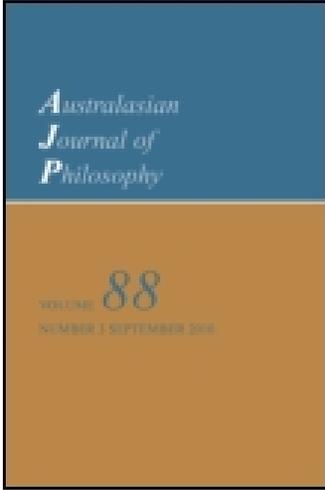


This article was downloaded by: [Texas Technology University]

On: 12 November 2014, At: 09:51

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954  
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH,  
UK



## Australasian Journal of Philosophy

Publication details, including instructions for authors  
and subscription information:

<http://www.tandfonline.com/loi/rajp20>

### Representation Theorems and the Foundations of Decision Theory

Christopher J. G. Meacham<sup>a</sup> & Jonathan Weisberg<sup>b</sup>

<sup>a</sup> University of Massachusetts, Amherst

<sup>b</sup> University of Toronto

Published online: 27 Oct 2010.

To cite this article: Christopher J. G. Meacham & Jonathan Weisberg (2011)  
Representation Theorems and the Foundations of Decision Theory, *Australasian Journal  
of Philosophy*, 89:4, 641-663, DOI: [10.1080/00048402.2010.510529](https://doi.org/10.1080/00048402.2010.510529)

To link to this article: <http://dx.doi.org/10.1080/00048402.2010.510529>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly

forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# REPRESENTATION THEOREMS AND THE FOUNDATIONS OF DECISION THEORY

Christopher J. G. Meacham and Jonathan Weisberg

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

The formalism of expected utility theory uses a probability function and a utility function to rank options. This formalism has many applications.<sup>1</sup> But philosophers have traditionally been interested in its use as a *decision theory*, where the probabilities and utilities in question represent an agent's degrees of belief and desires, and the ranking of options represents their preferences.

This decision theory has two interpretations, one descriptive and the other normative. On the descriptive interpretation, the theory says that agents have probabilistic degrees of belief and maximize expected utility. This is to understand decision theory as psychology, as a theory that characterizes the structure of belief and desire, and predicts and explains behaviour. On the normative interpretation, the theory says that agents ought to have degrees of belief satisfying the probability axioms and ought to maximize expected utility.<sup>2</sup> This is to understand decision theory as providing advice, or criteria of rational belief and choice.

Both interpretations face foundational challenges. The first challenge is to make sense of the notions of degree of belief and utility: we might be sceptical that there really are such things as belief and desire, that they can be meaningfully assigned numerical values, or that they are related to

<sup>1</sup>Rosenberg [1992] distinguishes some applications in economics, and Ross [2005: 325–9] discusses the theory as a general tool for modelling phenomena as diverse as the behaviour of neurons, the behaviour of social animals, and the sub-personal interests of people.

<sup>2</sup>It's natural to take both descriptive and normative decision theory to include claims about the structure of utilities as well. Since this matter has received little discussion in the literature, we won't explicitly discuss it in the text. But our discussion applies to it as well.

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 is needed 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. But many authors instead look to *representation theorems* to provide precise and general characterizations of these notions.<sup>3</sup> We'll call this response to the first challenge *characterizational representationalism*, or CR.<sup>4</sup> Characterizational representationalism holds that representation theorems play a crucial role in characterizing the notions of degree of belief and utility, the graded notions of belief and desire that appear in our folk, descriptive and normative theorizing.

Making sense of degrees of belief and utilities is a challenge for both the descriptive and normative interpretations of decision theory, but normative decision theory faces an additional challenge: justifying 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 claims, assess how well they agree with considered intuitions about cases, and so on. But again, many authors look instead to representation theorems to provide a more decisive and rigorous defence of these claims. We'll call this immodest response *normative representationalism*, or NR.<sup>5</sup> Normative representationalism holds that representation theorems play a crucial role in justifying the normative theses of probabilism and expected utility maximization.

