
Reviewed by Lara Buchak, UC Berkeley

Decision theory is used for a variety of purposes: decision makers use it to guide their own actions, and theorists use it both to normatively assess decision makers and to predict and explain their decisions. This book investigates whether the theory can fulfill all three of these purposes. In particular, Bermúdez explores three questions that decision theory must answer under any guise: How should we understand utility and preference? How finely should we individuate the possible outcomes in a decision problem? And how should choice be constrained over time? He argues that there are no answers to these questions that allow decision theory to serve all three purposes.

Decision theory begins with a theoretical core: this includes, among other features, a utility and probability function and a set of axioms about preferences that together imply that the utility function is of expectational form. Before we can put the theory to use, however, we have to fill it out further, which requires responding to three questions Bermúdez explores. To avoid confusion, I will use the term core decision theory to refer to the theoretical core. When I refer to some particular way of filling out the theory – which might involve interpreting or defining the terms involved, adding to the standard core, or even subtracting from it – I will call it a supplemented decision theory.

At the outset, it’s worth noting that this book does an admirable job of surveying and engaging the different possible responses to the challenges it addresses, so it is particularly worth reading to get a nuanced picture of the state of the field. But of course, the book does more than that: the central claim that none of the responses allows a supplemented decision theory to play all three roles simultaneously is, for the most part, convincing. Bermúdez draws a further conclusion: that because of the relationship between the roles, a supplemented decision theory cannot play one of them without simultaneously playing the others as well, and therefore that no supplemented decision theory can play any of the roles. I am less convinced by this conclusion, as I will explain, but Bermúdez’s discussion of it is instructive nonetheless.

The book is divided into five chapters. In the first chapter, Bermúdez presents his characterization of the three different roles decision theory purports to play, and how rationality relates to each role. The action-guiding role is fairly straightforward: decision theory as an action-guiding theory allows an agent to reason his way to a solution to a decision problem – that is, to a course of action, among those he believes to be available to him, that is best given his beliefs and desires. Decision theory in its normative role allows us to assess whether an agent has in fact reasoned his way to such a course of action; additionally, it allows us to assess whether the agent has formulated the decision problem as he should (e.g., identified all the available acts, ranked the outcomes in an acceptable way, and appropriately assessed the relationship between the acts and the outcomes). Finally, the third use of decision theory is to predict and explain behavior. To see how rationality fits into this dimension, note that we explain and predict action by characterizing an agent’s mental states – his beliefs, desires, and so forth – in a way that makes clear why they would give rise to that action; and note that this characterization must be guided by an assumption that the agent is understandable (we wouldn’t, for example, attribute to an agent beliefs and desires to which the action is directly and obviously opposed). The assumption of
rationality is the link between behavior and mental states, since we cannot access the agent’s mental states directly.

The next three chapters deal with Bermúdez’s three mandatory questions in turn. First is the challenge of how we should understand utility and preference. One way to understand these concepts is to give utility an operational interpretation, on which utility is simply a measure of preference, and to claim that preference is revealed by choice; that is, that there is no gap between what an agent chooses and what he prefers. An agent can use decision theory thus supplemented to guide his own actions, at least to a certain extent: this version of decision theory enjoins an agent to render his preferences internally consistent. However, the argument goes, decision theory thus supplemented has insufficient normative force because it cannot restrict choice in novel situations. Additionally, it is useless as an explanatory tool— we cannot explain what causes the agent to make the choices he makes—and can only serve a limited predictive function. We might modify this interpretation by pairing the operational interpretation with a view on which choice does not always straightforwardly reveal preference, such as the “considered preference” view, on which preference is only revealed when choice and verbally expressed preferences agree. The problem for this new interpretation, Bermúdez claims, is that a regress prevents the resulting theory from serving either the action-guiding or the normative role: when choice and verbally expressed preference disagree, the agent faces a new decision problem about which to endorse as his considered preference—but he cannot solve this problem without first determining his considered preferences about which of these two preferences to endorse.

Finally, we might reject the operational interpretation entirely: instead of taking the relation decision theory explicates to be the subjective preference relation, we might take it to be an objective relation. For example, we might take it to be the betterness relation and claim that what utility illuminates is not preference but rather the relative goodness of outcomes. Only this supplemented decision theory answers the first challenge for the normative role, because this theory can assess the rationality of novel choices. Unfortunately (according to Bermúdez), it cannot guide action, because the betterness relation cannot be deployed in practical reasoning: the only access we have to what is good is filtered through our beliefs (and if we try to instead use a subjective notion of goodness, we will face the same problems that plagued the operational interpretation). Furthermore, because agents do not have direct access to what is good, we will not be able to use this theory to explain their behavior.

