Objects of Choice

Wolfgang Schwarz[*]

Forthcoming in *Mind*, 2020

Abstract. Rational agents are supposed to maximize expected utility. But what are the options from which they choose? I outline some constraints on an adequate representation of an agent's options. The options should, for example, contain no information of which the agent is unsure. But they should be sufficiently rich to distinguish all available acts from one another. These demands often come into conflict, so that there seems to be no adequate representation of the options at all. After reviewing existing proposals for how to construe decision-theoretic options and finding them all wanting, I suggest that our model of rational agents should include a special domain of "virtual" option propositions to serve as formal objects of deliberation and choice.

1 Introduction

According to Bayesian decision theory, rational agents choose options that maximize expected utility, relative to the agents' credences and utilities. But even if we know the credence and utility function of a rational agent in a given situation, we cannot predict what the agent will do, unless we also know what options are available: what are the alternatives whose expected utility she is meant to compare?

The question can perhaps be ignored if we understand decision theory as an abstract study of decision problems. But it has to be addressed if the theory is to serve as a psychological model – an idealized, high-level model of the connection between (graded) belief, desire, and action. Such a model may, in turn, be understood in different ways. It could be a descriptive model, purporting to explain choices made by actual humans; it could be a normative model, prescribing the choices people ought to make; or it could be a constitutive model, implicitly defining what it is to be an agent with such-and-such beliefs and desires. On each interpretation, we can't assume that the available options are simply part of a well-defined decision problem.[1]

[*] Thanks to Romy Jaster, Theodore Korzukhin, two anonymous referees, and the audience of seminars at the ANU and Saarland University for helpful comments on earlier versions.

[1] The problem of defining the available options also arises for rivals to Bayesian decision theory; I will stick with Bayesian decision theory for the sake of concreteness and familiarity.

So what are the options available to an agent in a given decision situation? The naive answer is that the options are simply the acts the agent can perform. I think this almost correct. But the acts must be represented in a suitable way. If you don't know that the lottery ticket you are buying is a losing ticket, it would be wrong to represent the chosen act as *buying a losing ticket*.

Intuitively, an adequate representation of an available act should not reveal anything about the world of which the agent is unsure (such as whether the ticket is a losing ticket). However, the representation should also be strong enough to distinguish the act from any other act the agent could choose. As we will see, these demands can come into conflict.

Another challenge arises from the supposed "subjectivity" of decision theory. Assume an agent believes her options are A and B, while in fact they are A and C; each of B and C has greater expected utility than A. Decision theory, as a theory of subjective rationality, arguably should not predict (or prescribe) that the agent does C, given that she is not aware that C is an option. But we also don't want to predict (or prescribe) that the agent does B, which is not in her power.

I will lay out these problems more carefully in the next few sections, in the course of proposing some general constraints on an agent's options. I will then discuss two attractive ideas for how to construe decision-theoretic options. The first identifies options with inner acts of trying, intending, or deciding to perform a certain act. The second construes options as probability measures over a range of acts. I will argue that neither proposal is fully satisfactory. My own proposal will draw on both of these earlier proposals. In short, I suggest that decision-theoretic models should postulate a primitive domain of option propositions to serve as objects of practical deliberation. Every decision an agent can make, and therefore also every act she can perform, corresponds to an option proposition. But that proposition does not transparently describe or represent the decision, in physical or functional terms. It does not transparently describe or represent anything at all.[2]

2 What we can do

Let's start with the simple and intuitive idea that an agent's options are the acts she can perform. As I said, I think this is almost correct – we only need to ensure that these acts are represented in a suitable way. First, however, we need to clarify what we mean by 'act' and 'can'.

Ability modals are notoriously polysemous (see e.g. [Kratzer 1981]). There's a sense in which I can play the piano even when there's no piano around; there's a sense in which I can't accept an invitation if I have conflicting commitments; there's a sense in which

[2] In some respects, the present paper is a dual to [Schwarz 2018], which tackles an analogous puzzle about perception and defends an analogous solution.

determinism implies that we can't do anything except what we actually do. None of these senses are relevant to decision theory.

To home in on the relevant sense of 'can', compare falling out of a helicopter with reaching a junction on a hike. When you fall out of a helicopter, you will fall to the ground, and there is nothing you can do about that. By contrast, when a hike leads you to a junction, whether you end up turning left or right is sensitive to your psychological state: to your beliefs and desires, and to how these are put together to reach a decision. In that sense, what you do is under your control.

