Rejecting Representationalism

Christopher J. G. Meacham
Department of Philosophy
University of Massachusetts, Amherst

Jonathan Weisberg
Department of Philosophy
University of Toronto

Abstract Representation theorems are often taken to provide the foundations for decision theory. First, they are taken to characterize degrees of belief and utilities. Second, they are taken to justify two fundamental rules of rationality: that we should have probabilistic degrees of belief and that we should act as expected utility maximizers. We argue that representation theorems cannot serve either of these foundational purposes, and that recent attempts to defend the foundational importance of representation theorems are unsuccessful. As a result, we should reject these claims, and lay the foundations of decision theory on firmer ground.

1 Introduction

Decision theory begins with the notions of degree of belief and utility, and uses them to rank available options by their expected utilities. This framework has both a descriptive and a normative interpretation. On the descriptive interpretation, decision theory tells us that agents have probabilistic degrees of belief and act so as to maximize expected utility. This is to understand decision theory as psychology, as a theory that characterizes the structure of belief and desire, and which predicts and explains how agents behave. On the normative interpretation, decision theory tells us that agents ought to have degrees of belief that satisfy the probability axioms and ought to act so as to maximize their expected utility.[1] This is to understand decision theory as providing advice, or as providing criteria of rational belief and choice.

Both interpretations of decision theory face foundational challenges. The first challenge is to make sense of the notions of degree of belief and utility: we might be skeptical that there really are such things as belief

1 It's natural to take both descriptive and normative decision theory to include claims about the structure of utilities as well; i.e., to include the claims that agents do or ought to have utilities that are cardinal and commensurable. Since this matter has received little discussion in the literature, we won't explicitly discuss these claims in the text. But our discussion will apply to these claims as well.

and desire, that they can be meaningfully assigned numerical values in the way decision theory presupposes, or that they are related to preference by expected utility. A modest response to this challenge is to accept that these are theoretical notions derived from folk-psychological notions like confidence and desire, and that serious empirical work in psychology will need to be done before we can provide a more concrete account of what they are, how they behave, and indeed, before we can be sure that they are viable notions at all.[2] But a number of people have been dissatisfied with this modest response, and have looked to *representation theorems* to provide precise and general characterizations of these notions.[3] We'll call this immodest response *characterizational representationalism*, or CR.

Making sense of the notions of degree of belief and utility is a challenge for both the descriptive and normative interpretations of decision theory, but normative decision theory faces an additional challenge: it needs to justify its claims that degrees of belief ought to satisfy the probability axioms (*probabilism*) and that agents ought to act so as to maximize expected utility (*expected utility maximization*). A modest response to this challenge is to treat these normative claims like those we find in ethics, and to justify them in a similar way. That is, we may appeal to pros and cons of the normative assumption, discern our considered intuitions about how agents ought to behave in a variety of situations, assess how these intuitions mesh with the normative claim under evaluation, and so on. But again, a number of people have been dissatisfied with this response, and have looked to representation theorems to provide a more decisive and rigorous defense of these claims. We'll call this immodest response *normative representationalism*, or NR.

We think that characterizational representationalism (CR) and normative representationalism (NR) provide poor foundations for decision theory. We will argue that both CR and NR are implausible. The problems facing CR and NR are closely related. CR is problematic because it is in tension with psychology, both folk and empirical. NR is problematic because it implicitly depends on psychological claims similar to those made by CR. While the failure of CR has been recognized by some, especially in the psychological literature, the resulting failure of NR has not been appreciated. Our aim

2 See, for example, (Churchland 1981), (Stich 1983), and (Churchland 1986) for some who have argued that attitudes of belief and desire aren't viable scientific notions.

3 More specifically, we mean theorems that guarantee relatively unique expected utility representations of preference orderings, e.g. the theorems of (Ramsey 1926), (Savage 1954), and (Jeffrey 1983). There are representation theorems outside of expected utility theory, those of (Kahneman & Tversky 1979) and (Wakker & Tversky 1993) for example, which are not our topic of discussion here.

in this paper is to present the case against CR, and to show how similar problems undercut NR. By disposing of CR and NR, we hope to clear the way for a firmer foundation for decision theory.

We will proceed as follows. In section 2 we evaluate several versions of CR. In each case we argue that the characterization of degrees of belief and utilities provided is either inaccurate or uninteresting. In section 3 we show how CR's failure undercuts the standard argument for NR. In section 4 we examine some recent attempts to defend NR without presupposing CR, addressing the arguments of David Christensen (2001; 2004) and Lyle Zynda (2000). We summarize our conclusions in section 5.

2 Characterizational Representationalism

A typical representation theorem says that if a subject's preferences[4] satisfy certain constraints, then there is a unique[5] probability function and utility function whose expected utility ranking coincides with the subject's preferences:

Typical Representation Theorem: If an agent's preferences obey constraints C, then these preferences can be represented as resulting from a unique utility function u and probability function p by expected utility maximization.

The constraints in C typically include elementary requirements like transitivity as well as more substantive and technical assumptions. Different theorems make different assumptions, but our arguments are largely independent of the substance of those assumptions. So we will leave C as an unspecified placeholder for most of the discussion.

The characterizational representationalist wants to use the representation theorem to provide a precise and general characterization of the notions of degree of belief and utility—the graded notions of belief and desire that appear in our folk, descriptive and normative theorizing. This project has

4 There are many different conceptions of preference in play. We might think of preference in terms of dispositions to make choices (Savage 1954) or dispositions to welcome information (Jeffrey 1983). Alternatively, we might treat it as a device of behavioral interpretation (Maher 1993) or even as a basic propositional attitude (Joyce 1999). To avoid hanging our case on one or another of these views, we will leave it open which conception of preference is to be preferred. Our arguments are intended to apply across the board.

5 Strictly speaking, perfect uniqueness is not guaranteed. Savage's representation theorem yields a unique probability function and a utility function unique up to linear transformations, and Jeffrey's representation theorem does not guarantee a unique probability function. To simplify the discussion, however, we will speak as if perfect uniqueness is guaranteed.

several motivations: it allows us to work with decision theory without having to rely on vague and imprecise folk notions, it helps to provide empirical respectability for degrees of belief and utilities by characterizing them in terms of preferences, and it justifies attributions of numerical degrees of belief and utilities.[6]

But in order to employ the theorem to characterize our degrees of belief and utilities, some further assumptions are needed to bridge the gap between the p and u functions that appear in the theorem and our actual degrees of belief and utilities. And closing this gap is not as straightforward as it might appear to be.

First, the theorem only yields the desired p and u functions if the agent's preferences satisfy certain conditions: C. If the theorem is to bear on actual people, something further needs to be said about the status of these preference constraints. Are we assuming that all people's preferences always satisfy these constraints? Just some people some of the time? Or are we only providing a characterization of degrees of belief and utilities for agents who do satisfy these constraints, whether or not actual people are such agents?