We will argue that both characterizational representationalism (CR) and normative representationalism (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 psychology literature, the resulting failure of NR has not been appreciated. Our aim 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. §2 presents some background. §3 evaluates several versions of CR. In each case we argue that the characterization of degrees of belief and utilities provided is either inaccurate or uninteresting. We then turn to evaluate NR's prospects. §4 shows how CR's failure

<sup>3</sup>More specifically, we mean theorems that guarantee relatively unique expected utility representations of preference orderings, e.g. the theorem of Savage [1954]. Representation theorems outside of expected utility theory, like those of Kahneman and Tversky [1979] and Wakker and Tversky [1993], are not our topic of discussion.

<sup>4</sup>CR and the 'modest' response are not the only options, of course. Another alternative is offered by Hájek and Eriksson [2007], for example. While we favour the modest response, defending alternatives to CR is not our aim in this paper. We will lay out only the case against CR.

<sup>5</sup>As with CR, NR and the 'modest' response are not the only options. One could build one's decision theory on other foundations, in the manner of Levi [1997], for example. While here again we favour the 'modest' response, defending alternatives to NR is not our aim. Our aim is merely to lay out the case against NR.

undercuts the standard argument for NR. §5 then examines some recent attempts to defend NR without presupposing CR, specifically the arguments of David Christensen [2001, 2004] and Lyle Zynda [2000]. We summarize our conclusions in §6.

## 2. Representation Theorems

A typical representation theorem says that if an agent's preferences satisfy certain constraints, there is a unique probability function and utility function whose expected utility ranking coincides with the agent's preferences:<sup>6</sup>

*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 these variations. So we will leave  $C$  as an unspecified placeholder for most of the discussion.

In what follows, we will take an agent's preferences to encode their behavioural dispositions. We do this because normative decision theory is generally understood as an account of how an agent's degrees of belief and utilities determine how they ought to act. And for a representation theorem to yield these kinds of conclusions, it too must relate an agent's degrees of belief and utilities to their actions. Similarly, the kinds of descriptive claims NR implicitly relies on must link an agent's degrees of belief and utilities with their actions.

This is not the only way to understand preferences. For example, one might take preference to be a kind of mental state that's independent of behaviour—something like the mental state an agent is in when they truthfully assent to claims like 'I prefer going to the show over staying at home.' And one might understand the claim that one is an expected utility maximizer as a claim about how one's degrees of belief and utilities are related to this mental state, not to what decisions one makes. Likewise, one might understand the normative thesis of expected utility maximization as a claim about what this mental state ought to be, not what one ought to do.

But this kind of substitution will have little bearing on the arguments we present. Although this variation on CR makes different descriptive claims, the criticisms of CR we will raise apply to it in the much the same way.<sup>7</sup> Likewise, although this version of NR will make different normative claims,

<sup>6</sup>Strictly speaking, perfect uniqueness is not guaranteed. The standard theorems yield a unique probability function and a utility function unique up to linear transformations. To simplify the discussion, however, we will speak as if perfect uniqueness is guaranteed.

<sup>7</sup>For example, the empirical evidence that we don't act as expected utility maximizers and don't act in ways which satisfy conditions  $C$  (see §3.1.1) is also evidence that we're not expected utility maximizers with respect to these kinds of mental states, and that these kinds of mental states won't satisfy conditions  $C$ .

the criticisms of NR we raise will apply *mutatis mutandis* to this approach as well.

### 3. Characterizational Representationalism

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 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.<sup>8</sup>

But to use the theorem to characterize our degrees of belief and utilities, 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.<sup>9</sup>

First, the theorem yields the desired  $p$  and  $u$  functions only if the agent's preferences satisfy the conditions in  $C$ . 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 providing a characterization of degrees of belief and utilities only for agents who do satisfy these constraints, whether or not actual people are such agents?

Second, for agents who do satisfy the constraints in  $C$ , 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.<sup>10</sup> 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.<sup>11</sup>

<sup>8</sup>It is questionable whether preferences are significantly less problematic than degrees of belief, but many characterizational representationalists do seem to have held that they are. Ramsey, for example, clearly had some such view in mind [2001 (1931): 166].

<sup>9</sup>A further problem is that the standard representation theorems provide representations in terms of state-independent utilities, and so cannot correctly represent agents with state-dependent utilities; agents (arguably) like ourselves. See Schervish *et al.* [1990] and Howson and Urbach [2005: 58–9] for some discussion.

<sup>10</sup>A similar point is made by Hájek [2008].

<sup>11</sup>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 rationale. 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 (un)justified.