The fact that no strategy works here for the dual purposes of action guidance and assessment highlights, to me, an important contrast between the two dimensions as Bermúdez sees them: any theory that an agent can use to guide his actions must be irreducibly subjective; that is, the considerations an agent uses to guide his actions must be filtered through his own beliefs (this is why goodness alone cannot guide action). Certain facts about what an agent ought to do (e.g., facts about the good) cannot be brought to bear, since they are facts unknown to the agent (or not directly available for use in his reasoning). Therefore, if a supplemented decision theory relies on these facts, the theory cannot be used to guide action; and if it does not, it cannot have normative force.

The second challenge deals with how finely outcomes should be individuated. I will illustrate this problem using two examples that Bermúdez discusses. The first, due to Amos Tversky and Daniel Kahneman, is a study illustrating the effects that the description of a decision problem has
on preferences. Decision makers are asked to imagine that a disease threatens the lives of 600 people, and to choose which option from among each pair is least bad:

A1: 200 people are saved.
A2: There is a 1/3 probability that 600 people will be saved and 2/3 probability that nobody will be saved.

B1: 400 lives are lost.
B2: There is a 1/3 probability that no lives are lost and 2/3 probability that 600 lives are lost.

Although A1 and B1 are equivalent, and A2 and B2 are equivalent, people tend to prefer A1 to A2 and B2 to B1. In the second example, due to Frederic Schick, a doctor is operating on a patient who is a heinous murderer – and the doctor assigns different utilities to “the patient lives” and “the murderer lives” even though these report the same state of affairs. What these examples show is that sometimes the same options become more or less appealing when described differently. A supplemented decision theory needs a principle that determines when two options should figure into a rational decision maker’s deliberations in the same way. Bermúdez examines several proposals. He claims that each one results in a theory that allows an agent to guide his own action, so we can set the action-guiding dimension aside for the time being. The first possibility is to introduce an invariance principle: two options can only count as different if they are not materially equivalent (though we may make minor revisions, e.g., they may count as different if the agent is unable to recognize them as materially equivalent). So the options in the examples cannot count as different options. Decision theory thus supplemented certainly has a clear basis for normative assessment, but it cannot fully explain and predict what agents will do, since, given what we know about human psychology, the explanation for why agents choose as they do in the examples will invoke the way the options are described. Furthermore, we wouldn’t be able to predict how agents would choose in the examples above – or know what motivates them to make those choices – without knowing how the options are described.

By contrast, Bermúdez considers two principles that would each supplement decision theory so that it could explain and predict action, and that each initially leave open whether agents ought not have the preferences in the examples. The first is Schick’s, and it adds to the traditional belief-desire picture the concept of “understandings.” These represent how the agent sees an outcome and the motivating force it has for him: so, for example, the doctor assigns different utilities to the same outcome, depending on how he understands it. This requires abandoning the invariance principle, and since the invariance principle follows from the expected utility principle (as Bermúdez argues) it also amounts to abandoning the standard axiomatic basis of core decision theory, the basis which ordinarily helps give the theory normative force (once it is supplemented). Additionally, this principle leaves us without a perspective – at least without a perspective internal to the supplemented theory – from which to assess an agent’s understandings. The other possibility is John Broome’s principle of individuation by justifiers, which states that two outcomes can count as different only if they differ in a way that can make it rational to have a preference between them. This is different from Schick’s principle in that Schick’s principle creates a multiplicity of ways of understanding one outcome, but Broome’s principle creates a multiplicity of (materially equivalent) outcomes. However, to apply the principle in normative

---

assessment, we need to know whether it is rational for an agent to have a preference between any two particular outcomes, and so we need to know whether an agent ought to have those preferences; but of course we cannot know that without applying an already (fully) supplemented theory. So, unfortunately, although Schick’s and Broome’s principles allow each resulting decision theory to be a theory of explanation and prediction, Bermúdez concludes that neither resulting theory can serve the normative role.

When it comes to individuating options, the project of normative assessment and the project of explanation and prediction seem to have been at odds to begin with. As Bermúdez puts it, “The descriptive project of explanation/prediction is in prima facie tension with the prescriptive project of normative assessment. In explaining and predicting behavior we are interested in how things are. When engaged in normative assessment our interest is in how things ought to be.” In other words, we have a problem as soon as we think there are cases in which an agent ought not treat an outcome as different under different descriptions, but does: a normative theory should treat it as a single outcome, but a theory of explanation and prediction should treat it as two. We simply can’t satisfy both of these demands.

Recall my earlier point that action-guiding supplementations of decision theory and normative supplementations are at odds because the former need to include principles that are motivating from an agent’s perspective and the latter need to include principles that are independent of whether they motivate any particular agent. We are now in a position to see how the third dimension fits into Bermúdez’s picture. To explain and predict, theories must include both of these requirements: they must include some principle that is motivating for the agent – otherwise how could we explain his action? – but since they must include a method of predicting novel choice, they must bring facts to bear that the agent doesn’t currently have access to. That they must bring facts to bear that the agent cannot access explains why explanatory and predictive versions of the theory are at odds with action-guiding versions when we are stating what utility and preference are, and that they must bring facts to bear that motivate the agent explains why they are at odds with normative versions when we are picking a principle that says when outcomes can be individuated.