These are the kinds of choices studied in decision theory. They do not require a strong, libertarian kind of freedom. Decision theory is widely used in artificial intelligence, where it is taken for granted that the (artificial) agent's actions can be predicted from its internal architecture and the inputs it receives.

As a first pass, we might say that an agent *can* ϕ, in the sense relevant to decision theory, iff some variation of the agent's psychological state would bring about that the agent ϕs. However, on that conception, *believing that Sydney is the capital of Australia* is something I can do, for there is some variation of my psychological state – namely my state of belief – that would result in me having this belief. But my beliefs about capitals are not in the right way under my volitional control. I can't make myself believe that Sydney is the capital of Australia simply by an act of the will.

To define the acts that are under an agent's volitional control, we need to invoke another type of mental state (or event) besides beliefs and desires. Intuitively, when an agent's beliefs and desires cause her to perform a certain act, they do so by first causing a *decision*, which in turn causes the act. Let's assume that rational behaviour always involves such decision states mediating between the agent's standing attitudes and the resulting actions. If an agent faces a choice, then the decision she makes is sensitive to her beliefs and desires: if the agent had suitably different beliefs and desires, she would make a different decision. Let's say that a decision is *available* if some possible variation of the agent's beliefs and desires would lead to the decision.

As a second pass, then, I suggest that an agent *can* ϕ (in the sense relevant to decision theory) iff some available decision would bring about that the agent ϕs. This is still a little vague[3], and it might need a few epicycles to deal with various trouble cases[4], but the required complications would (I hope) not affect the issues I want to discuss; so I

[3] For example, I have not specified what counts as a 'possible variation' of the agent's beliefs and desires. If you have severe arachnophobia, is wanting to touch the spider on the wall a 'possible variation' of your desires?

[4] Trouble might arise, for example, if there is an indeterministic connection between attitudes and decisions, or between decisions and acts, so that there is no decision that *would* come about as a result of certain attitudes, or no act the agent *would* perform as a result of a certain decision. Relatedly, there are difficult questions about how these counterfactuals should be understood. We should probably exclude 'backtracking' interpretations, as made vivid in [Gallois 2009].

will move on.[5]

Next, 'act'. Suppose one Sunday afternoon you come across a blueberry shrub and decide to try one of the berries. So you *eat a berry from the bush*. You thereby also *eat a blueberry from the bush*, and you *eat a berry on a Sunday afternoon*. Action theorists debate whether these are three different acts or three descriptions of a single act. Without purporting to settle any dispute about the ordinary conception of acts, I will adopt the second usage: your act of eating a berry is the very same act as your act of eating a blueberry.

On this usage, we can't assume that rational agents have perfect knowledge about every aspect of the acts they are performing. You may not know that it is Sunday, or that the berry you're trying is a blueberry.

This matters for decision theory. Let's say you give some credence to the possibility that the berry is a poisonous tutsan berry. When you consider eating the berry, you should take this possibility into account. That is, even though the act you perform is (in fact) an act of eating a blueberry, this is not how it should be represented in your decision problem. To assess the expected utility of your choice, we must somehow bracket the fact that the berry is a blueberry.

So we must distinguish the all-inclusive *acts* an agent can perform from the *options* whose expected utility should guide the agent's choice. The option you choose should be neutral on whether the berry is a blueberry or a tutsan berry.

So what kind of thing is an option? The two main answers in Bayesian decision theory go back to Leonard Savage [1954] and Richard Jeffrey[6] [1965].

Savage models options as abstract functions from "states of nature" to "outcomes". Your act of eating the blueberry, for example, could be represented by a function mapping states in which the berry is a blueberry to the pleasant outcome of eating a blueberry, and states in which the berry is a tutsan berry to the less pleasant outcome of eating a tutsan berry. As desired, this representation no longer reveals that the berry is a blueberry.

Jeffrey rejects the distinction between acts and states of nature: "the human agent is taken to be part of nature and his acts are thus ingredients in states of nature" [Jeffrey 1992: 226]. Formally, Jeffrey models states, acts, and outcomes all as propositions. Your act of eating the berry might be represented as the proposition *that you eat a berry from*

[5] You may wonder why I haven't adopted the more familiar conditional analysis, on which an agent can ϕ just in case she would ϕ if she decided (or tried, or intended) to ϕ. Part of the answer is that I want to avoid some well-known problems for the conditional analysis (see [Maier 2018]). I also don't want to assume that decisions are individuated as propositional attitudes with a particular content ϕ. This will become important later.

