in favour of the mentalistic? The answer to this question, I will argue, is 'it depends'. In particular, it depends on whether one wishes to use decision theory for normative or descriptive purposes. So there is an interesting interaction between the normative/descriptive issue and the behaviouristic/mentalistic issue. I argue below that standard anti-behaviourist considerations do gain traction if one wishes to use decision theory for descriptive purposes; but for normative purposes matters are rather different.

[1] For reviews of this work at different stages of its development see Machina (1987), Starmer (2000) or Wakker (2010).

[2] The contrast between these two interpretations of decision theory has been discussed many times, under various labels; see in particular Hansson (1988), Bermudez (2009) and Buchak (2013). Dietrich and List (forthcoming) study the mentalist/behaviourist opposition (under those labels) in relation to microeconomics more generally.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

The structure of this paper is as follows. Section 2 is a brief outline of the origins of modern decision theory, the point of which will become clear. Section 3 discusses the contrast between behaviourist and mentalistic interpretations. Section 4 contrasts normative and descriptive uses of decision theory. Sections 5 and 6 ask how the behaviouristic/mentalistic distinction relates to the descriptive/normative distinction. Section 7 is a critique of other philosophers' views on the foundations of decision theory. Section 8 concludes.

2. THE ORIGINS OF EXPECTED UTILITY THEORY

The origins of EU theory lie in Daniel Bernoulli's attempt to explain the observation that in games of chance, people typically do not maximize expected monetary value (Bernoulli 1738). This observation, made dramatic by the 'St Petersburg paradox', today enjoys the status of a well-confirmed empirical fact. For example, agents typically prefer £5 for sure to a gamble which pays either £10 or nothing depending on the flip of a fair coin, despite both options having the same expected monetary value. Bernoulli argued that people's choices instead maximize expected *utility*, and suggested that an agent's utility function is the logarithm of money.

Two points about Bernoulli's theory deserve note. First, Bernoulli offered no argument for *why* a person should try to maximize their expected utility (nor for why their utility function should be logarithmic). Had he been asked to explain why an agent should not care about the variance in utility of a gamble, as well as its expected utility, for example, he would have had no answer. In so far as Bernoulli's aim was simply to describe or explain agents' actual choices, his inability to answer this question may not matter. But if the principle of expected utility maximization is construed as normative then the question cannot be ducked.

Second, Bernoulli understood 'utility' in a mentalistic way, i.e. as a measure of how much happiness an agent gets from a given amount of money. This raises an immediate question. How do we know that happiness or satisfaction can be quantified at all, and in particular why think it should be measurable on a cardinal scale, as it must if the idea of expected utility maximization is to make sense? Bernoulli offered no answer to this question.

Modern decision theory borrowed the idea of EU maximization from Bernoulli but developed it in a different way, and in the process supplied answers, of a sort, to the two questions above. The key idea, common to the treatments of Ramsey, Savage and von Neumann & Morgenstern, is to deduce EU maximization from a more fundamental basis. Their starting point is an agent's preference relation between certain options, which

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

can in principle be discovered by observation. This preference relation is assumed to satisfy certain axioms that are meant to be requirements of rationality. It is then shown, via a representation theorem, that an agent whose preferences satisfy the axioms necessarily behaves *as if* they are an EU maximizer, and vice versa. Thus the EU principle does not have to be taken as a primitive, but can be deduced from something supposedly more basic. Though this style of argument will be familiar to many readers, it is worth spelling out in more detail.

2.1 von Neumann & Morgenstern's theory

In von Neumann & Morgenstern's theory, an agent is faced with a choice between lotteries, where a lottery is an objective probability distribution over a finite set O of outcomes, or prizes. Thus for example, O might be a set of holiday destinations, e.g.{France, Spain, Italy}. The prizes are strict alternatives, i.e. only one will occur. A lottery specifies the probability (i.e. objective chance) that each member of O has of occurring; thus for example (France, 1/3; Spain 1/3, Italy 1/3) is the lottery in which each of the three holiday destinations occurs with equal probability. An outcome may be identified with the degenerate lottery in which it receives probability 1. The set of all lotteries over O is denoted L.

The agent is assumed to have a (weak) preference relation R over the lottery set L; 'xRy' means that the agent weakly prefers lottery x to y, where $x, y \in L$. This is often interpreted to mean that the agent would never choose y over x if both options were available.[3] The preference relation R is required to satisfy three conditions.[4] Firstly, R should be a complete and transitive binary relation, i.e. a weak ordering; secondly, R should satisfy a technical condition called continuity; and thirdly, R should satisfy the famous independence axiom, which says, in effect, that the agent's preference between x and y should not depend on which other alternatives they are 'mixed' with.[5] The second condition is needed to make the maths work; the first and third may be argued to be constraints that any rational agent's preference relation should satisfy. How compelling these constraints are is a much-debated question that we need not enter into for now.

From this starting point, von Neumann & Morgenstern then prove their celebrated expected utility theorem. This theorem says that if and

[3] The idea that preference can be reduced to hypothetical choice in this way is a key tenet of 'revealed preference theory' (cf. Sen 1971); see Section 2 below.

[4] Here I follow the simplification of the von Neumann & Morgenstern axioms originally devised by J. Marschak (1950) and standardly found in modern textbooks.

[5] Formally, the independence condition says that for any lotteries x, y, z and any probability $p > 0$, xRy iff $(x, p; z, 1-p)\ R\ (y, p; z, 1-p)$, where '$(x, p; z, 1-p)$' is the compound lottery that yields x with probability p and z with probability $1-p$.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

only if an agent's preference relation R satisfies the three conditions above, then there exists a real-valued utility function on O, such that the agent evaluates lotteries in L in accordance with the expectation of that utility function, and always prefers the lottery with the highest expected utility. The utility function is unique up to positive linear transformation. Thus so long as an agent's preference relation obeys the three conditions, then she is behaving *as if* she assigned numerical utilities to the prizes and was consciously computing expected utility.

2.2 Savage's theory

The von Neumann & Morgenstern theory has limited scope, since it deals only with decision making under risk, i.e. where the objective chances of the outcomes are known. Gamblers in casinos face decision problems of this sort, but most real decisions are done in the face of uncertainty rather than risk, i.e. where objectives chances are not known. Both Ramsey (1931) and Savage (1954) devised versions of expected utility theory to cover such decisions; here I focus on Savage's version.

The primitives of Savage's theory are a set O of outcomes and a set S of possible states of the world, used to represent the agent's uncertainty. Thus for example S might be {rainy, sunny, cloudy} and O might be {£10, £5, £0}.[6] An agent is faced with a choice between *acts*, where an act is any function from S to O. Thus for example one act is (£5 if rainy, £5 if sunny, £0 if cloudy), while another is (£10 if rainy, £0 if sunny or cloudy). Intuitively, the agent's preference between these two acts will depend partly on their utility function for money, and partly on their subjective beliefs about which state of the world is most likely to occur. The set of all acts, i.e. functions from S to O, is denoted A.