After all, an agent with badly incoherent degrees of belief who often acts contrary to expected utility maximization can still have preferences that satisfy  $C$ . For example, suppose Holmes is an expected utility maximizer, and thus his preferences satisfy the constraints in  $C$ . Now suppose that Holmes is struck on the head, and as a result comes to have the same credences and utilities as Moriarty, though Holmes's preferences remain unchanged. Holmes's psychology is now in a very strange state: he has a very high credence that a certain course of action will put Irene Adler in peril and he attaches a high utility to placing her in peril, yet his preferences favour a course of action that will keep her safe. And since Holmes acts in accordance with his preferences, he is no longer an expected utility maximizer. But since his preferences are the same, the theorem will still assign him the same probabilities and utilities he had to start with, even though those probabilities and utilities no longer describe his actual state of mind.<sup>12</sup>

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.

To bring the theorem to bear on our degrees of belief and utilities, the characterizational representationalist needs to make assumptions that tie the  $p$  and  $u$  functions that appear in the representation theorem to the degrees of belief and utilities of actual people. Our argument against CR will be a dilemma about the status of these assumptions. The characterizational representationalist's assumptions must be either empirical or non-empirical. If they are empirical, we'll argue, the resulting view is unjustified. If they are non-empirical, we'll argue, 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. Instead we will consider 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.

### 3.1 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

<sup>12</sup>One might object that the stipulated possibility is not a genuine one. Perhaps such a psychological state is nomologically impossible because it violates a law of psychology; maybe it's even metaphysically or conceptually impossible, violating some necessary/conceptual connection between credence, utility, and preference. But to make this objection is to posit precisely the sort of additional assumption we mean to highlight here, and to criticize in the following sections. A major burden of the remainder of §3 is to argue against the sorts of empirical/*a priori* assumptions that might close the logical gap this example illustrates.

theorem assigns to them describe their actual degrees of belief and utilities. We'll then consider a less ambitious stance, the *approximate representation* view, which 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 is the *counterfactual 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.

*3.1.1 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 psychological research that tells against them. To bring out these conflicts, we'll consider four consequences 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. Such a tight connection does not sit well with our folk conceptions of belief, desire, and preference.

Plausibly, *some* relations tend to hold between belief, desire, and preference. After all, we often explain someone's behaviour 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 several 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's implausible that just one of those connections is paramount. With all the pushes and pulls that beliefs and desires are entangled in, we should not expect there to be a rigid and straightforward connection between degree of belief, utility, and preference.

*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 that we violate expected utility maximization all the time.

Cases of akrasia are classic examples of non-maximizing behaviour. 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 impulse or temptation. And

we tend to describe our behaviour 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.<sup>13</sup>

Other, less dramatic examples of expected utility violations are common. Some people will drive three miles further to buy petrol where it is \$0.10 cheaper per gallon, even though the extra petrol 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 computer we want is \$2285 instead of \$2295. It is hard to tell a plausible story that reconciles this behaviour with expected utility maximization.

It is not impossible to tell such a story, of course. Perhaps the subject gets so much satisfaction out of a good bargain, or loathes paying the middleman's jacked-up prices so fiercely, that these behaviours really do maximize her expected utility. Similarly, in cases of akrasia, 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 view makes the empirical claim that the  $u$  function the representation theorem assigns correctly describes our utilities—the graded desires that appear 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 behaviour 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 behaviour as symptomatic of an idiosyncratic utility function; we criticize it, and try to correct our behaviour in the future.

In addition to these commonplace observations of non-maximizing behaviour, there is much 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 [1979], for example,

<sup>13</sup>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.

We think cases of akrasia nevertheless put serious strain on the expected utility hypothesis. Limitations of space prevent a thorough defence of this view, but we will briefly note two observations. First, Jeffrey's view cannot allow that the indulgent dieter's action is criticizable, since he maximizes expected utility. On Jeffrey's view, he has an unsatisfied second-order preference, to have abstinent first-order preferences, it's just that he cannot change his first-order preferences to satisfy this second-order preference. He can only take long-term steps to adjust such preferences in the future. But 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 not criticizable. Second, because Jeffrey holds that the indulgent dieter maximizes expected utility, he must distort the notion of utility in unacceptable ways. This is a more general problem we'll discuss shortly.