[6] Jeffrey actually has two accounts of options. Here I present the first; I will discuss the second in section 6.

the shrub. Again, this leaves open whether the berry is a blueberry or a tutsan berry. To choose an option, on Jeffrey's model, is to make the relevant proposition true.[7]

Jeffrey's framework has some advantages over Savage's, which is why it will be assumed in what follows. For example, if acts, states, and outcomes are all propositions, we can take into account logical and probabilistic connections between these elements, which is put to use not only in Jeffrey's own formulation of decision theory but also in many of its rivals, including [Gibbard and Harper 1978], [Lewis 1981], [Sobel 1986], [Skyrms 1984], and [Joyce 1999]. Without appeal to such connections, it is hard to give a satisfactory account of Newcomb Problems (as argued in [Joyce 1999: 117ff.]), or to model the way rational belief changes through deliberation (see [Skyrms 1990], [Joyce 2012]).

So here is where we stand. For any agent in a concrete choice situation, there is a range of acts the agent can perform, in the sense that there is an available decision that would bring about the act. These acts may be understood as the agent's "objective options": they provide a complete and accurate account of what the agent can do. Real agents rarely have perfect information about their objective options. Hence we need a more "subjective" conception of options to serve as objects of expected utility. Following Jeffrey, I will assume that subjective options are propositions. But this doesn't tell us *which* propositions should count as (subjective) options in a given choice situation. Is the proposition *that you eat a berry from the shrub* a suitable option in the berry situation? If so, why? If not, why not? What are the general rules?[8]

3 Options and actions

We might expect there to be some connection between an agent's decision-theoretic options and the acts she can perform (her objective options). Let's try to get clear about this connection.

To begin, it is natural to assume that every (subjective) option is a correct, but typically incomplete, description of some act the agent can perform. For example, the proposition *that you eat a berry from the shrub* is an incomplete description of the act you choose in the berry scenario. Let's call this the *Ability condition* on decision-theoretic options.

> **Ability.** A proposition A is an option (for an agent in a given choice situation) only if the agent can make A true (in the sense that there is an available decision that would render A true).

[7] By a 'proposition', I mean the kind of thing that is in the domain of an agent's subjective probability function. To avoid stylistic awkwardness, I will often use infinitival verb phrases ('eating a blueberry') instead of that-clauses ('that you eat a blueberry') to express candidate option propositions, but nothing important will hang on that – though see [Perry 1979] and [Lewis 1979] for arguments that verb phrases are actually more perspicuous representations of the relevant propositions.

[8] In Savage's framework, the analogous question is which functions from states to outcomes should count as options. I will briefly return to this question in section 9.

To illustrate, suppose you face a choice between turning left and turning right. These are your objective options. But suppose, in violation of the Ability condition, an adequate representation of your decision problem includes a further option, which no available decision would render true – say, the proposition that you lift off into the sky. Depending on your beliefs and desires, this option may maximize expected utility. And then decision theory will predict or recommend that you lift off. But you can't. By assumption, any decision you can make will bring about that you turn left or that you turn right. It seems wrong to say that you are irrational for not lifting off. The Ability condition ensures that the decision-theoretic 'ought' implies 'can'.

Another natural condition goes in the other direction, demanding that different acts an agent can perform should be represented by different options.

> **Cover.** For any proposition B which the agent can make true (in the above sense) there is some option A such that, if the agent were to choose A then B would be the case.

The Cover condition ensures that if you face a choice between turning left and turning right, then an adequate decision matrix includes an option whose choice would make you turn left and another that would make you turn right. It would not be OK if the matrix only had an option for turning left.

Consider what it would take, in general, for the Cover condition to fail. There would have to be some proposition B that would become true as a result of some available decision D, yet no option A in the agent's decision problem would, if chosen, lead to the truth of B. It follows that no option in the agent's decision problem would, if chosen, result in decision D. In effect, the Cover condition therefore says that every "available decision" is the result of choosing some option in the agent's decision problem. What else could an "available decision" be?

Another condition is needed to rule out a certain kind of redundancy among an agent's options. If you have the option of raising your left hand and the option of raising your right hand, then we may not want to count *raising a hand* as an additional, third option. This could be achieved by the following condition.[9]

> **Maximality.** A is an option only if there is no other option that entails A.