The agent has a preference relation R on the set A, on which Savage imposes a number of axioms. Thus for example one axiom requires that R be complete and transitive, while another requires that R satisfy the famous 'sure thing' principle (the analogue of the von Neumann and Morgenstern independence postulate). Again, there is considerable debate about whether these axioms are indeed rationally compelling.

Savage then proves that if and only if the agent's preference relation satisfies his axioms, then there exists a unique probability distribution over S, and a utility function over O unique up to positive linear transformation, such that the agent's preference between any two acts is always for the one with the highest expected utility. Therefore the agent behaves as if they have probabilistic beliefs about the states of the world

[6] For technical reasons, Savage's own construction requires that the set S be at least countably infinite, but the essence of his theory can be illustrated with finite S. Wakker (2010) proves a version of Savage's result for finite S (theorem 4.6.4 on p. 112).

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

and a utility function over the outcomes, *modulo* which they choose the act that maximizes expected utility. So Savage succeeds in extracting both utilities and (subjective) probabilities from the agent's preferences.

2.3 Significance of the modern expected utility theories

The significance of the Savage and von Neumann & Morgenstern results is that the input to the theorems – the agent's preference ordering – is in principle amenable to observation and introspection. A suitably placed observer, in favourable conditions, could in principle infer from observing an agent's choice behaviour whether they prefer lottery x to y, or act a to b. And the agent themselves, by careful introspection, could presumably infer their (hypothetical) preference between any two lotteries or acts. Such inferences are fallible, for both observer and agent; but even so, the situation is clearly quite different from a theory such as Bernoulli's which posits a real-valued utility function, measurable on a cardinal scale, straight off.

This means that Savage and von Neumann & Morgenstern are able to answer the two questions that Bernoulli's early theory prompted. The first question was how to justify the principle of EU maximization. Bernoulli had no real answer to this question, but Savage and von Neumann & Morgenstern do, for in their theories the principle is not taken as primitive, but deduced from what are arguably more primitive rationality constraints. They show that if an agent cannot be represented as an EU maximizer, then their preferences must violate at least one of the axioms. Presuming the axioms are accepted, this provides a normative basis, of a sort, for EU maximization.

This contrast can be sharpened by considering the question: why should an agent not care about the *variance* of the utilities associated with a given lottery or act, in addition to the expectation? On a theory such as Bernoulli's this question is a pressing one. But for von Neumann & Morgenstern and Savage, the question simply does not arise. So long as the agent's preference between lotteries/acts satisfies the relevant axioms, then there exists a utility function such that the agent necessarily behaves as if they are trying to maximize its expectation. That an agent should care only about an act's expected utility, and not about its variance in utility, is thus built into the construction of the utility function, on the modern view.

The second question facing Bernoulli's theory was: what licenses the assumption that utility can be quantified on a cardinal scale? If utility is primitive, then treating it as real-valued and cardinally measurable is to make a substantive psychological assumption that needs justification. Perhaps some such justification could be given; but von Neumann & Morgenstern and Savage bypass this problem altogether, for they prove the existence of a real-valued cardinal utility function from the preference

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

axioms. So the device of beginning with preferences over risky/uncertain options and 'reverse engineering' an agent's utility function permits a neat answer to the challenge to justify the assumption of cardinally measurable utility.

3. MENTALISTIC VERSUS BEHAVIOURISTIC INTERPRETATIONS OF EU THEORY

The difference between Bernoulli's and von Neumann & Morgenstern's and Savage's theories illustrates the contrast between mentalistic and behaviouristic interpretations of EU theory. On the mentalistic interpretation, the utility that an outcome has for an agent is regarded as a psychological fact about that agent. Various accounts might be given of what sort of fact this is. For example, one might think of utility in hedonic terms à la Bentham, or alternatively as a measure of subjective or objective well-being (cf. Broome 1991); and one might affirm, or deny, that an agent's utility is always accessible to introspection. The key point is that on this view, facts about an agent's utility function are not reducible to, or logical constructions out of, or re-descriptions of, facts about their preferences or choices. Rather they are self-standing psychological facts. The mentalistic interpretation of EU theory combines this notion of utility with the principle of EU maximization.

On the behaviourist view, utility is simply a convenient mathematical device, introduced by the theorist, for re-describing an agent's preferences. No psychological reality attaches to talk of utility, and utility maximization is strictly an 'as if' story. So to say that an agent has a particular utility function over a set of outcomes is simply a way of summarizing their preference ordering over those outcomes. In economists' jargon, utility is simply a representation of preference. This 'representationalist' thesis about the relation between utility and preference is orthodox among neo-classical economists, and applies more generally than to preferences over acts or lotteries. The behaviourist interpretation of EU theory combines this representationalist notion of utility with the principle of EU maximization.

A further behaviourist idea is that preference itself is reducible to (hypothetical) choice, i.e. an agent's preference ordering over a set of alternatives is a summary of her binary choices between them. Thus an observer could literally deduce an agent's preference ordering from observations of her choices. This idea, which forms part of 'revealed preference theory', is often found in the company of what I am calling the behaviourist interpretation of EU theory, but is logically distinct from it. To see this, note that someone who rejects revealed preference theory, e.g. who holds that preferring *a* to *b* is a *sui generis* mental state that causes an agent to choose *a* over *b* (so is not reducible to the latter), could still

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

hold that the agent's utility function is a mere representation of their preferences.[7] Thus there are two issues: the relation between utility and preference, and the relation between preference and choice. Our concern here is with the former issue. Thus mentalism and behaviourism, in this essay, denote alternative positions about the relation between utility and preference, which are compatible with different views about how preference relates to choice. The label 'behaviourist' is appropriate since on any view, an agent's preference ordering is 'closer' to her observable behaviour than is a real-valued utility function.

To sharpen the mentalist/behaviourist contrast, consider the question of whether EU theory can explain why an agent prefers a to b by saying that the expected utility of a exceeds that of b. On the behaviourist interpretation the answer is 'no', presuming that explain means *causally* explain. For a behaviourist, the claim that $EU(a) > EU(b)$ is equivalent to a statement about the agent's preferences; specifically, it means that the agent's preference ordering over the outcome set O of which a and b are members can be represented by the expected value of a (utility function, credence function) pair, *modulo* which $EU(a) > EU(b)$. In other words, what makes it the case that $EU(a) > EU(b)$ is that the agent exhibits a particular pattern of preferences. This obviously precludes the fact that $EU(a) > EU(b)$ from constituting a causal explanation of any of those preferences. If utility is construed mentalistically, by contrast, then this is a potentially valid causal explanation.