Second, for agents who do satisfy the required preference constraints, the representation theorem entails that they can be *represented as* expected utility maximizers with probabilistic degrees of belief. But this gives us no reason to think that such agents *are* expected utility maximizers with probabilistic degrees of belief. Compare: the bank machine at my local bank can be represented as a human bank teller with a tireless work ethic. But that does not entail that it is. Representation is one thing, reality is another, and by itself, the theorem gives us no reason to think that the expected utility maximizing representation of an agent's preferences describes her actual state of mind.[7]

There are, after all, arbitrarily many non-expected-utility representations of those same preferences. An agent with badly incoherent degrees of belief,

6 It is questionable just how empirically unproblematic preferences are, and whether they are significantly less problematic than degrees of belief. Nevertheless, many characterizational representationalists do seem to have held that they are significantly less problematic. Ramsey, for example, clearly had some such view in mind when he presented his representation theorem. (Ramsey 1926: p. 166)

7 One could take an anti-realist view of degrees of belief and utilities, and reject any demand for reasons to think the representation is true of any reality. But anti-realists agree with realists that the postulation and description of theoretical entities is beholden to empirical research, even if they disagree about the why and how. Anti-realists can thus understand our arguments against CR, which are largely based on empirical data and theorizing, as showing that CR is empirically unjustified in whatever way she thinks scientific claims about theoretical constructs are ordinarily justified.

who often acts contrary to expected utility maximization, can have preferences that satisfy the theorem's constraints. For example, suppose Holmes is an expected utility maximizer whose preferences satisfy the required constraints. The theorem will treat him in exactly the same way if he replaces his current degrees of belief and utilities with Moriarty's and yet retains his original preferences, even though he is no longer an expected utility maximizer. Without additional assumptions, the theorem allows degrees of belief and utilities to float free of preferences. All the theorem shows is that Holmes's preferences correspond to those of a possible expected utility maximizer. But the existence of such a representation tells us little about Holmes's actual epistemic or evaluative state, or his actual decision making algorithm.[8]

To bring the theorem to bear on our degrees of belief and utilities, the characterizational representationalist needs to adopt some further assumptions that tie the functions that appear in the representation theorem to the degrees of belief and utilities of actual people. In what follows, we'll present a dilemma for CR. The assumptions the characterizational representationalist employs must be either empirical or non-empirical. If they are empirical, we'll argue that the resulting view is unjustified. If they are non-empirical, we'll argue that the resulting view is uninteresting.

Each horn of the dilemma allows for a number of different assumptions, leading to a number of different views. Indeed, this is one of the difficulties with providing a decisive refutation of CR: there are so many potential positions that it is impossible to address them all individually. So what we'll do is look at several views which are representative of the empirical and non-empirical approaches, respectively. Then we'll use what we've learned to assess the general problems facing CR.

2.1 Empirical Views

We'll first look at views that understand CR as an empirical conjecture. We'll begin with the most straightforward approach, the *actual representation* view. This view holds that actual people typically satisfy the required preference constraints, and that the probability and utility functions the theorem assigns to them describe their actual degrees of belief and utilities. We'll then consider a less ambitious stance, the *approximate representation* view. This

8 A further problem along these lines is that the standard representation theorems provide representations in terms of state-independent utilities, and as such, seem unlikely to correctly represent agents like ourselves. For a discussion of state-independent utilities and related issues, see (Schervish et al. 1990) and (Howson & Urbach 2005: pp. 58-9).

view holds only that the preferences of typical people approximately satisfy the desired constraints, and that some probability and utility functions the theorem assigns approximate their degrees of belief and utilities. The third view we'll examine is the *modal representation* view, which abandons any claims about actual people's preferences. This position holds only that if a person's preferences *were* to satisfy certain constraints, then the resulting probability and utility functions would describe her degrees of belief and utilities.

2.1.1 The Actual Representation View

The actual representation view takes the p and u functions assigned by the representation theorem to be accurate descriptions of the degrees of belief and utilities of actual people. The view requires two empirical claims: (i) people's preferences satisfy the constraints in C, and (ii) the p and u functions the representation theorem assigns for such people correctly describe their degrees of belief and utilities. But these claims conflict with our folk conceptions of belief, desire, and preference, and there is a long tradition of formal psychological research that tells against them. To bring out these conflicts, let us consider four consequences that the actual representation view is committed to.

1. Preference determines belief and utility.

According to the actual representation view, there is a surprisingly strong connection between preference, belief, and utility: your degrees of belief and utilities are fully determined by your preferences. But such a tight connection does not sit well with our folk conceptions of belief, desire, and preference.

It is plausible that *some* relations tend to hold between belief, desire, and preference. After all, we often explain someone's behavior by appealing to her degrees of belief and utilities ("she bet on the Broncos because she was confident they would win"). But, as Christensen (2001) argues, belief and desire have important connections to a host of mental states besides preference, like anger and fear. These emotions can cause beliefs, be caused by beliefs, and even interfere with belief's effect on preference and action; knowing full well that cockroaches are harmless, many people would sooner leave the room than brush a cockroach off their desk. Given that beliefs have connections to so many mental states besides preference — emotions, perception, memory, and more — it is oddly arbitrary to take just one of those connections as paramount. With all the pushes and pulls that beliefs and desires get entangled in, a rigid and straightforward connection between degree of belief, utility, and preference seems unlikely.

2. We are expected utility maximizers.

If our degrees of belief and utilities are identified with the p and u functions that appear in the theorem's expected utility maximizing representation of our preferences, then we are expected utility maximizers. But it seems like we violate expected utility maximization all the time.

The most classic examples of non-maximizing behavior observable from the armchair are cases of akrasia, or "weakness of will". We naturally describe these cases as ones where we act contrary to what we know is in our best interest because we succumb to some more immediate impulse or temptation. And we tend to describe our own behavior in these cases as irrational precisely because we know that, by giving in to temptation, we are not acting in the way most likely to achieve what we want.[9]

Other, less dramatic armchair observations of expected utility violations are common. Some people will drive three miles further to buy gasoline at a station where it costs $0.10 cheaper per gallon, even though the extra gas they use costs more than the amount they save. And many of us will drive across town to shop at a grocery store where our groceries will be $40 instead of $50, but won't drive across town to shop at a computer store where the

9 Expected utility accounts of akrasia have been offered. Jeffrey (1974; 1983), for example, suggests that akrasia be understood as a tension between first- and second-order preferences. The dieter who indulges a craving prefers having a treat to not, but simultaneously prefers not having that first-order preference.

Nevertheless, cases of akrasia put serious strain on the expected utility hypothesis. On a story like Jeffrey's, the subject places greater value, all things considered, on indulging than not indulging. This alone is not terribly plausible: the indulgent dieter does not do what is in his best interest all things considered, he just gives in to momentary temptation. Because a story like Jeffrey's must say that we always maximize expected utility, it forces us to incorporate into our conception of utility every impulse, temptation, or emotion that proves strong enough to motivate action, no matter how fleeting. Such stories distort our conception of utility to the point that it becomes uninteresting. Utility no longer captures what is of value to the agent; instead it combines all possible motivational factors, weighing them solely according to their causal efficacy.

Also, Jeffrey's story cannot account for our sense that the indulgent dieter's action is criticizable. Because the akratic agent always maximizes expected utility, he does not do anything wrong. Jeffrey will try to accommodate us by saying that our dissatisfaction with the subject's situation stems from an appreciation for his unsatisfied second-order preference, to have abstinent first-order preferences. Still, according to Jeffrey, he does not fail to serve his interests when he does not adjust his first-order preferences to satisfy his second-order preference, because adjusting his first-order preferences is not an option for him. He can take long-term steps to adjust such preferences in the future, but he does not have immediate control over his preferences. But if adjusting his first-order preference is beyond his control, then he does not fail to do what is in his best interest; what is in his best interest fails to happen to him. So he is still not criticizable.

computer we want is $2,285 instead of $2,295. It is hard to tell a plausible story that reconciles this behavior with expected utility maximization.