Together, Ability, Cover, and Maximality paint the following picture of the connection between subjective and objective options.

Take an agent in a concrete decision situation. For any available decision D, there is a comprehensive proposition $@D = \bigcap\{P : D \mathrel{\Box\kern-1.5pt\rightarrow} P\}$ specifying everything that would be

[9] The Maximality condition is often assumed in discussions of (objective) moral permission and obligation, see e.g. [Brown 2018]. Not all arguments for and against adopting the condition in that context carry over to the present context; see fn. 12 below.

the case if that decision were made. Each act the agent can perform – each objective option – is completely described by some such proposition. On some accounts, @D picks out a single possible world, on others it picks out a narrow range of worlds; we can remain neutral on this.[10] In any case, the objective options will form a set of disjoint, tiny regions in logical space. The agent's subjective options are a "widening" of this set, expanding each region while keeping the regions disjoint.[11]

The remaining task is to say *how* the regions should be expanded – intuitively, how the acts an agent can perform should be described by her options.

To be clear, I have not established that this is the right way to approach an agent's options. I have not given any proofs of the Ability, Cover, and Maximality condition. Nor could I. All I claim is that these conditions have some intuitive appeal, and that they combine to an elegant picture of how subjective options are related to objective options.

Let's return to the "remaining task", of saying how the acts an agent can perform should be described by her options.

We have already seen that some descriptions are too narrow: they include too much information about the relevant act. In the berry scenario, where you don't know whether you're looking at blueberries or tutsan berries, the proposition *that you eat a blueberry*, for example, contains too much information. If this isn't obvious, let's look at the computation of expected utilities.

In "evidential" decision theory (see [Jeffrey 1965]), the expected utility of an option X can be defined as the credence-weighted average of the utility of any possible outcome O conditional on X:

$$EU(X) = \sum_O U(O)Cr(O/X).$$

Let T be the undesirable outcome of having consumed a tutsan berry. If B is the proposition *that you eat a blueberry*, then $Cr(T/B)$ is plausibly close to zero; so the badness of possible poisoning does not figure in $EU(B)$. If we treat B as one of your options, it could easily maximize expected utility, even if you are pretty sure that the berry is a tutsan berry. In that case, however, the right thing to do in light of your beliefs and desires is clearly not to eat a berry. So B is not one of your options.

[10] The issue turns on questions like the following. Suppose D is an available decision that would cause you to toss a coin, and you don't actually choose D; if you had chosen D, would the coin have landed heads? Would it have landed tails? If the answer is 'no' both times (by whatever standards the counterfactual is best understood in the present context), then @D will be true at worlds where the coin lands heads and also at worlds where the coin lands tails; so it will not pick out a single world (compare [Vessel 2003]).

[11] David Lewis [1981: 308] seems to have a similar picture in mind. He suggests that an agent's options are given by the partition of logically strongest propositions which the agent "can make true at will". Lewis does not explain this crucial phrase, nor why we should expect the relevant propositions to form a partition.

The same problem arises in "causal" decision theory, where expected utility may be defined in terms of the credence of the outcomes on the *subjunctive* supposition that the relevant option *would be* chosen (see [Joyce 1999]; the backslash indicates subjunctive supposition):

$$EU(X) = \sum_O U(O)Cr(O\backslash X).$$

Again, $Cr(T\backslash B)$ is plausibly around zero, and so $EU(B)$ does not take into account the possibility of poisoning.

The proposition *that you eat a blueberry* is too specific, providing too much information about the relevant act. Other propositions are too unspecific, providing too little information. The proposition *that you eat something*, for example, is too unspecific, as it does not settle whether you eat a berry from the bush or a cracker from your bag. One might hope that the combination of Cover and Maximality rules out such propositions. But propositions can also be too specific without violating the above conditions. For a simple (albeit artificial) example, consider a situation in which you are rewarded for unintentionally closing your eyes but punished for intentionally closing them. Having no means of causing yourself to unintentionally close your eyes, you rationally decide not to close your eyes. As a result, the most likely circumstances under which you would close your eyes might be circumstances under which you do so unintentionally. *Closing your eyes* then has comparatively high expected utility (in both evidential and causal decision theory, as $Cr(\textit{Reward}\,/\,\textit{Close})$ and $Cr(\textit{Reward}\setminus\textit{Close})$ are both high). So if we count *closing your eyes* as an option, we would wrongly conclude that it would have been rational to choose it.[12]

Some act descriptions are too specific, others are too unspecific. What, in general, should be included in an option proposition?