It is perhaps unsurprising that most philosophers have favoured the mentalistic interpretation of EU theory. For the opposition between behaviourist and mentalist interpretations looks like an instance of the more general clash between behaviourism and mentalism in psychology, where it is widely held that behaviourism has been refuted. Moreover, many philosophers have wanted to view EU theory as a formalization of ordinary folk psychology, with utility and credence being the quantitative counterparts of desire and belief respectively, and the principle of maximizing expected utility corresponding to the Humean 'belief-desire' law.[8] Thus for example David Lewis regards EU theory as 'the very essence of our common sense theory of persons, elegantly distilled and systematized' (1974: 337). Lewis's idea really only makes sense on a mentalistic interpretation. For it is a standard view, among philosophers and the folk themselves, that desires and beliefs *do* cause actions, and that citing an agent's desires and beliefs *does* serve to causally explain their behaviour. If decision theory is to be regarded as a formalization of

[7] Here I am indebted to Richard Bradley.

[8] Hampton (1994) provides some telling criticisms of this supposed connection between EU theory and belief-desire psychology.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

folk psychology, which is an undeniably attractive idea, the behaviourist interpretation cannot be sustained.

However the founders of modern EU theory were adamant that their theory was to be interpreted behaviouristically.[9] One clear statement of this is by Friedman and Savage (1948) who wrote that the von Neumann & Morgenstern expected utility hypothesis

> asserts that individuals behave as if they calculated and compared expected utility and as if they knew the odds … the validity of this assertion does not depend on whether individuals know the precise odds, much less on whether they say that they calculate and compare expected utilities or think that they do, or whether psychologists can uncover any evidence that they do, but solely on whether it yields sufficiently accurate predictions about the class of decisions with which the hypothesis deals. (1948: 282)

In a similar vein, J. Harsanyi (1977) wrote, in relation to Savage's theory:

> a Bayesian need not have a special desire to maximize expected utility per se. Rather, he simply wants to act in accordance with a few very important rationality axioms, and he knows that this fact has the inevitable mathematical implication of making his behaviour equivalent to expected-utility maximization. As long as he obeys these rationality axioms, he simply cannot help acting as if he assigned numerical utilities … to alternative possible outcomes of his behaviour, and assigned numerical probabilities … to alternative contingencies that may arise, and as if he tried to maximize his expected utility in terms of these utilities and probabilities (1977: 381).

Here Harsanyi explicitly endorses the behaviourist claim that to describe an individual as maximizing EU is simply to say that their preferences satisfy certain axioms.

Another facet of the behaviourist interpretation is the insistence that an agent's utility function, as defined by the EU theorems, has nothing to do with the cardinal utility of the 19th century economists. The latter held, with Bentham, that 'utility' was a measure of the pleasure or happiness that an agent gets from a good, with the cardinality deriving from the assumption that the *intensity* of pleasure can be quantified. On this traditional view, cardinal utility had nothing in particular to do with risk or uncertainty, and was supposed to apply to choices between certain outcomes. However both Savage and Arrow claimed that in EU theory, the 'utility' of an outcome must *not* be identified with the amount of happiness an agent would get from its certain receipt. Thus Arrow (1951: 425) wrote that utility conceived of in this way was 'a meaningless concept', while Savage (1954: 93) cautioned against confusing his notion

[9] A possible exception is Frank Ramsey (1931), whose views on the behaviourism versus mentalism issue are not easy to discern. See Bradley (2004) for an insightful discussion of Ramsey's position.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

of utility with 'the now almost obsolete notion of utility in riskless circumstances'.

In their influential *Games and Decisions* (1957: 32), R. Luce and H. Raiffa discuss a number of common 'fallacies' about EU theory. The first fallacy is that an agent prefers lottery x to lottery y because $EU(x) > EU(y)$. This 'is the exact opposite of the truth', they argue. The second is that a rational agent might care about the variance of utilities, as well as the expectation. This is a 'completely wrong interpretation of the utility notion', they claim. Luce and Raiffa argue that both of these fallacies stem from 'a failure to accept that preferences precede utilities'. They insist that EU theory should not be treated as a theory about what is going on in the heads of rational decision makers: 'there is no need to assume, or to philosophize about, the existence of an underlying subjective utility function, for we are not attempting to account for the preferences or the rules of consistency. We only wish to devise a convenient way to represent them.'

The behaviourist interpretation of EU theory is still pervasive among contemporary economists, and routinely taught as part of microeconomic orthodoxy.[10] Indicative of this is the textbook treatment of risk aversion. Empirical work shows that people's preferences over monetary gambles are generally risk averse, i.e. they prefer to receive $\$x$ for certain to a gamble with expected monetary value of $\$x$. One might think that EU theory can explain this fact, by hypothesizing, with Bernoulli, that agents' utility functions for money are concave. However this is only a legitimate explanation on a mentalistic interpretation of EU theory. Economics textbooks, by contrast, simply *define* risk-aversion as concave utility – the standard 'coefficient of risk aversion' is simply a measure of the concavity of an agent's utility function for money.[11] So there is no question of concave utility *explaining* risk-averse preferences; they are simply one and the same thing, on the textbook view.

Given that the founders of EU theory insisted on the behaviouristic interpretation, it is striking that contemporary philosophers almost always adopt the mentalistic interpretation. Thus for example D.H. Mellor (2005: 140), in a discussion of 'subjective decision theory' (SDT), writes: 'actions of whose aetiology SDT is true are not only thereby causally explained by the subjective credences and utilities which cause them; those actions are also thereby rationalized, since credences and utilities which cause actions as SDT says are *reasons* for acting in a quite standard

[10] See for example Gilboa (2009), or Mas-Collel *et al*. (1995) which defend the behaviourist view and take it for granted, respectively. See also the section devoted to the 'causal utility fallacy' in Binmore (2008: 19–22).

[11] This coefficient, known as the 'Arrow-Pratt' coefficient of risk aversion, is defined as $u''(x)/u'(x)$ where 'x' is money, i.e. the second derivative of the utility function divided by the first. Since $u'(x)$ is always positive – the agent prefers more money to less – the sign of the coefficient of risk aversion is given by the sign of $u''(x)$.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

sense'. In describing the credences and utilities of Bayesian decision theory as 'causes', Mellor takes himself to be stating an obvious and widely held view; yet this view was explicitly rejected by the originators of Bayesian decision theory, as we have seen.

The prevalence of the mentalistic interpretation in the philosophical literature is manifest in how decision theory is often presented. The typical philosopher's presentation begins with a utility and a credence function, often in the form of a 2×2 matrix, then explains that decision theory prescribes the rule of maximizing expected utility. Often there is little discussion of what the utility numbers mean, or where they come from, and in particular no mention of the idea that they are derived from preferences via a representation theorem. This is true of Nozick's (1969) paper on Newcomb's problem and much of the ensuing literature on that topic. For example D. Lewis (1981) introduces what he takes to be standard decision theory as follows: 'a rational agent has at any moment a credence function and a value function … each world *W* has a value V(*W*), which measures how satisfactory it seems to the agent for *W* to be the actual world … decision theory … prescribes the rule of V-maximizing, according to which a rational choice is one that has greatest expected value' (1981: 6). Lewis makes no mention of preferences at all; he appears to conceive the agent's 'value function' (i.e. utility function) mentalistically. However he says nothing about what why it should be supposed cardinally measurable, nor why rationality prescribes maximization of its expectation.