It is not impossible to tell some such story, of course. Perhaps the subject gets so much satisfaction out of finding a good bargain, or loathes paying the middle-man's jacked-up prices so fiercely, that these behaviors really do maximize her expected utility. One can insist that we really do act so as to maximize expected utility, it is just that our utilities are altered by momentary temptations or impulses. But if utility includes all the impulses, biases, habits, etc. that we ordinarily think of as *getting in the way of* the rational pursuit of what we desire, then the notion of utility becomes much less interesting. On this conception, utility is just the residue of preference once belief is factored out, which trivializes the claim that we do or ought to maximize expected utility.

Moreover, these stories are of little help to the actual representation view. The actual view makes the empirical claim that the u function the representation theorem assigns correctly describes our utilities — the graded notion of desire that appears in our folk, descriptive and normative theorizing. But the notion of utility being considered doesn't line up with the notion we use in our theorizing. Both folk and empirical psychology acknowledge that biases, impulses, heuristics, and more interfere with the formation of preferences based on beliefs and desires. We do not simply chalk up apparently irrational behavior to a highly unstable utility function; instead we try to explain it by citing interfering factors. Similarly, in the normative domain, we do not write-off apparently irrational behavior as symptomatic of an idiosyncratic utility function; we criticize it, and try to correct our behavior in the future.[10]

In addition to these armchair observations of non-maximizing behavior, there is a good deal of experimental evidence that we are not expected utility maximizers. Our tendency to violate expected utility maximization is so pervasive and well recognized that explaining why and how we do it is big business for psychologists. Kahneman and Tversky (Kahneman & Tversky 1979), for example, found reliable violations of expected utility maximization in Allais-type decision problems. The choice-patterns uncovered by these cases are at odds with expected utility maximization, but are explained by Kahneman and Tversky's Nobel Prize-winning Prospect Theory. People also habitually violate expected utility maximization by violating stochas-

10 Likewise, one can insist that we really do act so as to maximize expected utility, and accommodate the apparent discrepancies by adopting a non-standard notion of belief. But again, this move is of little help to the actual view. The notion of belief being proposed is just the residue left over when desire is subtracted from preference, and this isn't the notion of belief the actual view is concerned with.

tic dominance (Birnbaum & McIntosh 1996; Birnbaum & Navarette 1998). Again, these results are at odds with the expected utility maximization, but are explained by the RAM model (Birnbaum & Navarette 1998) and the TAX model (Birnbaum & Chavez 1997). Of course, no existing theory fits all of the evidence perfectly. Nonetheless, these results, among many others, suggest that we are not expected utility maximizers, but instead follow decision rules that are better modeled by other accounts, such as cumulative prospect theory (Tversky & Kahneman 1992), rank-dependent utility theories (Quiggin 1982), and weighted utility theories (Fishburn 1983).

We cannot pretend to adjudicate the dispute between expected utility theory and its opponents here. Decisive refutations of empirical hypotheses are hard to come by and this case is no exception.[11] What we have done is to survey the most prominent armchair and formal data that tell against the expected utility hypothesis, and responded to some of the most popular replies that we have heard. We have also appealed to expert authority: contemporary research in psychology regards the expected utility hypothesis as a failure that has been supplanted by more successful alternatives. In our view, the record cannot justify standing by the expected utility hypothesis.

3. Our preferences satisfy C.

The actual representation view holds that our preferences satisfy the constraints in *C*. But empirical evidence suggests that our preferences reliably violate those constraints. Kahneman and Tversky (1979) famously found that subjects consistently fall prey to versions of Allais' (1979) paradox, thereby violating Savage's (1954) Independence Axiom. Lichtenstein and Slovic's (1971; 1973) discovery of robust preference reversal suggests that people habitually violate the Transitivity Axiom as well.

4. Normative decision theory becomes uninteresting.

So far we have examined worries about the plausibility of actual representationalism. But problems of a different kind arise when we consider the consequences for normative decision theory. In particular, adopting these characterizations robs normative decision theory of much of its interest. Normative decision theory only applies to agents who have degrees of belief and utilities. But if actual representationalism is true, then agents who have degrees of belief and utilities are automatically probabilistic expected utility maximizers, and the further question of whether they ought to be probabilistic expected utility maximizers is merely academic.

To sum up, the actual representation view is not a plausible empirical

11 For some defenses of expected utility as a descriptive hypothesis and/or criticisms of alternatives, see (Watt 2002), (LeRoy 2003) and (Levy & Levy 2002). For some responses to these defenses, see (Rabin & Thaler 2001), (Monti et al. 2005) and (Wakker 2003).

hypothesis. It is at odds with our folk conceptions of belief, desire, and preference, as well as a substantial body of psychological research. It also threatens to trivialize normative decision theory. As such, one should feel uneasy about resting the foundations of decision theory on the actual representation view. For these reasons, many proponents of CR have rejected the actual view in favor of weaker views, like the approximate or modal representation views. We turn to these now.

2.1.2 The Approximate Representation View

The approximate representation view attempts to avoid the problems that face the actual representation view by weakening its empirical claims. The approximate representation view claims only that (i) the preferences of typical people approximately satisfy the constraints in C, and (ii) the p and u functions that the theorem assigns to nearby sets of preferences will approximate their actual degrees of belief and utilities. The move to approximation avoids the empirical conflicts that face the actual view, since discrepancies can be ascribed to the gap between the approximation and reality. It also avoids trivializing normative decision theory, since maximizing expected utility is no longer an automatic consequence of having degrees of belief and utilities.

Trying to come up with an idealized model that successfully approximates actual phenomena is standard scientific practice. But if we understand "approximately" loosely enough, the approximate view is trivially true. In order for it to be true in an interesting way, it needs to be the case that there isn't another theory at the same level of abstraction that provides a better characterization of the relevant phenomena. And a problem for this view is that there *are* other theories that are more empirically successful. For example, cumulative prospect theory, rank-dependent utility theories and weighted utility theories all provide better approximations of how our degrees of belief, utilities and preferences are related. Though couched at the same level of abstraction, these theories make more accurate predictions about human preferences than expected utility theory.[12] (Kahneman & Tversky 1979; Tversky & Kahneman 1992; Knetsch 1989; Birnbaum & McIntosh 1996; Birnbaum & Chavez 1997; Birnbaum & Navarette 1998; Prelec 2000; Tuthill & Frechette 2002)

A second, methodological problem for the approximate view is that its weakness undermines one of CR's central motivations. The goal was to

12 Though, again, there are dissenting opinions; see footnote 11.

provide a *precise* and general characterization of our degrees of belief and utilities, so we don't have to rely on vague folk notions in order to make sense of decision theory. But the approximate view only makes a rough claim about how our degrees of belief, utilities, and preferences are related. It doesn't provide us with anything like an analysis of degrees of belief and utilities, nor a precise characterization of them via their role in a psychological theory. Indeed, it's hard to see how the approximate view is an advance over folk theory. To the extent that the expected utility model makes novel claims — e.g., that willingness to stake utility increases linearly with degree of belief — the empirical studies suggest that it is wrong. And the rough truths it does manage to capture — e.g., that willingness to stake utility tends to increase with degree of belief — are things we already know from folk psychology.