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 stochastic dominance [Birnbaum and McIntosh 1996; Birnbaum and Navarette 1998]. Again, these results are at odds with expected utility maximization, but are explained by the RAM [Birnbaum and Navarette 1998] and TAX [Birnbaum and Chavez 1997] models. 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 modelled by other accounts, like cumulative prospect theory [Tversky and 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.<sup>14</sup> But 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] showed 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: actual representationalism robs normative decision theory of its interest. Normative decision theory applies only to agents who have degrees of belief and utilities. But if actual representationalism is true, agents who have degrees of belief and utilities are automatically probabilistic expected utility maximizers, and the further question whether they ought to be probabilistic expected utility maximizers is moot.<sup>15</sup>

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

<sup>14</sup>For some defences of expected utility as a descriptive hypothesis and/or criticisms of alternatives, see Watt [2002], LeRoy [2003], and Levy and Levy [2002]. For some responses to these defences, see Rabin and Thaler [2001], Montuori *et al.* [2005], and Wakker [2003].

<sup>15</sup>A natural reply: we should understand normative decision theory as the injunction to *have* degrees of belief and utilities, and thus, *a fortiori*, to *be* agents who maximize expected utility. And *this* injunction isn't trivial when applied to agents who don't have degrees of belief and utilities.

But this proposal represents a substantive shift in the content of normative decision theory. We're no longer dealing with the same norms, and these replacements can't do the same work as the originals. For example, normative decision theory is supposed to say which of an agent's options she ought to take. But the injunction to have degrees of belief and utilities will be silent on this, since every option will maximize expected utility relative to some pair of probability and utility functions.

threatens to trivialize normative decision theory. At the very least, one should feel uneasy about resting the foundations of decision theory on the actual representation view. Many proponents of CR have rejected the actual view for such reasons, favouring weaker views like the approximate and counterfactual representation views. We turn to these now.

*3.1.2 The Approximate Representation View* The approximate representation view tries to avoid the preceding problems by weakening its empirical claims. It 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.

Constructing 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. For it to be true in an interesting way, there must be no other theory at the same level of abstraction that better describes the target phenomena. 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 [Kahneman and Tversky 1979; Tversky and Kahneman 1992; Knetsch 1989; Birnbaum and McIntosh 1996; Birnbaum and Chavez 1997; Birnbaum and Navarette 1998; Prelec 2000; Tuthill and Frechette 2002].<sup>16</sup>

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

<sup>16</sup>Again, there are dissenting opinions; see note 14.

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.

*3.1.3 The Counterfactual Representation View* The counterfactual representation view is even more conservative than the approximate view. Unlike the actual and approximate views, the counterfactual view does not make claims about our actual preferences. Instead, it holds that if our preferences *were* to satisfy *C*, 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 counterfactual 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.<sup>17</sup> Viewed in this light, the counterfactual 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.

As an analogy, consider what it would take for humans 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 counterfactual view is that, like the approximate view, it undermines a central motivation for adopting

<sup>17</sup>See Howson and Urbach [2005: 57–60] for a recent survey of the role of state-dependence assumptions in representation theorems.

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. While the approximate view failed to be precise, the counterfactual view fails to be general. The counterfactual view provides a characterization of degrees of belief and utilities that applies only to agents who satisfy the required preference constraints. This characterization doesn't apply to agents like ourselves whose preferences don't satisfy these constraints. And while it asserts counterfactual facts about what our degrees of belief and utilities would be, were we to have different preferences, it tells us nothing about what our actual degrees of belief and utilities are. So the counterfactual view is of little help in characterizing the kinds of degrees of belief and utilities had by agents like ourselves, and does little to lessen our reliance on intuitive folk notions when using decision theory.

The counterfactual view avoids empirical refutation 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. Restricting itself to remote possibilities, it is of little help in characterizing actual degrees of belief and utilities, doing little to diminish our dependence on folk notions when using decision theory.

### 3.2 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.

**3.2.1 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 actual

people aren't expected utility maximizers with respect to their degrees of belief and utilities. The stipulative view escapes this worry, because it does not follow that people are not expected utility maximizers with respect to their degrees of belief\* and utilities\*. After all, degrees of belief\* and utilities\* needn't 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 degrees of belief\* and utilities\* are of no use to the characterizational representationalist's project. If they don't correspond to the notions we ordinarily work with in our theorizing, then it is misleading to suggest that the representation theorem characterizes, illuminates, or legitimates the subject matter of that theorizing.