This observation is not necessarily a criticism of those philosophical discussions to which it applies. For some purposes, it may not matter whether utility is treated as psychologically real or as a representation of preference; it seems likely, for example, that the 'causal versus evidential' issue that is the focus of much philosophical interest will arise anyway. Moreover, it may be pedagogically useful to introduce the principle of EU maximization by simply positing probabilities and utilities *ab initio*, as for example R. Jeffrey does in the first chapter of *The Logic of Decision* (1990), whatever one's view of the behaviourist versus mentalism issue. And finally, it may be that there are sound arguments against the behaviourist interpretation anyway; so that in presenting EU theory in mentalistic guise, philosophers are taking themselves to be offering a more methodologically acceptable version of that theory, free from unnecessary behaviouristic shackles.

This last suggestion raises two important questions. First, do the standard anti-behaviourist considerations in psychology and philosophy of mind, which most contemporary philosophers accept, tell against the behaviourist interpretation of EU theory? Second, if so, can we simply adopt EU theory but divest it of the behaviouristic interpretation that its early 20th century pioneers gave it? A number of recent discussions

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

suggest that the answer to both questions is 'yes'. Christensen (2001) and Eriksson & Hájek (2007) both deploy standard anti-behaviourist arguments against the behaviouristic construal of the credences and utilities that feature in subjective EU theory; while Joyce (1999) advocates divesting decision theory of its 'pragmatist' (i.e. behaviourist) commitments. Bermudez (2009) also offers a qualified endorsement of this view. However I think the situation is not quite this simple. The answer to these two questions depends crucially on what use we want to put decision theory to.

4. NORMATIVE VERSUS DESCRIPTIVE USES OF EU THEORY

It is a familiar point that EU theory can either be interpreted normatively or descriptively. These options are not exclusive: the same theory may be able to do both jobs. The normative/descriptive contrast is muddied slightly by the fact that EU theory involves a measure of idealization; those who construe the theory descriptively regard it as a useful model of actual choices, not a literally correct description. However the contrast is still reasonably sharp: a description, even idealized, is different from a prescription.

Economists are typically interested in the descriptive construal of EU theory; their main concern is to describe and predict actual human behaviour (cf. the quote from Friedman and Savage (1948) above). This is not to say that considerations of rationality play no role, but it is derivative. Economists who employ rational choice models do so out of a conviction that humans' economic behaviour, at least in some domains, is largely rational; where this can be shown not to hold, they develop alternative models. Thus the discovery in the 1970s that experimental subjects exhibit preferences that systematically violate EU theory led to the rapid development of 'non-expected utility' theories, in an attempt to improve on EU's predictive fit to the data. This is a large and ongoing area of research (cf. fn 1).

Philosophers by contrast are usually interested in EU theory construed normatively, i.e. as prescribing how choices should be made. Thus much of the philosophical interest has been in hypothetical cases where EU theory appears to offer counterintuitive advice, and what to say about them. Indicative of this difference in focus is that the literature on experimental violations of EU theory has received little attention from philosophers, and the non-expected utility models have hardly been discussed at all.[12] This is not necessarily a lacuna: if one's concern is with

[12] Exceptions to this generalization include Bermudez (2009), Okasha (2011) and Buchak (2013).

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

principles of *rational* choice, one is unlikely to be concerned with theories whose stated intent is to describe *departures* from ideal rationality.

How does the normative versus descriptive issue relate to the mentalistic versus behaviouristic issue? I argue below that if one is interested in EU theory construed descriptively, as are most economists, then either the mentalistic or the behaviouristic interpretations is in principle available; but that general anti-behaviouristic considerations tell in favour of the former. However if one is interested in EU theory construed normatively, as are most philosophers, then the behaviourist interpretation is in a sense mandatory. If this is correct then there is a considerable irony. For most philosophers have favoured the mentalistic interpretation of EU theory, but construed the theory normatively. This is the one option that is not available, if I am right.

5. BEHAVIOURISM VERSUS MENTALISM ON THE DESCRIPTIVE CONSTRUAL

Consider firstly the descriptive construal of EU theory. For concreteness, let us focus on Savage's version. So construed, the theory says that an agent's preferences over uncertain options, i.e. 'acts' in Savage's sense, do in fact satisfy certain axioms; and then proves the existence of a unique credence and utility function, *modulo* which the agent may be represented as maximizing expected utility. Now suppose for the sake of argument that this theory was descriptively successful, i.e. that people's actual preferences satisfy the axioms. (Empirically we know that this is not so; but leave that aside for the moment.) Let us then ask: should we interpret the theory mentalistically or behaviouristically? In particular, should an agent's credence and utility function be treated as fictions, or as psychologically real?

Though Savage insisted on the behaviouristic interpretation, from a modern vantage point this looks untenable. Almost all sciences introduce theoretical posits that go beyond, and are meant to explain, the data; few philosophers today are tempted by an instrumentalist or fictionalist attitude towards such posits. This is as true in psychology as anywhere else; since the 'first cognitive revolution', psychologists have been happy to posit unobservable mental states and processes, many of them inaccessible to consciousness, that are meant to explain behaviour. And in philosophy of mind, it is a commonplace to regard an agent's intentional attitudes, such as beliefs and desires, as internal causes of the agent's behaviour.

Against this background, there seems every reason to regard an agent's credence and utility functions, as defined by Savage's theorem, as psychologically real, and as capable of explaining her preferences and choices. For consider the question: what explains the fact that our agent's

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

preferences satisfy the Savage axioms? This is a legitimate question, and a pressing one given that the vast majority of possible preference orderings over acts do *not* satisfy those axioms. On the mentalistic interpretation, there is potentially a simple answer: the agent has a credence function over states of the world, and a utility function over outcomes, which she uses to compute the expected utility of the available acts before choosing the one with the highest. If the credence and utility functions are viewed as psychologically real, this constitutes a potential causal-computational explanation of why the agent's observed preferences satisfy the Savage axioms. The computations will presumably be occurring at the sub-personal level, but this is quite standard in cognitive psychology. But on the behaviourist interpretation this explanation is unavailable; that the agent's preferences satisfy the Savage axioms is left as an unexplained fact.

Note that this argument is an application of a standard anti-behaviourist line of thought, according to which behaviourism, by refusing to posit unobservable mental states and processes, is left unable to explain certain salient observable regularities. A similar argument is often levelled against instrumentalism about unobservable posits in general, not just psychological posits. There seems no reason why such arguments should not apply to decision theory (construed descriptively). From this perspective, those who insist on the behaviouristic interpretation appear to be cleaving to an outdated, positivistic philosophy of science; this point is argued in detail by Dietrich and List (forthcoming).