To sum up, the approximate view avoids the problems the actual view faces by weakening its empirical claims. But even these enervated claims are implausible, and adopting them undermines the rationale for pursuing CR.

2.1.3 The Modal Representation View

The modal representation view is even more conservative than the approximate view. Unlike the actual and approximate views, the modal view does not make claims about our actual preferences. Instead, it holds that if our preferences *were* to satisfy the desired constraints, then the p and u functions the theorem assigns *would* describe our degrees of belief and utilities.

This view succeeds in weakening its empirical claims enough to avoid conflicting with empirical evidence. Our current evidence suggests that human decision making is a complex, multi-variate process that makes use of a range of methods and heuristic shortcuts, and is unlikely to follow any simple rule precisely. Furthermore, it indicates that the preferences that result from this process reliably violate C in a number of ways. Given this, the possibility where our preferences satisfy C is remote; too remote for our current empirical theories to say anything interesting about.

But this doesn't make the modal view empirically trustworthy. After all, there are lots of different ways we could have preferences that satisfy C. We might have probabilistic degrees of belief, but have our beliefs and utilities relate to our preferences via some rule other than expected utility maximization. We might have non-numerical degrees of belief and utilities, and have them relate to our preferences by some non-numerical rule. We might even be one of the many probabilistic expected utility maximizers with C-satisfying preferences who don't line up with the theorem's representation because they have state-dependent utilities, which the theorem does not

allow.[13] Viewed in this light, the modal view is a rather incredible claim: on the basis of no evidence, it makes precise empirical claims about agents of a kind we've never encountered. It seems ill-advised to try to rest the foundations of decision theory on such a claim.

By way of analogy, consider what it would take for beings like us to be perfect utilitarians. Perhaps we would need to be free of bodily desires, because such desires would bias us towards our own concerns. Or perhaps we would need to develop an indomitable will to do what we think is right. Or perhaps we would need to establish a strong bond of friendship with every other sentient being. Now suppose someone claimed that if we were perfect utilitarians, we would be free from bodily desires. We don't have any evidence against this claim. But we don't have any evidence for it either: this possibility is too remote for psychology to say anything interesting about. And given that we can imagine any number of ways to be a perfect utilitarian *with* bodily desires, it would be foolish to rest the foundations of moral psychology on this claim.

A second, methodological problem for the modal view is that, like the approximate view, it undermines a central motivation for adopting CR. The goal was to provide a precise and general characterization of our degrees of belief and utilities, which relieves us from having to rely on vague folk notions in order to make sense of decision theory. While the approximate view failed to be precise, the modal view fails to be general. The modal view only provides a characterization of degrees of belief and utilities for agents who satisfy the required preference constraints. And since agents like us don't satisfy these constraints, this is of little help in working out what our degrees of belief and utilities are, and does little to lessen our reliance on intuitive folk notions when using decision theory.

The modal view avoids empirical disagreement by restricting the scope of its claims to cases we don't have evidence about. While this move allows it to avoid empirical conflicts, it also cuts it off from empirical support. By restricting itself to remote situations, it is of little help in characterizing our degrees of belief and utilities, and does little to diminish our dependence on intuitive folk notions when using decision theory.

13 See (Howson & Urbach 2005: pp. 57-60) for a recent survey of the role of state-dependence assumptions in representation theorems.

2.2 Non-Empirical Views

The empirical views we've looked at are empirically unjustified. Non-empirical views avoid this problem by eschewing empirical claims. The simplest non-empirical view, the *stipulative* view, treats 'degree of belief' and 'utility' as newly defined terms, stipulating that an agent's degrees of belief and utilities are the values of the p and u functions the representation theorem assigns to her. A slightly stronger view, the *explicative view*, makes a further claim: that these stipulatively defined terms explicate the corresponding folk notions of belief and desire, capturing what is good and useful about the folk notions while leaving behind what is imprecise, problematic, or otherwise undesirable.

These non-empirical views avoid many of the empirical worries we've raised by avoiding empirical claims. Moreover, these views provide a precise characterization of degrees of belief and utilities. Nonetheless, we take these views to be unsatisfactory. To avoid homophonic confusions, we will call the stipulated notions *degrees of belief** and *utilities**. Our objections to these views can then be framed as follows.

2.2.1 The Stipulative View

If one simply defines 'degree of belief*' and 'utility*' as the stipulative view does, one is obliged to give some reason to think that these terms are related to the topics of interest. And if degrees of belief* and utilities* are defined in a way which makes the worries we've raised irrelevant, they risk becoming irrelevant themselves.

To see this, consider how the stipulative approach deals with the problems we raised for the actual representation view. One worry was that ordinary people don't appear to be expected utility maximizers with respect to their degrees of belief and utilities. The stipulative view escapes this worry, because it doesn't follow that people are not expected utility maximizers with respect to their degrees of belief* and utilities*. After all, there's no reason to think that degrees of belief* and utilities* correspond to degrees of belief and utilities — the graded notions of belief and desire that appear in our folk, descriptive and normative theorizing. But then why are degrees of belief* and utilities* interesting? If they don't correspond to any of the notions that we ordinarily work with, what good are they?

A second worry was that ordinary people don't appear to satisfy the preference constraints required for the representation theorem to apply. But this isn't evidence that the stipulative view is false, just evidence that people

like us don't have degrees of belief* and utilities*. But then why bother with degrees of belief* and utilities*? Why define a possible mental state that we don't have? And so on.

The stipulative view overcomes the problems that afflict the other approaches by brute force: it ensures by stipulative definition that degrees of belief and utilities are the p and u functions the representation theorem assigns. Of course, one can define theoretical terms however one likes. But, absent any reason to think that the stipulated notions are the ones we use in our descriptive or normative theorizing about belief and decision, this is a hollow victory. Without some further empirical claims, we have no reason to consider degrees of belief* and utilities* relevant to the topic at hand.

2.2.2 The Explicative View

A natural response to these worries is to appeal to the explicative view: degrees of belief* and utilities* are interesting and relevant because they explicate existing notions. They offer precise and rigorous replacements for folk or quasi-formal notions like belief, desire, credence, etc. But the problems we raised for empirical views can be recast as reasons to think the stipulated notions will be poor replacements.

Following Carnap (1950), we can evaluate a proposed explication of a concept according to four criteria: (i) the explication should be simple, (ii) the explication should be precise, (iii) the explication should be fruitful, and (iv) the explication should be similar to the original concept; i.e., it should apply to most of the same cases as the original concept, and should be able to play the same roles. Degrees of belief* and utilities* do well with respect to the first two criteria. But they fail to do well enough with respect to the latter two criteria to be satisfactory explications of our original concepts.