4 The agent's perspective

Here is a tempting thought. In the berry scenario, the proposition *that you eat a blueberry* is not an adequate description of the chosen option because you don't know that you're looking at a blueberry shrub. Decision theory is supposed be a 'subjectivist" or "internalist" theory, telling us what an agent should do *from her own perspective*, *in*

[12] Similar examples are discussed in [Weirich 1983: 179] and [Hedden 2012: 354f.]. This kind of case also motivates the Maximality condition: if we treat both *closing* and *intentionally closing* as options, the former might come out as maximising expected utility. The literature on Maximality mostly focuses on cases in which a logically weaker option is an (exclusive) disjunction of more specific options. This makes the issue rather subtle, as one might conjecture that a disjunctive option has maximal expected utility only if all its disjuncts do, in which case including the disjunctive option would do no harm. [Sobel 1983] argues that this conjecture is true in evidential decision theory, but not in causal decision theory

light of her beliefs and desires. The recommendations of decision theory therefore should not draw on external matters of which the agent is unaware.

This thought can be fleshed out in different ways. We might, for example, impose the following condition.[13]

> **Modal Certainty.** A proposition is an option for an agent only if the agent assigns credence 1 to the hypothesis that she can make it true.

In the berry scenario, you are not certain that you can (make it true that you) eat a blueberry. So Modal Certainty correctly rules out *eating a blueberry* as an available option.

Instead of demanding credence 1, we could demand that the agent *knows*, or *believes* that she can make the relevant proposition true, appealing to a non-Bayesian concept of binary knowledge or belief (compare [Hedden 2012: 348]). I will not explore the pros also and cons of these alternatives, except to note that as long as you give positive credence to the tutsan berry hypothesis, it seems wrong to disregard this possibility when computing expected utilities. So mere knowledge or belief arguably won't suffice, unless they entail credence 1.

On reflection, it is not clear (to me) how these proposals are supposed to follow from the internalist character of decision theory. Loosely speaking, internalism demands that the available options should be "accessible" to the agent; the rules of decision theory should in principle be implementable by a local computational process whose output isn't directly sensitive to facts about the agent's environment, unmediated by the agent's beliefs. But this doesn't mean that the agent must be certain of (or know) her options. Compare beliefs and desires. These surely count as "accessible" in the relevant sense. However, a rational decision-maker need not be certain of her own beliefs and desires (see [Skyrms 1980]). Why then would she have to be certain of her options?

Admittedly, some authors seem to think that agents who conform to decision theory must consciously compute the expected utilities of all their options. This might require some kind of knowledge of the available options, and of the relevant credences and utilities. But that interpretation of decision theory makes the theory completely implausible as a descriptive, normative, or constitutive model. Real people don't consciously compute expected utilities whenever they face a choice, and there is no reason to think that they should, or that ideally rational people would.[14]

Support for Modal Certainty may come from the study of rational deliberation. Intuitively, if you know that you can raise your arm, and you decide to raise it, then you

[13] In print, Modal Certainty has been most prominently defended by J.H. Sobel; see [Sobel 1980: 178–80], [Sobel 1983: 199f.], [Sobel 1986: 155f.], [Sobel 1988: 8].

[14] Here I agree with [Pettit 1991: 167–169], [Jackson 1991: 468–471], [Maher 1993: 5–8], [Joyce 1999: 80], and many others.

know that your arm will go up; no further observations are needed. Decision-making seems to provide "knowledge without observation", as Elisabeth Anscombe memorably put it ([Anscombe 1957: sec.8]).[15] In the framework of Bayesian decision theory, the effect of decision-making on rational credence has been most thoroughly studied by Brian Skyrms (see esp. [Skyrms 1990]). Skyrms's models imply that deciding in favour of an option goes along with becoming certain of that option. This matches standard models in artificial intelligence, which assume that when an agent makes a choice then her probability function is conditionalized on the chosen option (see e.g. [Russell and Norvig 2010: ch.17]).

We can turn these models around (as noted in [Jeffrey 1968]): if you couldn't become certain of a proposition merely through rational deliberation, then the proposition isn't one of your options. So we get the following constraint.