In effect, this is to suggest that standard maxims of scientific inference tell in favour of the mentalistic interpretation. Faced with a very specific pattern in our data – preferences that satisfy the Savage axioms – we hypothesize the existence of entities – credence and utility functions that combine in a particular way – to explain the data. This explanation is elegant, and the rule of combination – mathematical expectation – is highly intuitive. Of course there is no *guarantee* that the explanation is correct, but there is strong inductive evidence in favour of it. For unless the agent is performing expected utility calculations at the sub-personal level, how do we explain the remarkable fact that her preferences exhibit the pattern that they do?

This argument tallies with the fact that in recent years certain neuroscientists, notably Paul Glimcher and colleagues, claim to have discovered a neural basis for expected utility maximization. Glimcher (1993) claims that different neural structures encode an agent's credence and utility functions, and that expected utility computations are actually taking place at the neural level. He and his co-authors write: 'the available data suggest that the neural architecture actually does compute a desirability for each available course of action. This is real physical

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

computation, accomplished by neurons, that derives and encodes a real variable' (Glimcher *et al.* 2005: 220). In effect, Glimcher is pursuing a version of the methodological strategy recommended above: treating EU theory as a candidate descriptive theory of how agents actually choose in the face of uncertainty, but refusing to countenance the behaviourist interpretation of that theory.

The foregoing argument is complicated slightly, but only slightly, by the fact that EU theory is known not to fit the available experimental data as well as the non-EU alternatives. For the general point – that there is no more reason to regard credences and utilities as fictions than other scientific posits – extrapolates easily to non-expected utility models. Take for example the well-known 'Choquet expected utility' model of Schmeidler (1989), which represents an agent's preferences by means of a utility function over outcomes and a (non-additive) belief function over events; the model was designed to accommodate observed violations of the EU axioms. In so far as the model provides a good descriptive fit to the data, the anti-behaviouristic considerations outlined above apply just as well. Standard maxims of scientific inference suggest that we should accept these functions as psychologically real, and capable of explaining an agent's preferences. The same applies to other non-EU models too, such as cumulative prospect theory, which currently appears to fit the experimental data better than the alternatives (cf. Wakker 2010).

Two possible behaviourist responses to this argument spring to mind. First, one might deny that general anti-instrumentalist considerations support a mentalistic construal of the utility and credence functions of decision theory. These entities have a specific mathematical structure, and in this respect are different from the internal states and processes that cognitive psychology usually traffics in. Even if one is happy to posit sub-personal internal states to explain behaviour, one might have qualms about positing internal states that satisfy certain specific measurability assumptions. The suggestion is thus that there is a particular reason for construing utility and credence functions in an 'as if' way, that does not apply more generally.

This is a coherent suggestion, but would need considerable elaboration to be convincing. It is true that a real-valued utility function, measurable on a cardinal scale, is unlike the typical internal state that cognitive psychology posits; but if positing such a state is needed to explain the data there can surely be no *a priori* objection. This is an empirical matter. However this suggestion does raise one important issue, which is that proponents of the mentalistic interpretation need to clarify the relation between credences and utilities, conceived as psychologically real, and ordinary beliefs and desires. More generally, what is the relation between an agent's own explanations of their choices in terms of beliefs and desires, and the explanations of the EU (or non-EU) theorist in

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

terms of credences and utilities? This is a pressing issue for anyone who construes decision theory descriptively and adopts the mentalistic interpretation; see Section 6 below.

Second, a proponent of the behaviourist interpretation might argue that for their purposes, it makes no difference whether utilities and credences are psychologically real or not, so there is no need to assume that they are. Economists' primary interest is in choice behaviour and its consequences; the psychological causes of that behaviour are sometimes regarded as irrelevant to their concerns.[13] From this perspective, to interpret decision theory mentalistically is to append to it a gratuitous piece of metaphysics. An argument of this sort is suggested by Friedman and Savage's insistence that the hypothesis of EU maximization can only be falsified by choice data.

This is a coherent response (though reliant on a controversial conception of economic methodology). However at most it shows that for certain intellectual purposes the behaviourist interpretation is all that need be assumed. It does not show that proponents of the mentalistic interpretation are guilty of a gross methodological error, or have violated some canon of reasonable inference. Indeed a main lesson of post-positivist philosophy of science is that in practice, theories rarely achieve close fit to the observed data unless they have the causal structure of the world at least approximately right (cf. Dietrich and List (forthcoming)). There seems no reason why this general moral should not apply to theories of decision-making.

To sum up, if EU theory is construed descriptively, as a theory about peoples' actual preferences or choices, there seems no particular reason to interpret the theory behaviouristically rather than mentalistically, unless one endorses a general instrumentalist attitude towards science that few contemporary philosophers find plausible. On the contrary, to the extent that the theory fits the data, there seems good reason to adopt a realistic attitude to the utilities and credences which the theory posits. The same applies to the 'non-EU theories' that attempt to model departures of actual behaviour from the predictions of EU.

6. BEHAVIOURISM VERSUS MENTALISM ON THE NORMATIVE CONSTRUAL

Construed normatively, EU theory is a theory of rational rather than actual behaviour; again, I focus on Savage's version for concreteness. So construed, the theory claims that on pain of irrationality, an agent's

[13] This traditional view is defended by Gul and Pesendorfer (2008), but opposed by Camerer (2008) among others. These and the other papers in Caplin and Schotter (2010) offer differing perspectives on this issue. See Clarke (2014) and Dietrich and List (forthcoming) for useful discussion.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

choices between acts should satisfy certain axioms, *modulo* which she is representable as an expected utility maximizer. I claim that, in an important sense, the behaviourist interpretation is actually mandatory when the theory is construed normatively.

The key point is a simple one. The fundamental normative requirement of EU theory is a constraint on an agent's *preferences* – that they should conform to the Savage axioms – and not on the agent's credences, utilities or rule for combining them into a choice criterion. As we know, it follows from Savage's theorem that if an agent's preferences satisfy this constraint, then she behaves *as if* maximizing the expected value of a utility function on the outcomes with respect to a credence function on the states of nature. But it is quite wrong to view the normative content of the theory as saying that an agent should maximize expected utility relative to a psychologically real utility and credence function. For that normative requirement, even if we grant that it makes sense, is *logically stronger* than the requirement that the agent's preferences satisfy the axioms, and is no part of EU theory correctly understood.

To see this point, suppose we have some grounds for attributing to an agent a 'psychological' utility function, real-valued and cardinally measurable, over a set of outcomes. (For example, perhaps the agent can introspect this utility function.) Suppose we also have grounds for attributing to the agent a credence function over a set of states of nature that satisfies the laws of probability, again interpreted psychologically. Now consider the normative injunction: 'choose between acts so as to maximize expected utility, relative to your psychological utility and credence function'. If the agent follows this injunction, then her resulting preference relation over acts will satisfy Savage's axioms. *But the converse is not the case*. Even if the agent violates this injunction, it is perfectly possible that her preference relation satisfies Savage's axioms. If so, then one of two things must be true. Either her Savage utility function, as defined by the EU representation theorem, is non-linear with respect to her 'psychological' utility function; or her Savage credence function is not identical to her psychological credence function, or both. In either case, her preferences are perfectly rational (in the sense of conforming to the Savage axioms).