In particular, let us focus on the fourth criterion. Degrees of belief* and utilities* will not apply to most of the same cases as our original notions. Recall how the non-emprical views escape the worries that face the actual representation view: evidence that subjects like us don't satisfy the preference constraints required for the representation theorem to apply isn't evidence that these views are false, just evidence that people like us don't have degrees of belief* and utilities*. But if degrees of belief* and utilities* don't apply to subjects like us, then they will not apply to most of the same cases as the original concepts.

Likewise, degrees of belief* and utilities* fail to play the same roles as the original concepts they're intended to replace. Since we often don't have degrees of belief* and utilities*, in virtue of violating the constraints in C,

they cannot play a useful role in descriptive accounts of our mental states, predictive accounts of our behavior, or prescriptive accounts of what our behavior ought to be.

Even putting these kinds of issues aside, degrees of belief* and utilities* are unable to usefully play many of these roles. For example, degrees of belief* and utilities* are unable to play a useful role in normative decision theory. This is because adopting degrees of belief* and utilities* trivializes normative decision theory. Normative decision theory only applies to agents who have degrees of belief and utilities. But agents who have degrees of belief* and utilities* are automatically probabilistic expected utility maximizers with respect to them. So suppose we take the "degrees of belief" and "utilities" that appear in normative decision theory to be degrees of belief* and utilities*. Then it will be true by definition that all agents subject to the norms of decision theory satisfy them.

So adopting degrees of belief* and utilities* fails to provide satisfactory explications of the graded notions of belief and desire that play a role in our folk, descriptive and normative theorizing. And the explicative view is no more successful than the stipulative view in providing a satisfactory account of characterizational representationalism.

2.3 General Considerations

We've looked at the problems facing five kinds of characterizational representationalism. But there are other varieties we haven't considered. We can get a feel for the problems that will face other versions of CR by diagnosing the source of the problems we have encountered.

Every version of CR uses the representation theorem to posit a relation of some kind between preferences on the one hand, and degrees of belief and utilities on the other. Versions of CR differ in the conceptions of the relata they work with, and in the relation taken to hold between those relata. Ultimately though, the characterizational representationalist wants to provide a precise and general characterization of degrees of belief and utilities that will fit into our descriptive and normative theorizing about these notions. The basic problem is that the relation suggested by the representation theorem — expected utility maximization — doesn't fit the relation we need in our descriptive and normative theorizing.

The actual view maintains that the relation does fit, but this conflicts with our empirical evidence. The other views try to avoid empirical conflicts, but each gives up part of what a satisfactory account of CR requires. The approximate view gives up on trying to provide a precise characterization.

The modal view gives up on trying to provide a general characterization. The non-empirical views give up on trying to provide a characterization of the degrees of belief and utilities that we were originally concerned with. And despite these sacrifices, only the non-empirical views avoid worries regarding empirical justification.

It's these tradeoffs that yield the structure of the dilemma between the empirical and non-empirical approaches to CR. At the end of the empirical horn, we have a theory which has the desired properties, but conflicts with empirical evidence. At the end of the non-empirical horn, we have a theory which avoids empirical entanglements, but is disconnected from the topics of interest. And the range of views in-between trade off between these difficulties.

So CR's prospects look bleak. Although different versions of the view can avoid particular problems, the source of these problems remains: the psychological picture at the heart of characterizational representationalism is false.

3 Normative Representationalism

We turn now to the second application of the representation theorem: justifying the normative theses of probabilism and expected utility maximization. How is the representation theorem supposed to provide such a justification? As a first pass, the reasoning goes something like this. Our preferences ought to satisfy *C*, and the representation theorem assures us that if they do then we will be probabilistic expected utility maximizers. So we ought to be probabilistic expected utility maximizers.

As given, this reasoning is invalid. The representation theorem shows that if one's preferences satisfy *C*, then one is representable as a probabilistic expected utility maximizer. But given that our preferences ought to satisfy *C*, this only suggests that we ought to be *representable as* probabilistic expected utility maximizers.[14] It doesn't yield the stronger claim that we ought to *actually be* probabilistic expected utility maximizers. Suppose, for example, that Jones wants nothing more than to drink the glass of aged, single malt whisky in front of him. Given only the weaker claim about representability, there is nothing irrational in him pouring it down the sink untouched, as long as his behavior can be represented as that of some (puritanical) expected utility maximizer.

14 Why only "suggests"? Because there are reasons to think that even this inference is invalid. See the remarks concerning further problems for these kinds of arguments, below.

CR and NR face parallel problems: there is a gap between representation and reality that must be bridged for either view to get what it wants out of the representation theorem. Because of its descriptive aspirations, CR faced an additional challenge, since it had to say something about the case where the theorem does not provide a representation because *C* is not satisfied. But NR is not concerned with what your degrees of belief and utilities are when *C* is not satisfied, since it is only concerned with the case where you do what you ought to do, i.e. where you satisfy *C*. So NR only needs it to be the case that the canonical representation is the correct one when it exists. Following Christensen (2001; 2004), we'll label this assumption:

Representation Accuracy: If an agent's preferences can be represented as resulting from a unique utility function u and probability function p by expected utility maximization, then p and u describe the agent's degrees of belief and utilities.

Representation Accuracy in hand, the argument then goes as follows: we ought to satisfy *C*, and the representation theorem guarantees that if we do, we can be represented as probabilistic expected utility maximizers. Given Representation Accuracy, it then follows that if we satisfy *C* we will be probabilistic expected utility maximizers, so we ought to be probabilistic expected utility maximizers.

Adding Representation Accuracy eliminates the lacuna in the argument, but it also betrays the argument's dependence on problematic assumptions regarding the relation between degrees of belief, utilities and preferences. As with CR, the normative representationalist faces a dilemma between empirical and non-empirical understandings of Representation Accuracy. If she understands Representation Accuracy as an empirical claim, it's unjustified; if she understands Representation Accuracy as a non-empirical claim, the normative conclusions of the argument become uninteresting.

First consider the empirical horn of the dilemma. If we take p and u to describe the degrees of belief and utilities we discuss in our normative and descriptive theorizing, then Representation Accuracy becomes similar to the modal representation view, and suffers from similar problems. Like the modal view, it escapes conflicting with our evidence because it confines itself to cases we don't have evidence about. While this allows it to avoid empirical conflicts, it also cuts it off from empirical support. Since there are a number of ways to have preferences with a unique expected utility maximizing representation, and yet not have the psychological states that the representation attributes, the argument's use of Representation Accuracy is unjustified.

On the non-empirical side of the dilemma, the problems parallel those facing the non-empirical take on CR. For example, suppose that Representation Accuracy is true because we define 'degrees of belief' and 'utilities' to be the values delivered by the representation theorem; i.e., degrees of belief* and utilities*. Then we need some reason to care about the corresponding normative requirements, *probabilism** and *expected utility maximization**. To make degrees of belief* and utilities* relevant to epistemology and normative decision theory, these states must be linked to the states that are the topic of our normative theorizing in these domains. And since agents like us generally don't have degrees of belief* and utilities*, it's hard to see how they're relevant. Moreover, these normative requirements are prescriptively useless. Agents who don't have degrees of belief* and utilities* won't be subject to these normative requirements, and agents who do have them will already satisfy these requirements by definition. So although we can repair the argument for NR by adopting a non-empirical understanding of Representation Accuracy, doing so threatens to trivialize its normative conclusions.