> **Deliberational Certainty.** A proposition A is an option for an agent only if the agent could rationally come to give credence 1 to A merely through making a decision.

Like Modal Certainty, Deliberational Certainty correctly implies that *eating a blueberry* is not an option in the berry scenario: since you're not sure if the berry is a blueberry, you couldn't rationally become certain that you'll be eating a blueberry merely on the basis of making a choice.

Deliberational Certainty looks somewhat better supported than Modal Certainty, being implied by standard models of Bayesian decision-making. One might, however, argue that Deliberational Certainty implies Modal Certainty. For suppose a proposition fails Modal Certainty, so that the relevant agent is unsure whether she can make the proposition true (perhaps because that depends on external facts outside the agent's control). Arguably, the agent then couldn't become certain merely through deliberation that she will make the proposition true. So the proposition also fails Deliberational Certainty.

I am not convinced by this argument, but I won't dwell on the matter. The converse direction clearly fails: Modal Certainty does not entail Deliberational Certainty. For example, suppose you have a choice between two berries; you know that one is a tutsan berry, the other a blueberry, but you don't know which is which. *Eating the blueberry* then plausibly satisfies Modal Certainty, at least if we understand 'can' along the lines I proposed in section 2: you may be certain that there is an available decision that would cause you to eat the blueberry. On the other hand, you could not rationally become certain that you will eat the blueberry merely through deliberation. So Deliberational

[15] Anscombe argued that intentionally performing an act implies knowing (without observation) that one performs the act. The condition I am about to spell out is related to, but logically independent of this claim.

Certainty (correctly) rules out *eating the blueberry* as an option; Modal Certainty does not.[16]

In any case, both Modal Certainty and Deliberational Certainty have an obvious flaw. Credence 1 is not easy to come by; many prominent Bayesians have even suggested that contingent propositions should never have credence 1 (see e.g. [Hájek 2012]).

Consider again your choice of trying a berry. By Deliberational Certainty, there would have to be a non-trivial proposition describing this act of which you become absolutely certain, and rationally so, merely through deliberation. What might that proposition be? Whatever it is, couldn't you have (perhaps misleading) evidence that the proposition is false? Would it really be rational to bet your life on it?

Similarly for Modal Certainty. Here we have to assume that there is a non-trivial proposition describing your act of which you are absolutely certain that you can make it true. But whatever that proposition is, it is a contingent fact that you can make it true; there are possible situations in which you can't make it true. Why must you be in a position to conclusively rule out these situations?[17]

The problem is exacerbated by the Cover condition, which requires that every act an agent can perform is represented by a distinct option proposition. An example from [Jeffrey 1968: 37] brings out the problem. Jeffrey considers an experienced marksman aiming at a distant target. The marksman's posture and the direction in which he is pointing are carefully attuned to his information about the wind, the quality of his rifle, and so on. He has fine-grained control over how he holds his rifle. He is, in effect, choosing between thousands of torque configurations in his joints (although that is obviously not how he conceptualizes his options). To model the marksman's decision problem, we need thousands of option propositions. By Modal Certainty, there must be thousands of relevant, distinct propositions of which the marksman is certain that he can make them true. By Deliberational Certainty, there must be thousands of such propositions of which the marksman could become rationally certain through deliberation. What could these propositions be?

One might respond that the marksman does not really face a choice between thousands of options. Perhaps he merely chooses to *aim at the target*, and some sub-personal process

[16] Modal Certainty might rule out the option if we stipulate that 'can' (in Modal Certainty) must be understood, say, in accordance with the conditional analysis. But that would lead to other problems. For example, arguably an agent could have the option of *intending to travel to Paris* even if she doesn't have the capacity to intend to intend to travel to Paris; Modal Certainty would implausibly demand that the agent must nonetheless be certain that if she intended to intend to travel to Paris, then she would succeed.