The point can equivalently be put as follows. An agent whose preference relation satisfies the Savage axioms can be represented as if maximizing the expected value of a utility function with respect to a unique credence function. But there is no guarantee that the utility and credence functions of which this is true bear any particular relation to the agent's psychological 'utility' and 'credence' functions, if such

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

things exist.[14] So it is quite wrong to construe EU theory as telling an agent to maximize expected utility with respect to some pre-existing 'credence' and 'utility' functions, that are defined independently of their choice behaviour. Even if such functions exist, and are measurable on the appropriate scales, maximizing expected utility with respect to them is *not* necessary for having preferences that satisfy the Savage axioms, which is the fundamental normative requirement.

It follows that if EU theory is construed normatively rather than descriptively, the behaviourist interpretation cannot simply be ditched. Suppose that, motivated by standard anti-behaviourist considerations, we interpret the utility concept psychologically, e.g. in hedonic terms à la Bentham. Suppose we then interpret EU theory to say that an agent should maximize the expected value of this hedonic utility function, with respect to their subjective beliefs (credences). Then, we immediately run into the two problems that Bernoulli's theory faced: to justify the assumption that this hedonic utility function is cardinally measurable, and to explain why a rational agent should care only about the expected utility of an act – rather than also attending to the variance in utility, for example. The first problem could conceivably be solved, but the second is intractable.

To see why, consider an agent who is simply risk-averse with respect to hedonic utility – they strictly prefer to receive 5 hedonic utils for sure to a fair coin flip on 10 hedonic utils or nothing. Intuitively this is perfectly rational, just as it is perfectly rational to strictly prefer 5 dollars for sure to a fair coin flip on 10 dollars or nothing (cf. Buchak 2013). And to repeat, it is perfectly possible that such an agent exhibit preferences that satisfy the Savage axioms, in which case her utility function defined à la Savage will simply be concave with respect to her hedonic utility function. So the normative injunction to maximize expected 'hedonic' utility is both intuitively unreasonable, in that it seems perfectly sensible to be risk-averse with respect to it, and unsupported by any axiomatic argument.

Note that this argument applies not just to hedonic utility but to any 'psychological' notion of utility at all, for all such notions, however exactly they are defined, cannot be guaranteed to coincide with the utility that comes from the EU representation theorems. So the injunction to maximize the expected value of a psychological utility function is necessarily more demanding than the requirement to have preferences that accord with the Savage axioms. Moreover, without further explanation of how 'psychological' utility relates to preference or choice, it is unclear how one would ever be able to tell whether an agent was obeying the injunction or not.

[14] This point was made by Harsanyi (1977: 286), who wrote that '*introspective* utility functions need not have any simple relationship to the utility functions inferred from people's behaviour'.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

This explains why I say that if EU theory is taken normatively, the behaviourist interpretation is mandatory. If one adopts the mentalistic interpretation, i.e. posits utility and credence functions that are defined independently of preference or choice, and then interprets the theory as enjoining an agent to maximize expected utility with respect to these functions, this in effect takes us 'back to Bernoulli'. It sacrifices the crucial conceptual advance of the von Neumann & Morgenstern and Savage theories: supplying a normative justification for the principle of maximizing expected utility. This justification is only available if one adopts the behaviourist interpretation.

It follows that standard anti-behaviourist considerations in the philosophy of mind gain no traction if our interest is in normative decision theory. The fundamental norm of modern decision theory is a requirement on preference (or choice, on a revealed preference view). Of course one might accept this norm, but also hold, in an anti-behaviouristic spirit, that agents have psychologically real credence and utility functions that explain their preferences or choices; and one might further hold that these credence functions are (or should be) probabilistic. There is nothing wrong with such a combination of views. But crucially, one must not then read EU theory as prescribing maximization of expected utility with respect to these credence and utility functions. This is a more stringent requirement than that preferences should accord with the Savage axioms, and one that modern decision theory, correctly understood, does not recognize.

To summarize: if credence and utility are taken as psychologically real, and defined independently of preference or choice, the normativity of the 'maximize EU principle' receives no support from the Savage axioms. Of course, there might be some other argument, not based on axiomatic conditions on preferences, for why an agent is rationally required to maximize EU with respect to her 'psychological' utility and credence functions. For example, an argument might be made that an agent who does not do this is almost certain in the long-run to end up with less psychological utility than one who does, or will end up committing to choices that her future self would regret, or will be guilty of some other sort of incoherence or mental instability. Such arguments are conceivable, but to my knowledge none has been spelled out in the literature. Were such an argument to succeed, it would be quite different to the argument for EU maximization based on the Savage axioms, as it would involve a different utility concept.

This implies that if one is interested in decision theory construed normatively, as are most philosophers, then one cannot simply eschew the behaviourist interpretation and hold onto the rest of theory intact. And yet this is precisely what many modern philosophers do, i.e. they start their discussion with utility and credence functions, understood psychologically and defined independently of preference or choice,

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

and then interpret decision theory as issuing the normative injunction to maximize expected utility with respect to them. But this is to fundamentally misconstrue the normative content of modern EU theory. I turn now to a critique of recent philosophical work which is guilty of this error.

7. CRITIQUE OF SOME PHILOSOPHICAL WORK

The view I am critiquing is too widespread to document exhaustively. Much of the literature on 'causal decision theory', from Nozick (1969) and Lewis (1981) onwards, proceeds by simply writing down credence and utility functions, with no mention of the idea of deriving them from preferences via a representation theorem (though Joyce (1999) is a notable exception). These credence and utility functions are generally assumed to be psychologically real, or at least to be legitimate idealizations of real psychological states, and to be capable to causally explaining an action's choices. Sometimes this is stated explicitly, as for example in Mellor (2005) quoted above.

Presenting decision theory this way is unexceptional if the aim is descriptive but the discussions in question have a normative focus. (If the aim were descriptive, it would be odd to focus exclusively on EU theory.) Typically, authors appear to interpret decision theory as issuing the normative injunction to maximize EU with respect to these psychological credence and utility functions. But even if we grant that this advice is meaningful, i.e. grant that such psychologically real functions exist and are measurable on appropriate scales, this is a stronger requirement than the true normative injunction of modern EU theory.

Two recent papers that illustrate this are Briggs (2010) and Meacham and Weisberg (2011).[15] Briggs writes: 'it is a platitude among decision theorists that agents should choose their actions so as to maximize expected value … I make absolutely no substantial assumptions about the nature of the good measured by the value function: 'value' may be read hedonically, morally, aesthetically, pragmatically, or in whatever other way suits the reader's fancy' (2010: 2). However, the only notion of 'value' (i.e. utility) of which it is a 'platitude' that agents should maximize its expectation is utility in the sense of von Neumann & Morgenstern and Savage, i.e. the utility function defined by the representation theorems. If 'value' is read as moral value, or aesthetic value, it is emphatically not a platitude that a rational agent should try to maximize expected value. On the contrary, risk aversion with respect to these 'values' is perfectly rational, and is compatible with having preferences that satisfy the EU

[15] I stress that both papers have considerable merit, and in the case of Briggs, the point I am critiquing is incidental to the main arguments of her paper.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

axioms. It is evident that Briggs construes decision theory as telling the agent to maximize expected utility with respect to some independently defined utility function; which as I have argued is a misconception.