This reliance on problematic assumptions regarding the relation between degrees of belief, utilities and preferences is the biggest problem with the argument for NR. But the argument faces other worries as well. Many have questioned whether we should accept the normative preference constraints it requires.[15] Some have noted that the argument appears to conflate instrumental and epistemic rationality.[16] A few have pointed out that the standard representation theorems assume state-independent utilities, and that if the argument for NR is correct then we ought to have state-independent utilities, an implausible claim.[17] A worry that has not been noticed is that the argument may commits a fallacy of deontic logic.[18]

We are sympathetic to some of these concerns. But they are largely

15 Savage's independence axiom has long been challenged on the grounds that it violates intuitions in the Allais paradox (1979) and the Ellsberg paradox (1961). More technical axioms are often accused of making unrealistic demands on real agents, e.g. that the agent have preferences over an infinitely divisible space of possibilities. For some contemporary discussion of the critical history here, and accompanying defenses of the Savage and Jeffrey axioms as normative requirements, see (Maher 1993) and (Joyce 1999).

16 E.g. (Christensen 1996, 2001, 2004) and (Joyce 1998).

17 For a discussion of state-independent utilities and related issues, see (Schervish et al. 1990) and (Howson & Urbach 2005: pp. 58-9).

18 The argument employs the following inference: (i) if you are representable as having probabilistic degrees of belief and maximizing expected utility, then you really are such; (ii) you ought to be so representable; so (iii) you ought to have probabilistic degrees of belief and maximize expected utility. This is an inference of the following form:

independent of the theme we have been developing. Our primary concern is with how the dependence on problematic assumptions regarding the relation between degrees of belief, utilities and preferences undermines a decision theory based on representation theorems. So we will not pursue these other points here.

4 Normative Representationalism Without Representation Accuracy

The standard argument for NR relies on Representation Accuracy, but one can try to defend NR without making this assumption. We'll consider one first-pass attempt to do so, before turning to two more sophisticated attempts that have appeared in the recent literature.

One way to defend the original argument for NR without employing contentious assumptions about the relation between degrees of belief, utilities and preferences is to weaken its conclusion. Although the original argument may not show that we ought to be probabilistic expected utility maximizers, it may show that we ought to act *as if* we were such. So the representation theorem does justify normative "as if" theses similar to probabilism and expected utility maximization.

But this response just gives up the game. If these "as if" theses are what we're after, then the representation theorem is beside the point: all the "as if" theses tell us is that our preferences should satisfy *C*, and we don't need the representation theorem to deduce this — it's one of the assumptions we start out with.[19] Likewise, insofar as we are interested in deciding between alternative theories of epistemic and practical rationality — theories based on sub-additive functions (Shafer 1976) or ranking functions (Spohn 1988), for example — the representation theorem is of little use. The only thing relevant

---

$$\frac{A \rightarrow B \quad O(A)}{O(B)}$$

where $O(\cdot)$ is the deontic 'ought' operator. This argument form is not valid — not in standard deontic logic, and not intuitively. Consider the following instance of it:

$$\frac{\text{If I wear a seatbelt, I'll drive faster.} \quad \text{I ought to wear a seat belt.}}{\text{I ought to drive faster.}}$$

In general, if doing A will lead to your doing B and you ought to do A, it does not follow that you ought to do B.

19 Actually, these "as if" theses will place constraints on our preferences that are slightly weaker than *C*. But these weaker constraints will be entailed by the assumptions we start out with.

to our evaluation is whether these theories provide a decision algorithm consistent with C, and the representation theorem doesn't help us assess that.[20]

We turn now to two recent attempts to defend NR without assuming Representation Accuracy: those of David Christensen (2001; 2004) and Lyle Zynda (2000). Since their proposals focus on the justification of probabilism, the remainder of our discussion focuses on probabilism alone, leaving expected utility maximization to the side.

4.1 Christensen

Christensen (2001; 2004) agrees that the standard argument for NR is unsound, and proposes to repair it by replacing Representation Accuracy with a normative claim. In particular, he proposes to employ:

> **Informed Preference:** If an agent has a higher degree of belief in B than in A, then she ought to prefer the option of getting a desirable prize if B obtains to the option of getting the same prize if A obtains.[21]

Informed Preference is an informal, normative version of Savage's *coherence condition*.[22] Coherence conditions are standard assumptions when proving representation theorems, serving to forge the connection between preferences and degrees of belief. While Savage treated his coherence condition as definitional of 'degree of belief', and others seem to treat it as a necessary connection between degrees of belief and preferences, Christensen's idea is to treat coherence conditions as normative constraints instead. On Christensen's view it is possible to have degrees of belief other than the representational ones. It is just not rational to do so, since you will violate Informed Preference.

20 It is also worth noting that a big part of the representation theorem and its proof become otiose in the argument for "as if" theses, since uniqueness is no longer important. If all we want to show is that, to satisfy C, one must behave as if one were a probabilistic expected utility maximizer, then we just need there to be at least one probabilistic expected utility maximizer for each C-satisfying preference structure. If there are multiple such representations, it is still true that all C-satisfiers behave like probabilistic expected utility maximizers. In a way this is a boon for the argument, since a representation theorem without uniqueness is easier to prove and requires fewer and less controversial assumptions. But the trouble that authors typically go to to ensure uniqueness in their theorems suggests that "as if" theses are not what they are after.

21 Christensen's formulation differs slightly, since it is couched in terms of ideally rational agents, instead of what an agent ought to do.

22 The "coherence" terminology is from Joyce (1999: pp. 89, 130).

Given Informed Preference, Christensen offers the following argument for probabilism: our preferences ought to satisfy C, and our degrees of belief ought to cohere with our preferences as required by Informed Preference. Given that the required preference constraints are satisfied, the representation theorem shows that there is a unique probability function which will satisfy Informed Preference with respect to those preferences. It follows that we ought to have probabilistic degrees of belief, since that is the only way to satisfy all of the constraints we ought to obey.

Unfortunately, this argument is invalid, turning crucially on an ambiguity in typical statements of representation theorems. Representation theorems do not show that there is a unique degree of belief function that coheres with your preferences, and that this function is a probability function. Rather, they show that there is only one *probabilistic* degree of belief function that coheres with your preferences. But there are many non-probabilistic degree of belief functions that will also cohere with your preferences, often in non-expected utility maximizing ways. For example, consider the following function $f : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$:[23]

$$\begin{aligned} f(p) &= \frac{1}{1-p} - 2 \quad \text{if } p \neq 1, \\ &= \infty \qquad\quad \text{otherwise.} \end{aligned}$$

If p is the probability function that coheres with your preferences, then $f(p)$ will cohere with your preferences as well. But $f(p)$ is not a probability function. Indeed, $f(p)$ violates *all* of the probability axioms: it is not additive, positive or bounded.[24] So the argument presupposes that there are no non-probabilistic degree of belief functions that cohere with the preferences of an agent who satisfies C, when there are actually arbitrarily many.[25]

At this point, one might be tempted by the following line of thought. We have shown that one needn't have probabilistic degrees of belief to satisfy all of Christensen's normative constraints, since one might have the degrees of belief given by $f(p)$ rather than those given by p. But one might try to maintain that the difference between p and $f(p)$ is merely notational or conventional. Maybe the apparent difference between p and $f(p)$ is just an

23 Where we take the algebraic operations and ordering relations to be extended over $-\infty, \infty$ in the usual way.

24 Consider a σ-algebra $\mathcal{A}$ over Ω. A probability function $p : \mathcal{A} \to \mathbb{R}$ must satisfy the three probability axioms: (i) $\forall A \in \mathcal{A}$, $p(A) \geq 0$, (ii) $p(\Omega) = 1$, (iii) if $A_1, A_2, \ldots$ are mutually exclusive, then $\cup_i A_i = \sum_i p(A_i)$.