The final challenge that Bermúdez addresses is how a supplemented decision theory should constrain preferences over time, e.g., how the theory should treat agents whose preferences are inconsistent with preferences they have expressed in the past or will express in the future. Here the main tension is between the action-guiding use on the one hand and both the normative assessment use and the explanatory and predictive use on the other. Bermúdez points out that a theory that guides action is indexed to the present moment: from the point of view of an agent deciding what he should do, his only reasons to act are determined by his current preferences and beliefs (though the content of these could include preferences at other times), and he can only choose which action to perform now. On the other hand, when we are assessing an agent’s actions, as well as when we are explaining and predicting his behavior, we treat him as a unified agent over time.

The three strategies Bermúdez considers are the sophisticated choice strategy, the rational preference change strategy, and the resolute choice strategy. The sophisticated choice strategy directs agents to consider how they will choose in the future and to choose a current action with that in mind: for example, if I don’t want to drink, but I know I’ll prefer to drink if I am around alcohol, I will not go to a bar, even if I now prefer going to the bar and not drinking to not going at all. This gives an agent clear guidance, but at the expense of choosing to override his future

---

4 Bermúdez, pg. 77.
preferences – even when it is the future preferences that ought to be satisfied, as Bermúdez constructs examples to show. Furthermore, since sophisticated choice is not always rational, adding a principle about what sophisticated choosers do to our theory will not help us predict the behavior of all rational agents. In response to these two worries, Bermúdez explores a second strategy that would distinguish between rational and irrational preference change; however, he rejects this solution too for the normative and explanatory/predictive dimensions. Finally, Bermúdez considers the resolute choice strategy, on which an agent must be committed to his previous decisions: so, in the example, I should go to the bar, but I must stick by my decision not to drink. While this gives us a compelling way to assess, explain, and predict action over time – because we are treating the agent in the unified way these projects require – it is plain that decision theory supplemented by this principle cannot be sufficiently action-guiding. After all, I either currently have preferences about how to take my past preferences into account, or I do not. If I do, then I am not really choosing resolutely, but if I don’t, then it is hard to see what could motivate me to defer to my past preferences. Again, there is a tension between the standpoint a decision maker must take when he is deliberating about what to do and the standpoint we as theorists take when we are figuring out what he ought to do and when we are explaining and predicting his behavior. Our assessment, explanation, and prediction of what he does must bring to bear a perspective – this time a perspective that is not anchored in any particular moment – that the agent himself cannot access at the time of his deliberation.

Bermúdez concludes that a supplemented decision theory cannot fill all three roles simultaneously; indeed, one cannot even fill any two roles simultaneously, since no supplementation that answers all three challenges works for more than one purpose. We might think that we could simply use different supplemented decision theories for each of the three purposes. However, Bermúdez claims that this is not possible. He claims, on the contrary, that decision theory cannot be supplemented to fill any of the three roles. He argues this by showing that the roles cannot be completely separated. Because of space limitations, I will focus mainly on the connection he posits between the action-guiding dimension and the normative dimension.

Bermúdez argues that if a supplemented decision theory can serve as a theory of normative assessment, it must also serve as an action-guiding theory. To say that a supplemented decision theory has normative force is to say that we ought to bring our probability and value judgments in line with its axioms. But it is only the case that we ought to do this because these judgments are the basis for our deliberation: they are what underlie our reasons for action. In other words, if a supplemented decision theory has normative force, it is only because of the role the theory has in deliberation – it is only because it guides us in making decisions. This is just to say that it can only satisfy the normative purpose if it can satisfy the action-guiding purpose. The force of a supplemented decision theory as a normative theory is inextricably tied to its force as an action-guiding theory.

This discussion is puzzling, for the following reason: the link Bermúdez forges between the action-guiding dimension and the normative dimension seems at odds with some of his earlier arguments that the responses to the three problems cannot work simultaneously for both of these dimensions. As stressed above, one reason that supplemented decision theories that work for the action-guiding dimension fail to work for the assessment dimension is that normative assessment involves not just evaluating how an agent acted given his understanding of the problem, but also evaluating his understanding of the problem. And one reason that supplemented decision theories that work for the normative dimension fail to work for the action-guiding dimension is that these theories introduce information that the agent is unable to use in deliberation, e.g., the betterness relation or past preferences. In other words, there is a mismatch here: we (the theorists assessing the agent) have access to different information than the deliberating agent does. If I am correctly
interpreting the way Bermúdez understands these two dimensions in the middle three chapters, then it can’t be the case that a supplemented decision theory only has normative force to the extent that it can be action-guiding. On the contrary, since what we assess includes features the agent himself cannot deliberate about (e.g., his parsing of the decision problem), assessment must be grounded in something other than decision theory’s role in deliberation.