The same is true of Meacham and Weisberg's paper, which is entitled 'Can representation theorems provide a foundation for decision theory?' They take 'decision theory' to mean the normative injunction to maximize expected utility with respect to psychological credence and utility functions. Quite reasonably, they then wonder what the basis for this normative injunction could be; they then consider and reject the suggestion that 'representation theorems' supply the answer. But this is to get things backwards. As I have stressed, the true normative injunction of EU theory is 'choose in accordance with certain axioms'; this is equivalent to maximizing expected utility with respect to the credence and utility functions defined by the EU representation theorem. So there is no question of this normative injunction receiving a 'foundation' in something more basic. It is only if one misconstrues EU theory in the way I have described that one might be tempted to ask this question.

Another case worth discussing is Joyce (1999), one of the most sophisticated philosophical works on decision theory. Joyce cannot fairly be accused of the misconstrual I have criticized in other authors, but there is a dialectical tension in his position which is related. Joyce's position is interesting because his concern is with normative issues, and he endorses the methodology of seeking representation theorems for EU maximization; but he also argues strongly that 'decision theory must throw off the pragmatist/behaviourist straitjacket that has hindered its progress for the past seventy years' (1999: 254). This combination of views is striking because the 'representationalist' approach to decision theory is intimately bound up with behaviourism. Indeed unless one is a behaviourist, it is hard to see why the orthodox methodology of deriving EU maximization from axioms on preferences would have much appeal. This prompts the question: given his rejection of behaviourism, what function does Joyce think that a representation theorem actually serves?

Joyce explicitly addresses this question. He argues that an EU representation theorem achieves two things: first to 'help us understand what the global mandate to maximize expected utility demands at the level of individual preferences'; and second to 'make it possible for proponents of expected utility maximization to rest their case on the plausibility of the local axioms rather than the expected utility principle itself' (1999: 82). (These correspond to the right-to-left and left-to-right directions of Savage's theorem respectively.) However the second of these achievements only makes sense on a behaviouristic view. For the 'plausibility of the axioms' only supplies a justification for expected utility maximization if by an agent's 'utility function' we mean the function that we get out of the representation theorem. On a mentalistic view of utility,

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

the injunction to maximize expected utility is strictly stronger than the injunction to have preferences that obey the Savage axioms; so proponents of expected utility who favour a mentalistic interpretation cannot 'rest their case on the plausibility of the local axioms', as Joyce suggests.

What about the first of the two achievements that Joyce credits to the EU representation theorem? This applies equally on a mentalistic or a behaviouristic view. On either, it is quite true that a representation theorem shows us what the requirement to maximize expected utility implies for an agent's preference relation. This in turn teaches us how we could conclusively show that an agent is *failing* to maximize expected utility – by showing that their preferences don't satisfy the axioms. This applies for any notion of utility, mentalistic or otherwise. But what achieves this is the mathematically trivial, right-to-left part of the representation theorem – which says that *if* an agent has a utility function and a probability function relative to which they maximize expected utility, then their resulting preferences will satisfy the axioms. The real interest in a representation theorem is the converse result that satisfying the axioms is *sufficient* for the agent to be representable as an EU maximizer, which is non-trivial.[16]

This suggests that Joyce's attempt to reconcile the 'representationalist' methodology of modern decision theory with his rejection of behaviourism does not succeed. The second of the two achievements that he credits to a representation theorem only makes sense on a behaviouristic view; while the first provides no reason to find axioms on preferences that suffice for, in addition to being required by, EU maximization. But the search for such axioms is at the heart of most modern decision-theoretic work, include Joyce's own.

Finally, I want to speculate on why so many philosophers have interpreted EU theory in the way I have criticized, i.e. as prescribing maximization of expected utility relative to 'psychological' utility and probability functions. Part of the answer, I suspect, is the tendency to regard decision theory as a formalization of belief-desire psychology, and the maximize EU principle as the quantitative counterpart of the Humean belief-desire law. Typical formulations of the latter read 'if an agent desires x, and believes that doing y is the best way of bringing about x, then they will do y, *ceteris paribus*'. Intuitively this 'law' has *some* connection with decision theory, given that credences and utilities are naturally regarded as 'graded' beliefs and desires. If one thinks that the belief-desire law has normative appeal, and if one interprets beliefs and desires non-behaviouristically, as most philosophers do, and if also one

[16] This point is made neatly by Dekel and Lipman (2010), who argue that some of the rationales often given for seeking representation theorems only require the trivial half of the theorem.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

thinks that the maximize EU principle is the natural formalization of the belief-desire law, then one will be led to interpret EU theory in the way criticized above.

However it is far from obvious that the EU principle *is* the uniquely correct formalization of the belief-desire law. To see this, consider an agent whose preferences violate the Savage axioms, and so who cannot be represented as maximizing EU on any notion of utility, mentalistic or otherwise. For concreteness, suppose that the agent has the well-known Allais preferences, so violates Savage's sure-thing principle, but satisfies the axioms of rank-dependent utility theory.[17] Does it really follow that such an agent is in breach of something like the Humean belief-desire law, i.e. that they are failing to choose the action that, by their own lights, will bring them the outcome they most want? It is not at all clear that this is so.[18] At the very least, the point would need careful argumentation, and some way of explicitly translating the language of beliefs and desires into that of credences and utilities. The loose conceptual connection between belief-desire psychology and decision theory is quite inadequate, on its own, to warrant treating the EU principle as a norm that applies relative to antecedently given probability and utility functions in the manner criticized above.

8. CONCLUSION

The pioneers of 20th century decision theory were adamant in their insistence on the behaviouristic interpretation of their theory, an attitude that is still prevalent among contemporary economists. However among many philosophers, the usual interpretation of decision theory is mentalistic, closer in many ways to the original theory of Bernoulli. I have argued that the correct interpretation depends on whether we want to use decision theory for descriptive or normative ends. Construed descriptively, as a theory of actual choice, there seems no reason not to interpret credence and utility as psychologically real, at least to the extent that the theory fits the data. But construed normatively, as a theory

[17] Rank-dependent utility theory was devised by Quiggin (1982) and Schmeidler (1989), for risk and uncertainty respectively (though under different names). See Wakker (2010) for good discussion.