25 In general, any order-preserving map from the unit interval will yield a degree of belief function that coheres with your preferences.

artifact of the particular numerical scheme or scale in which we choose to describe an agent's credences, as when we choose to measure temperature in Celsius rather than Fahrenheit. On this conception, the states of mind described by p and by $f(p)$ would be the same.

However, in order to argue that all the different degree of belief functions that will satisfy Christensen's normative constraints describe the same epistemic state, one must hold a quite coarse conception of degrees of belief. Any numerical degree of belief function that orders propositions the same as p does will cohere with the preferences p coheres with. Since one must be prepared to treat any ordinal equivalent of p as describing the same psychological reality, one must be prepared to sacrifice all cardinal facts about degrees of belief: absolute values, differences, and even ratios of differences, are all unreal on such a view.

Precisely such a view has been endorsed by Zynda (2000). We turn to this view next.

4.2 Zynda

The problem for Christensen's proposal was non-uniqueness; because many non-probabilistic functions cohere with a given set of preferences, Informed Preference does not require our degrees of belief to be probabilistic. In response to this difficulty, Lyle Zynda (2000) proposes to eliminate the non-uniqueness. He suggests that any two representations with the same preference ranking are notational variants, signifying the same psychological state. On Zynda's proposal, the only real features of your beliefs are those shared by every possible representation that coheres with your preferences. Claims like "the agent thinks A is more likely than B" describe something real, whereas claims like "the agent thinks A is .5 more likely than B" are artifacts of the particular numerical representation we choose.

How does Zynda's psychological proposal help normative representationalism? His proposal yields a coarse conception of degrees of belief on which quantitative constraints like the probability axioms don't make sense. Since numerical claims about degrees of belief don't describe anything real, it doesn't make sense anymore to require your beliefs to satisfy things like the probability axioms. But there are qualitative analogues of those axioms that make sense for coarse degrees of belief, de Finetti's axioms:

Normalization Tautologies are more likely than contradictions.

Boundedness A tautology is at least as likely as anything and anything is at least as likely as a contradiction.

Qualitative Additivity If A is more likely than B, and C is incompatible with both, then $A \vee C$ is more likely than $B \vee C$.

For a degree of belief function to cohere with a preference ranking that satisfies the usual constraints, it must rank propositions in accordance with these axioms. Thus we have a qualitative analogue of probabilism, and a representation theorem argument for it. The conclusion is weaker, since de Finetti's qualitative axioms are weaker than the usual quantitative ones. But if Zynda is right that degrees of belief are accordingly coarse, then this is as much as one could ask for. And de Finetti's axioms do have non-trivial content: they rule out a number of existing views about rational belief, such as Shafer's belief functions (Shafer 1976) and certain (non-standard) interpretations of Spohn's ranking functions (Spohn 1988).

That said, let's step back and assess this proposal. While Zynda's argument for the conclusion that our degrees of belief should satisfy de Finetti's axioms avoids non-uniqueness problems, his coarse view of degrees of belief is problematic. On Zynda's account it makes sense to say that an agent regards A as more likely than B, but it does not make sense to say that she thinks A is *much* more likely than B. Nor does it make sense to say that one agent is more confident of A than another agent, since the absolute utility an agent assigns to a proposition can vary as much as we like. What epistemic psychology is left with is just qualitative, agent-relative comparisons like "the agent thinks A is more likely than B." And it doesn't look like this is enough.

Without interpersonal comparisons, for example, it is hard to explain why one person assents to a proposition while another does not. Even restricting our attention to a single agent, we lose a lot if we give up facts about *how much* more likely an agent thinks one thing is than another. You hardly need to deliberate about whether to stake a prize on 2+2=4 or George W. Bush being an alien, but you would need to think a bit harder about whether to stake the same prize on Bush being an alien as opposed to Cheney. Presumably, the difference in deliberation time can be explained (at least partly) by facts about relative magnitudes in your degrees of belief. Magnitudes can also make a difference to preference and choice without there being any qualitative difference in degree of belief. If your degrees of belief in A and B are close enough, you might prefer to stake a desirable prize on A rather than B merely out of habit or some other arbitrary factor, when a greater difference might have led you to the opposite preference. Very likely, the heuristics we use to form preferences are sensitive to facts about magnitude that aren't captured ordinally.

Magnitudes are also important for Bayesian accounts of confirmation. According to the standard Bayesian resolution of the raven paradox, for example, the discovery of a non-black object should raise our degree of belief in the raven hypothesis, but only by very little. So the extra-ordinal structure contained in the standard Bayesian picture of degrees of belief is not idle. Magnitudes encode important features of our degrees of belief, and if we abandon this structure, degrees of belief lose much of their utility.[26]

4.3 The Last Gasp

Perhaps we should stick to a numerical conception of degrees of belief, but give up on the aim of justifying probabilism. Instead, we might settle for the more modest conclusion that our degrees of belief ought to satisfy de Finetti's axioms. Then we can couple Zynda's maneuver with Christensen's appeal to

26 Could Zynda enjoy the fruits of numerical degree of belief talk without committing to the reality of anything more than ordinal structure? Anti-realists about unobservable entities appeal to unobservables in their explanations, despite being agnostic about the existence of those entities. They take a view of explanation on which a satisfactory explanation can appeal to things we do not believe are true. (van Fraassen 1980) So perhaps Zynda can take a similar approach: believe only in the reality of ordinal degrees of belief, but talk the talk of numerical degrees of belief for the purposes of explanation and prediction.

This move is of little help to Zynda in the present context, however. First, the view is unmotivated, since our reasons for being realists about ordinal structure are equally reasons for realism about numerical structure. The claim that belief has ordinal structure is motivated by its explanatory and predictive success, and maybe also by introspection (we feel more certain of some things than others). But the same is true of the claim that degrees of belief have numerical structure. This is what our objections to Zynda's ordinal view show: magnitudes allow us to explain and predict facts about assertion, deliberation time, confirmation, etc. And introspection makes it plausible that some differences in certainty are greater than others: the difference between your certainties in "The sun will rise tomorrow" and "I am Napoleon" is much greater than the difference between your certainties in "I am not a brain in a blue vat" and "I am not a brain in a pink vat". Since the considerations that motivate realism about ordinal structure also motivate realism about numerical structure, a view which endorses realism about one but not the other is theoretically unstable.

Second, the difference between realism and anti-realism about numerical structure is immaterial for the purposes of our discussion. In order to appeal to numerical features of belief in explanations and predictions, the anti-realist must hold that a geniune distinction underlies different numerical representations that are ordinally equivalent. After all, she will take these different numerical descriptions to predict and explain different phenomena. But if the anti-realist can recognize the distinction, then she will be able to make sense of norms that track such distinctions, like NR (where, of course, the anti-realists's version of NR is cashed out in terms of the genuine facts that underlie this distinction). And with respect to the representation theorem argument for NR, the anti-realist will run into the same problems regarding non-uniqueness as the realist.