A second worry was that ordinary people don't appear to satisfy the preference constraints in *C*, so the representation theorem does not apply to them. 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\*. Here, again, the notions of degree of belief\* and utility\* are of no use to the characterizational representationalist, since actual people do not have degrees of belief\* and utilities\*. In so far as the characterizational representationalist seeks to characterize, illuminate, or legitimate the subject matter of our talk about actual people's degrees of belief and utilities, her stipulative definitions of 'degree of belief\*' and 'utility\*' are failures; actual people do not have degrees of belief\* and utilities\*.

One might suggest that, even if actual people do not have degrees of belief\* and utilities\*, possible people do—maybe even possible people whose preferences approximate those of actual people. But this is just to fall back on the approximate view or counterfactual view criticized above, and to incur the empirical commitments that made those views problematic.

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

*3.2.2 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 on the

first two criteria. But they do poorly enough on the latter two to be unsatisfactory explications of our original concepts.

In particular, let us focus on (iv). Degrees of belief\* and utilities\* will not apply to most of the same cases as our original notions. Recall how the non-empirical views escape the worries that face the actual representation view: evidence that people don't satisfy *C* isn't evidence that these views are false, just evidence that people don't have degrees of belief\* and utilities\*. But if people don't have degrees of belief\* and utilities\*, these terms 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 *C*, they cannot play a useful role in descriptive accounts of our mental states, predictive accounts of our behaviour, or prescriptive accounts of what our behaviour ought to be.

Even putting these 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, since adopting degrees of belief\* and utilities\* trivializes normative decision theory. Normative decision theory applies only 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. If we take the 'degrees of belief' and 'utilities' that appear in normative decision theory to be degrees of belief\* and utilities\*, it will be true by definition that all agents subject to the norms of decision theory satisfy them.

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

### 3.3 General Considerations

We've looked at the problems facing five kinds of characterizational representationalism. There are other varieties we haven't considered, but we can get a feel for the problems they will face 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 providing a precise characterization. The counterfactual view gives up on providing 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 trade-offs 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. The range of views between these two ends trade off between these difficulties.

So CR's prospects are bleak. Different versions of the view can avoid particular problems, but the source of the problems remains: the psychological picture at the heart of CR is false.

#### 4. 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 can be represented as 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*, one is representable as a probabilistic expected utility maximizer. But given only that our preferences ought to satisfy *C*, this merely suggests that we ought to be *representable as* probabilistic expected utility maximizers. It doesn't yield the stronger claim that we ought to *actually be* probabilistic expected utility maximizers. Suppose, for example, that Jones attaches maximal expected utility to drinking the glass of aged, single malt whisky in front of him. Given only the weak claim about representability, there is nothing irrational in his preferring to pour the whisky down the sink untouched, and acting accordingly. As long as his preferences and behaviour can be represented as those of some (puritanical) expected utility maximizer, he has satisfied all the normative constraints imposed on him so far.<sup>18</sup>

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

<sup>18</sup>A similar point is made in Hájek [2008].

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 concerned only 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 about the relations 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 it 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 counterfactual representation view, and suffers from similar problems. Like the counterfactual view, it escapes conflicting with our evidence because it confines itself to cases we don't have evidence about. But this also cuts it off from empirical support. We have no empirical reason to think that agents with  $C$ -satisfying preferences would have the psychological states the representation attributes to them. So the Representation Accuracy postulate is unjustified.

On the non-empirical side of the dilemma, the problems parallel those facing the non-empirical take on CR. For example, suppose 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 like Representation Accuracy is the biggest problem with the argument for NR. But there are other worries. Many question whether we should accept the normative preference constraints it assumes.<sup>19</sup> Some object that the argument conflates instrumental and epistemic rationality [Christensen 1996, 2001, 2004; Joyce 1998]. And a few point 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 [Schervish *et al.* 1990; Howson and Urbach 2005: 58–9].

We are sympathetic to some of these concerns, but they are largely independent of the theme we have been developing. Our primary concern is how the dependence on problematic assumptions about 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.

### 5. Normative Representationalism Without Representation Accuracy

One can defend NR without assuming Representation Accuracy. We'll consider one first-pass attempt to do so before turning to two more sophisticated attempts that appear in the literature.

One way to defend NR without employing an assumption like Representation Accuracy is to weaken the conclusion. If the original argument does 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.