[18] Here I am indebted to Christopher Clarke (2012) who argues persuasively that the true decision-theoretic analogue of the Humean belief-desire law is not the EU principle but rather the weaker principle of stochastic dominance, which says that if prospect x 'stochastically dominates' prospect y, then the agent should prefer x. Informally, this means that for any utility level, the probability of getting an outcome with at least that utility level is greater if one chooses x rather than y. Importantly, while the principle of stochastic dominance is satisfied by EU theory, it is also satisfied by most extant alternatives to EU theory, including rank-dependent utility theory and cumulative prospect theory. See Wakker (2010) for useful discussion.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

of ideally rational choice, matters are different. For the fundamental normative requirement in EU theory is on preferences; if one attempts to marry a mentalistic construal of utility and credence with the maximize EU norm, one produces a normative injunction that is strictly stronger than the requirement that preferences should conform to the theory's axioms, and that there is no particular reason to obey. This crucial point appears to have gone unnoticed in much of the literature.

ACKNOWLEDGEMENTS

Thanks to Christopher Clarke, Ken Binmore, Philippe Mongin, Richard Pettigrew, Jason Konek, Christian List, John Weymark, Richard Bradley and to two anonymous referees for comments and discussion. This work was supported by the European Research Council Seventh Framework Program (FP7/2007-2013), ERC Grant agreement no. 295449.

REFERENCES

Arrow, K. 1951. *Social Choice and Individual Values*. New York, NY: John Wiley and Sons.

Bermudez, J. L. 2009. *Decision Theory and Rationality*. Oxford: Oxford University Press.

Bernoulli, D. 1738. Exposition of a new theory on the measurement of risk. Reprinted in *Econometrica* 22: 23–26, 1954.

Binmore, K. 2008. *Rational Decisions*. Princeton, NJ: Princeton University Press.

Bradley, R. 2004. Ramsey's representation theorem. *Dialectica* 58: 483–497.

Briggs, R. 2010. Decision-theoretic paradoxes as voting paradoxes. *Philosophical Review* 119: 1–30.

Broome, J. 1991. *Weighing Goods*. Oxford: Blackwell.

Buchak, L. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Camerer, C. 2008. The case for mindful economics. In *The Foundations of Positive and Normative Economics*, ed. A. Caplin and A. Schotter, 43–69. Oxford: Oxford University Press.

Caplin, A. and A. Schotter. 2008. *The Foundations of Positive and Normative Economics*. Oxford: Oxford University Press.

Christensen, D. 2001. Preference-based arguments for probabilism. *Philosophy of Science* 68: 356–376.

Clarke, C. 2012. *The Role of Psychology in Economics*. Unpublished PhD dissertation, submitted to the University of Bristol.

Clarke, C. 2014. Neuroeconomics and confirmation theory. *Philosophy of Science* 81: 195–215.

Dekel, E. and B. Lipman. 2010. How (not) to do decision theory. *Annual Review of Economics* 2: 257–282.

Dietrich, F. and C. List. Forthcoming. Mentalism versus behaviourism in economics: a philosophy-of-science approach. *Economics and Philosophy*.

Eriksson, L. and A. Hájek. 2007. What are degrees of belief? *Studia Logica* 86: 183–217.

Friedman, M. and L.J. Savage. 1948. Utility analysis of choices involving risk. *Journal of Political Economy* 56: 279–304.

Gilboa, I. 2009. *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.

Glimcher, P. W. 2003. *Decisions, Uncertainty and the Brain*. Cambridge, MA: MIT Press.

Glimcher, P. W., M. C. Dorris and H. M. Bayer. 2005. Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behaviour* 52: 213–256.

Gul, F. and W. Pesendorfer. 2008. The case for mindless economics. In *The Foundations of Positive and Normative Economics*, ed. A. Caplin and A. Schotter, 3–42. Oxford: Oxford University Press.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press

Hampton, J. 1994. The failure of expected utility theory as a theory of reason. *Economics and Philosophy* 10: 195–242.
Hansson, B. 1988. Risk aversion as a problem of conjoint measurement. In *Decision, Probability and Utility*, ed. P. Gärdenfors and N. E. Sahlin, 136–158. Cambridge: Cambridge University Press.
Harsanyi, J. C. 1977. On the rationale of the Bayesian approach: comments on Professor Watkins' paper. In *Foundational Problems in the Special Sciences*, ed. R. E. Butts and J. Hintikka, 381–392. Dordrecht: Reidel.
Jeffrey, R. 1990. *The Logic of Decision*, 3rd edition. Chicago, IL: Chicago University Press.
Joyce, J. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.
Lewis, D. 1974. Radical interpretation. *Synthese* 23: 331–344.
Lewis, D. 1981. Causal decision theory. *Australasian Journal of Philosophy* 59: 5–30.
Luce, R. D. and H. Raiffa. 1957. *Games and Decisions*. New York, NY: Dover.
Machina, M. J. 1987. Choice under uncertainty: problems solved and unsolved. *Journal of Economic Perspectives* 1: 121–154.
Marschak, J. 1950. Rational behavior, uncertain prospects, and measurable utility. *Econometrica* 18: 111–141.
Mas-Collel, A., M. D. Whinston and J. R. Green. 1995. *Microeconomic Theory*. Oxford: Oxford University Press.
Meacham, C. J. G. and J. Weisberg. 2011. Representation theorems and the foundations of decision theory. *Australasian Journal of Philosophy* 89: 641–663.
Mellor, D. H. 2005. What does subjective decision theory tell us? In *Ramsey's Legacy*, ed. H. Lillehammer and D. H. Mellor, 137–148. Oxford: Oxford University Press.
Nozick, R. 1969. Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel*, ed. N. Rescher, 114–146. Dordrecht: Reidel.
Okasha, S. 2011. Optimal choice in the face of risk: decision theory meets evolution. *Philosophy of Science* 78: 83–104.
Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3: 323–343.
Ramsey, F. P. 1931. Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*, ed. R. Braithwaite, 156–198. London: Kegan Paul.
Savage, L. J. 1954. *The Foundations of Statistics*. New York, NY: Dover.
Schmeidler, D. 1989. Subjective probability and expected utility without additivity. *Econometrica* 57: 571–587.
Sen, A. 1971. Choice functions and revealed preference. *Review of Economic Studies* 38: 307–317.
Starmer, C. 2000. Developments in non-expected utility theory. *Journal of Economic Literature* 38: 332–382.
von Neumann, J. and O. Morgenstern. 1944. *Theory of Games and Economic Behaviour*. Princeton, NJ: Princeton University Press.
Wakker, P. 2010. *Prospect Theory*. Cambridge: Cambridge University Press.

BIOGRAPHICAL INFORMATION

Samir Okasha is Professor of Philosophy of Science at the University of Bristol. He is the author of *Evolution and the Levels of Selection* (Oxford University Press, 2006) and *Philosophy of Science: A Very Short Introduction* (Oxford University Press, 2002). He is currently the Principal Investigator on a five year project funded by the European Research Council entitled *Darwinism and the Theory of Rational Choice*.

https://doi.org/10.1017/S0266267115000346 Published online by Cambridge University Press