Something has to give: either there is a way to supplement decision theory to answer all the challenges for both dimensions at once, or a theory can be a normative theory without being useful in guiding action. This second possibility seems more compelling to me: although Bermúdez is right that the reason we ought to bring our beliefs and desires in line with the axioms of decision theory is that beliefs and desires are the basis for deliberation, it is not the case that all of the features that are normatively assessed are features we can deliberate about. For example, how the agent sets up the problem is important because of the role the setup plays in allowing the agent to decide on the correct action – but on Bermúdez’s picture, the agent cannot use any decision theory to deliberate about this, so we cannot infer that a theory has normative force in the matter of problem formulation only to the extent that it can guide action in this matter. Alternatively, one might be convinced by the argument of the concluding chapter and might want to preserve the strong link between normative assessment and action guidance. If this is the case, we might resolve the dilemma by concluding that all that should enter into our assessment of an agent is what he does with those facts available to him; consequently, some of the supplemented theories that already work for action guidance will work for normative assessment. Admittedly, this will weaken the notion of normativity involved – it will be more agent-relative than we might have initially wanted – but it is not clear how bad this is: perhaps whatever we think is wrong about, say, formulating a decision problem in a bad way or missing an available option simply falls under the scope of a different charge than decision-theoretic irrationality.

Though I do not have space to discuss it fully here, I believe that a similar problem arises for Bermúdez’s argument that a supplemented decision theory fills the deliberative role only to the extent that it fills the explanatory and predictive role. Briefly, Bermúdez argues that our reason for thinking that any decision theory is a useful tool for deliberation – that the theory is a formalization of our folk notions of belief and desire – would also be a reason for thinking it explains and predicts rational behavior. However, his discussion in the middle three chapters implies that the basis of an action-guiding theory is different from that of a theory of explanation and prediction: like a theory of normative assessment, a theory of explanation and prediction must bring to bear some facts that the agent himself does not have access to. What accounts for the tension between the argument in the concluding chapter and the earlier discussion? One possibility is that although decision theories supplemented for each of the two purposes must each include these folk concepts, we build in different assumptions about the concepts depending on whether we are predicting an agent’s actions or he is engaging in deliberation to choose them: if we want to predict an agent’s actions, then we must make some assumptions about, say, how an agent will choose in novel circumstances, but if the agent is trying to decide what to do, he might need to see this choice as indeterminate. Again, we might stress this difference and note that even if we have reason to think that a supplemented decision theory is useful in guiding action, we might not have reason to think that it can help predict and explain all action; in particular, we might not have reason to think that it can explain novel choice or behavior when preferences are inconsistent over time. Or we might tighten the link between the action-guiding dimension and the explanatory and predictive dimension by weakening the notions of explanation and prediction involved, so that it is not necessary for a decision theory itself to explain and predict novel choice and rational preference change.
This discussion brings out a general tension between Bermúdez’s argument (in the middle three chapters of the book) that each supplemented theory fails to fill some role and his argument (in the last chapter) for the connections between the roles. I worry that the one argument undermines the other – or vice versa. Here’s a reason to think that it is the argument in the concluding chapter that has gone wrong, rather than the earlier argument: if it is correct that a supplemented decision theory can only serve one role to the extent that it can serve the others, then any incompatibility between a supplemented decision theory and one of the dimensions should produce an incompatibility with the other dimensions. So, for example, if adding Schick’s “understandings” doesn’t allow the resulting decision theory to serve the normative role, then it shouldn’t allow the theory to serve the action-guiding role either; but it seems that it can serve this role (and Bermúdez himself claims that it can).

Decision theory considered purely as a formal theory is incomplete: it needs to be filled out with additional principles and its theoretical entities need to be interpreted, in order for it to be suited to the purposes for which it is used. Bermúdez argues that no single way of filling out the theory will make it suitable for more than one purpose. He also argues that any way of filling out the theory so that it can serve one purpose must make it suitable for serving the others. These claims are in tension – Bermúdez recognizes this, which leads to his pessimistic conclusion that decision theory faces a crisis – and I worry that the argument for one claim undermines the argument for the other. However, since both arguments have some intuitive pull, it is well worth attending to Bermúdez’s careful discussion in both parts of the book. Figuring out which assertions we are more committed to will help elucidate what we want decision theory to do under each guise. Consequently, the book gives us not only a thorough assessment of the various challenges facing decision theory, but also a clear way to frame discussion of the relationships among the different purposes decision theory might serve.