This move gives up the game, however. If 'as if' theses are all we're after, 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 an assumption we start out with.<sup>20</sup> Moreover, if we are interested in deciding between alternative theories of epistemic and practical rationality—e.g. between theories based on additive vs. super-additive functions [Shafer 1976]—the representation theorem is useless. All

<sup>19</sup>Savage's independence axiom has long been challenged on the grounds that it violates intuitions in the Allais and Ellsberg paradoxes [Allais 1979; Ellsberg 1961]. More technical axioms are often accused of making unrealistic demands on real agents. For some contemporary responses to these objections, see Maher [1993] and Joyce [1999].

<sup>20</sup>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.

that matters is whether the competing theories provide a decision rule consistent with  $C$ , and the representation theorem doesn't help us assess that.

We turn now to two recent attempts to defend NR without assuming Representation Accuracy: those of Christensen [2001, 2004] and Zynda [2000]. Since their proposals focus on the justification of probabilism, the remainder of our discussion leaves expected utility maximization to the side.

### 5.1 Christensen

Christensen agrees that the standard argument for NR is unsound, and proposes to repair it by replacing Representation Accuracy with a normative claim. He proposes:

*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.<sup>21</sup>

Informed Preference is an informal, normative version of Savage's *coherence condition* (to borrow Joyce's terminology [1999: 89, 130]). 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.

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 say that there is a unique degree of belief function that coheres with your preferences, and that this function is a probability function. Rather, they say 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-

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

expected utility maximizing ways.<sup>22</sup> For example, consider the function  $f: \mathbb{R} \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ :

$$f(p) = [1/(1-p)] - 2 \text{ if } p \neq 1, \\ = \infty \text{ otherwise.}$$

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.<sup>23</sup> 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.<sup>24</sup>

One might suggest that the difference between  $p$  and  $f(p)$  is merely notational or conventional. On this view, the apparent difference between  $p$  and  $f(p)$  is just an 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. But the states of mind described by  $p$  and by  $f(p)$  are really the same, just described in different terms.

But 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 way  $p$  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. Lyle Zynda [2000] endorses precisely this view.

## 5.2 Zynda

The problem for Christensen's proposal was non-uniqueness: because many non-probabilistic functions cohere with a given preference ordering, Informed Preference does not require our degrees of belief to be probabilistic. Zynda proposes to eliminate the non-uniqueness by suggesting that any two representations of 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 0.5 more likely than  $B$ ' are artifacts of the particular numerical representation we choose.

<sup>22</sup>See Zynda [2000] and Hájek and Eriksson [2007] for discussions of non-probabilistic representations.

<sup>23</sup>Consider a  $\sigma$ -algebra  $\mathcal{A}$  over  $\Omega$ . A probability function  $p: \mathcal{A} \rightarrow \mathbb{R}$  must satisfy the three probability axioms: (i)  $\forall A \in \mathcal{A}, p(A) \geq 0$ , (ii)  $p(\Omega) = 1$ , (iii) if  $A_1, A_2, \dots$  are mutually exclusive, then  $p(\cup_i A_i) = \sum_i p(A_i)$ .

<sup>24</sup>In general, any order-preserving map from the unit interval will yield a degree of belief function that coheres with your preferences.

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 to require your beliefs to satisfy 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, this is as much as one could ask for. And de Finetti's axioms do have non-trivial content: they rule out competing views about rational belief, such as Dempster-Shafer Theory [Shafer 1976] and non-standard interpretations of Ranking Theory [Spohn 1988].

While Zynda's argument avoids the 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 we are left with are just qualitative, agent-relative comparisons like 'The agent thinks  $A$  is more likely than  $B$ '. And this is not 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 Dick 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. For example, the standard Bayesian resolution of the raven paradox says that 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.<sup>25</sup>

Zynda defends NR by adopting a psychological proposal that makes the relevant notion of belief qualitative and agent-relative. If we adopt this view, the resulting beliefs cannot do the work demanded by much of our normative and descriptive theorizing. Since Zynda's defence of NR comes at the cost of embracing a notion of degree of belief that is too weak to do the work we want it to, the price of adopting this defence is too high.