Informed Preference to set up an argument for de Finetti's axioms. An agent's preferences ought to obey the constraints assumed by the representation theorem, and her degrees of belief ought to cohere with those preferences in the sense of obeying Informed Preference. If she does these things, her degrees of belief will obey de Finetti's axioms. So she ought to obey said axioms. What should we think of this argument?

We think this line of argument is unappealing. For one thing, many of the additional problems about normative representation theorem arguments that we have not been appealing to still apply (see section 3 above). Putting these worries aside, we question the interest of the argument. A compelling argument has premises that are more plausible than its conclusion. But de Finetti's axioms are more plausible than the preference axioms the argument employs. Recall de Finetti's axioms: Normalization, Boundedness, and Qualitative Additivity. The first two are uncontroversial, so the third is the only real prize. By contrast, the representation theorem argument requires several rather awkward assumptions about rational preference, plus Informed Preference. So if we don't already accept de Finetti's axioms, it is unlikely that this argument will convince us.

5 Conclusion

Representation theorems have been taken to secure the foundations of decision theory in two ways: they are taken to characterize the notions of degree of belief and utility (characterizational representationalism), and they are taken to justify the key tenets of normative decision theory (normative representationalism). We have argued that both of these claims are problematic.

Characterizational representationalism can be cashed out in a number of ways, but each of these precisifications is either unjustified or uninteresting. The source of these difficulties is that the central claim that a satisfactory account of CR needs — that the representation theorem provides an accurate characterization of the degrees of belief and utilities that play a role in our descriptive and normative theorizing — is false. The moral is that we should accept a more modest response to the characterizational challenge: we should accept that degrees of belief and utilities are quasi-theoretical notions inherited from folk-psychology, and that serious empirical work in psychology will need to be done before we can provide a more concrete account of what they are.

Normative representationalism founders as well. The source of these problems is that the preference constraints provided underdetermine the epistemological and decision-making methods the agent employs: there are

a number of different ways of representing the agent that yield the same preferences. The standard way to eliminate these unwanted representations is to adopt something like Representation Accuracy, but this assumption is problematic. Christensen attempts to eliminate the underdetermination by adding Informed Preference, but this addition is insufficient, since many alternative representations remain. Zynda eliminates the underdetermination by claiming that it is psychologically unreal, but this coarsens the notion of degree of belief in undesirable ways. The moral is that here, too, we should settle for modesty. We should treat these normative claims like any others, and justify them in similar ways.

CR and NR have often been taken to provide the foundations of decision theory. But these foundations are unstable, and by resting decision theory on them, we place it in peril. We should reject these foundations, and dispose of CR and NR. By doing so, we clear the way for a firmer foundation for decision theory.

References

Allais, Maurice. 1979. The so-called allais paradox and rational decisions under uncertainty. In Maurice Allais & Ole Hagen (eds.) *Expected Utility Hypotheses and the Allais Paradox*. D. Reidel.

Birnbaum, Michael & A. Chavez. 1997. Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes* 71.

Birnbaum, Michael & William McIntosh. 1996. Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes* 67.

Birnbaum, Michael & Juan Navarette. 1998. Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty* 17.

Carnap, Rudolph. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Christensen, David. 1996. Dutch-book arguments depragmatized. *Journal of Philosophy* 93(9).

Christensen, David. 2001. Preference-based arguments for probabilism. *Philosophy of Science* 68.

Christensen, David. 2004. *Putting Logic in its Place*. Oxford University Press.

Churchland, Patricia S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press.

Churchland, Paul M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78.

Ellsberg, Daniel. 1961. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* 75.

Fishburn, P C. 1983. Transitive measurable utility. *Journal of Economic Theory* 31: 293–317.

Howson, Colin & Peter Urbach. 2005. *Scientific Reasoning: A Bayesian Approach*. Open Court.

Jeffrey, Richard C. 1974. Preference among preferences. *Journal of Philosophy* 71(13): 377–391.

Jeffrey, Richard C. 1983. *The Logic of Decision*. University of Chicago Press, 2nd edn.

Joyce, James. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science* 65(4).

Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kahneman, Daniel & Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47.

Knetsch, Jack L. 1989. The endowment effect and evidence of nonreversible indifference curves. *The American Economic Review* 79(5).

LeRoy, S. 2003. Expected utility: A defense. *Economic Bulletin* 7: 1–3.

Levy, Moshe & Haim Levy. 2002. Prospect theory: Much ado about nothing? *Management Science* 48: 1334–1349.

Lichtenstein, Sarah & Paul Slovic. 1971. Reversals of preferences between bids and choices in gambling decisions. *Experimental Psychology* 89.

Lichtenstein, Sarah & Paul Slovic. 1973. Response-induced reversals of preference in gambling decisions: An extended replication in las vegas. *Journal of Experimental Psychology* 101.

Maher, Patrick. 1993. *Betting on Theories*. Cambridge University Press.

Monti, Martin M., Simon Grant & Daniel N. Osheron. 2005. A note on concave utility functions. *Mind and Society* 4: 85–96.

Prelec, Drazen. 2000. Compound invariant weighting functions in prospect theory. In Daniel Kahneman & Amos Tversky (eds.) *Choices, Values, and Frames*. Cambridge University Press.

Quiggin, J. 1982. Defending expected utility theory. *Journal of Economic Behavior and Organization* 3: 323–343.

Rabin, Matthew & Richard H. Thaler. 2001. Anomalies: Risk aversion. *Journal of Economic Perspectives* 15(1): 219–232.

Ramsey, Frank P. 1926. Truth and probability. In Richard B. Braithwaite (ed.) *The Foundations of Mathematics and other Logical Essays*. Routledge and

Kegan Paul, 1931.

Savage, Leonard J. 1954. *The Foundations of Statistics*. Wiley Publications in Statistics.

Schervish, Mark J., Teddy Seidenfeld & Joseph B. Kadane. 1990. State-dependent utilities. *Journal of the American Statistical Association* 85(411).

Shafer, Glenn. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Spohn, Wolfgang. 1988. Ordinal conditional functions: A dynamic theory of epistemic states. In William Leonard Harper & Brian Skyrms (eds.) *Causation in Decision, Belief Change, and Statistics II*. Kluwer.

Stich, Stephen P. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Bradford Books.

Tuthill, Jonathan & Darren Frechette. 2002. Non-expected utility theories: Weighted expected, rank dependent, and, cumulative prospect theory utility. In *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*.

Tversky, Amos & Daniel Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Rish and Uncertainty* 5: 297–323.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford University Press.

Wakker, Peter & Amos Tversky. 1993. An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* 7(7): 147–176.

Wakker, Peter P. 2003. The data of levy and levy 'prospect theory: Much ado about nothing?' actually support prospect theory. *Management Science* 979–981.

Watt, R. 2002. Defending expected utility theory. *Journal of Economic Perspectives* 16: 227–230.

Zynda, Lyle. 2000. Representation theorems and realism about degrees of belief. *Philosophy of Science* 67.