### 5.3 *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 manoeuvre with Christensen's appeal to Informed Preference to set up an argument for de Finetti's axioms. An agent's preferences ought to obey *C*, and her degrees of belief ought to cohere with those preferences by 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 it is unappealing. For one thing, many of the additional problems for NR that we have not been appealing to still apply (see §4 above). But more, 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

<sup>25</sup>Could Zynda enjoy the fruits of numerical degree of belief talk without committing to the reality of anything more than ordinal structure? After all, anti-realists about unobservable entities appeal to unobservables in their explanations, despite being agnostic about the existence of those entities. Similarly, we might 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. First, the view is unmotivated, since our reasons for being realists about ordinal structure are equally reasons for realism about structure encoding the ratios of differences. 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). And the same is true of the claim that degrees of belief have structure encoding the ratios of differences: magnitudes allow us to explain and predict facts about assertion, deliberation time, and confirmation, introspection suggests that some differences in certainty are greater than others, and so on.

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 genuine distinction underlies different numerical representations that are ordinally equivalent. But if the anti-realist can recognize the distinction, she will be able to make sense of norms that track such distinctions, like NR (where, of course, the anti-realist's version of NR is cashed out in terms of the genuine facts that underlie this distinction).

argument employs. Normalization and Boundedness are uncontroversial, so Qualitative Additivity is the only real prize. By contrast, the representation theorem argument requires several rather awkward assumptions about rational preference, plus Informed Preference. If we don't already accept de Finetti's axioms, it is unlikely that this argument will convince us.

## 6. Conclusion

Representation theorems have been taken to secure the foundations of decision theory in two ways: by characterizing the notions of degree of belief and utility and by justifying the key tenets of normative decision theory. 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.

Normative representationalism founders as well. The preference constraints assumed by the standard argument underdetermine the epistemological and decision-making methods the agent employs: there are many ways of representing the agent that yield the same preferences. We can eliminate these unwanted representations by adopting 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.

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.<sup>26</sup>

*University of Massachusetts, Amherst*  
*University of Toronto*

Received May 2009  
Revised July 2010

<sup>26</sup>We are grateful to Frank Arntzenius, Maya Eddon, Hilary Greaves, Alan Hájek, John Hawthorne, Barry Loewer, Adam Sennet, Michael Strevens, and several anonymous referees for helpful discussion and feedback.

## References

- Allais, Maurice 1979. The So-Called Allais Paradox and Rational Decisions Under Uncertainty, in *Expected Utility Hypotheses and the Allais Paradox*, ed. Maurice Allais and Ole Hagen, Dordrecht: Reidel: 437–681.
- Birnbaum, Michael and A. Chavez 1997. Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence, *Organizational Behavior and Human Decision Processes* 71/2: 161–94.
- Birnbaum, Michael and William McIntosh 1996. Violations of Branch Independence in Choices Between Gambles, *Organizational Behavior and Human Decision Processes* 67/1: 91–110.
- Birnbaum, Michael and Juan Navarette 1998. Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence, *Journal of Risk and Uncertainty* 17/1: 49–79.
- Carnap, Rudolph 1950. *Logical Foundations of Probability*, Chicago: University of Chicago Press.
- Christensen, David 1996. Dutch-Book Arguments Depragmatized, *Journal of Philosophy* 93/9: 450–79.
- Christensen, David 2001. Preference-Based Arguments for Probabilism, *Philosophy of Science* 68/3: 356–76.
- Christensen, David 2004. *Putting Logic in its Place*, New York: Oxford University Press.
- Ellsberg, Daniel 1961. Risk, Ambiguity, and the Savage Axioms, *Quarterly Journal of Economics* 75/4: 643–69.
- Fishburn, P. C. 1983. Transitive Measurable Utility, *Journal of Economic Theory* 31/2: 293–317.
- Hájek, Alan 2008. Arguments for—or Against—Probabilism?, *British Journal for the Philosophy of Science* 59/4: 793–819.
- Hájek, Alan and Lina Eriksson 2007. What are Degrees of Belief?, *Studia Logica* 86/2: 185–215.
- Howson, Colin and Peter Urbach 2005. *Scientific Reasoning: A Bayesian Approach* (3rd Edition), Chicago and La Salle: Open Court.
- Jeffrey, Richard C. 1974. Preference Among Preferences, *Journal of Philosophy* 71/13: 377–91.
- Jeffrey, Richard C. 1983. *The Logic of Decision* (2nd Edition), Chicago and London: University of Chicago Press.
- Joyce, James 1998. A Nonpragmatic Vindication of Probabilism, *Philosophy of Science* 65/4: 575–603.
- Joyce, James 1999. *The Foundations of Causal Decision Theory*, New York: Cambridge University Press.
- Kahneman, Daniel and Amos Tversky 1979. Prospect Theory: An Analysis of Decision Under Risk, *Econometrica* 47/2: 263–91.
- Knetsch, Jack L. 1989. The Endowment Effect and Evidence of Nonreversible Indifference Curves, *The American Economic Review* 79/5: 1277–84.
- LeRoy, Stephen F. 2003. Expected Utility: A Defense, *Economics Bulletin* 7/7: 1–3.
- Levi, Isaac 1997. *The Covenant of Reason: Rationality and the Commitments of Thought*, New York: Cambridge University Press.
- Levy, Moshe and Haim Levy 2002. Prospect Theory: Much Ado About Nothing?, *Management Science* 48/10: 1334–49.
- Lichtenstein, Sarah and Paul Slovic 1971. Reversals of Preferences Between Bids and Choices in Gambling Decisions, *Experimental Psychology* 89: 46–55.
- Lichtenstein, Sarah and Paul Slovic 1973. Response-Induced Reversals of Preference in Gambling Decisions: An Extended Replication in Las Vegas, *Journal of Experimental Psychology* 101/1: 16–20.
- Maher, Patrick 1993. *Betting on Theories*, New York: Cambridge University Press.
- Monti, Martin M., Simon Grant and Daniel N. Osherson 2005. A Note on Concave Utility Functions, *Mind and Society* 4/1: 85–96.
- Prelec, Drazen 2000. Compound Invariant Weighting Functions in Prospect Theory, in *Choices, Values, and Frames*, ed. Daniel Kahneman and Amos Tversky, New York: Cambridge University Press: 67–92.
- Quiggin, John C. 1982. A Theory of Anticipated Utility, *Journal of Economic Behavior and Organization* 3/4: 323–43.
- Rabin, Matthew and Richard H. Thaler 2001. Anomalies: Risk Aversion, *Journal of Economic Perspectives* 15/1: 219–32.
- Ramsey, Frank P. 2001 (1931). Truth and Probability, in *The Foundations of Mathematics and other Logical Essays*, ed. Richard B. Braithwaite, London: Routledge: 156–98.
- Rosenberg, Alexander 1992. *Economics—Mathematical Politics or Science of Diminishing Returns?*, Chicago: University of Chicago Press.
- Ross, Don 2005. *Economic Theory and Cognitive Science*, Cambridge: MIT Press.
- Savage, Leonard J. 1954. *The Foundations of Statistics*, New York: John Wiley and Sons, Inc.
- Schervish, Mark J., Teddy Seidenfeld and Joseph B. Kadane 1990. State-Dependent Utilities, *Journal of the American Statistical Association* 85/411: 840–7.
- Shafer, Glenn 1976. *A Mathematical Theory of Evidence*, Princeton: Princeton University Press.
- Spohn, Wolfgang 1988. Ordinal Conditional Functions: A Dynamic Theory of Epistemic States, in *Causation in Decision, Belief Change, and Statistics II*, ed. William Leonard Harper and Brian Skyrms, Dordrecht: Kluwer Academic Publishers: 105–34.
- Tuthill, Jonathan and Darren Frechette 2002. Non-Expected Utility Theories: Weighted Expected, Rank Dependent, and Cumulative Prospect Theory Utility, *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*, URL = <<http://www.farmdoc.uiuc.edu/ncccl34>>
- Tversky, Amos and Daniel Kahneman 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty* 5/4: 297–323.

- Wakker, Peter and Amos Tversky 1993. An Axiomatization of Cumulative Prospect Theory, *Journal of Risk and Uncertainty* 7/7: 147–76.
- Wakker, Peter P. 2003. The Data of Levy and Levy ‘Prospect Theory: Much Ado About Nothing?’ Actually Support Prospect Theory, *Management Science* 49/7: 979–81.
- Watt, R. 2002. Defending Expected Utility Theory: Comment, *Journal of Economic Perspectives* 16/2: 227–30.
- Zynda, Lyle 2000. Representation Theorems and Realism About Degrees of Belief, *Philosophy of Science* 67/1: 45–69.