[17] One might try to avoid the problem by adopting a contextualist account of "credence 1" (as in [Greco 2017]). The idea would be that even though falsifying scenarios for the relevant propositions – say, about your abilities – exist, and are not ruled out by your evidence, they can often be ignored for the purpose of a given conversation; if they are, we can truly say that you assign rational credence 1 to these propositions. But what are your options if we don't ignore the falsifying scenarios?

selects the appropriate torque configurations in response to the marksman's information about the quality of the rifle etc. But this doesn't seem right. There are different ways of aiming at the target, and the marksman has voluntary control over these ways. That he points the rifle in a particular direction, as opposed to the opposite direction, is a voluntary choice; it has a rational explanation in terms of the marksman's goals and information.[18]

To be sure, the marksman may not consciously think to himself that he is choosing this option rather than another. But as I said before, I do not think decision theory is usefully understood as a theory of conscious decision-making. Decision theory tells us which of the options available to an agent best fit the agent's goals in light of the agent's information. It does not matter what, if anything, goes through the agent's mind as they make their choice.

Let me sum up the problems we have encountered. First, to secure the desired internalism of decision theory, we would like to adopt a constraint along the lines of Modal Certainty or Deliberational Certainty. But these conditions imply that a decision-maker must be absolutely certain of contingent propositions which don't seem to be conclusively established by the available evidence. Let's call this the *evidence problem*.

The evidence problem is related to what I'll call the *cover problem*. This is the problem that we need thousands of distinct options for the marksman's decision problem. More generally, the problem is that the specification of an agent's options is subject to opposing constraints. On the one hand, an adequate representation of an agent's options should not turn on external facts of which the agent is unsure. Since rational agents may be unsure about many things, this means that the option propositions may have to be rather weak. On the other hand, any option proposition must distinguish the act it represents from every other act the agent can perform. This means that the option propositions have to be rather strong, especially if the agent has a lot of options.

A third, but again related problem arises from our assumption that a proposition is an option for an agent only if the agent can make the proposition true (by the Ability condition). Arguably, whether an agent can make a given proposition true always depends on cognitively external matters of which the agent may be unsure or ignorant. The Ability condition therefore seems to clash with the desired internalism of decision theory.[19] Let's call this the *ability problem*.

[18] I am not saying that every aspect of the marksman's movements is under his voluntary control. But what's under his control is a lot more fine-grained than *aiming at the target*.

[19] Sobel ([1980], [1986]) stipulates that a rational agent is certain that she can ϕ only if she really can ϕ. The Ability condition would then be entailed by the (internalistically acceptable) Modal Certainty condition. But why should we accept Sobel's stipulation, especially if we assume (with Modal Certainty) that rational agents often assign credence 1 to contingent propositions about what they can do? ([Hedden 2012: 352f.] offers a possible explanation; see footnote 22 below.)

5 Decisions

We might hope to escape the problems from the previous section by construing an agent's options not as (propositions describing) overt acts, but as (propositions describing) internal events of *trying*, *intending*, or *deciding*. Proposals along these lines have been made in [Sobel 1971], [Sobel 1983], [Weirich 1983], [Joyce 1999: ch.2], and [Hedden 2012] (see also [Pollock 2002] for critical discussion).

It is easy to see why this move can seem appealing. Right now, for example, I'm pretty sure that I could, if I wanted, raise my arms. But I'm not entirely sure. There's a small chance that the muscles in my arms have just stopped working. I'm not even sure that I *have* arms: a small part of my credence goes to the hypothesis that I'm an armless brain in a vat. As a consequence, *raising my arms* should not count as one of my options. For suppose I give low utility to the skeptical scenarios in which I can't raise my arms. If *raising my arms* were an option, it would probably have greater expected utility than not raising my arms simply because it rules out the undesirable skeptical scenarios in which I'm paralysed or envatted. (Since *Raise-Arms* entails $\neg$*Brain-in-Vat*, $Cr(\neg \textit{Brain-in-Vat}/\textit{Raise-Arms})$ and $Cr(\neg \textit{Brain-in-Vat}\backslash \textit{Raise-Arms})$ are both 1.) But surely my desire not to be paralysed or envatted gives me no reason to raise my arms. I'm not irrational for not raising my arms.

So raising my arms is not one of my options. Intuitively, I can *try* or *decide* or *intend* to raise my arms, but whether I succeed depends on whether I'm a brain in a vat and on whether my muscles still work – things I can't simply make true or false by an act of the will.

There are other reasons to allow for intentions or decisions as options (as highlighted in [Weirich 1983]). We frequently make decisions not only about what we are going to do right now, but also about what we are going to do later: how we'll spend the afternoon, what we'll cook for dinner, where we'll go on holiday. The immediate outcome of these decisions is not an overt act, but an intention. Indeed, sometimes the main reason for deciding to perform an act lies not in features or consequences of the act, but in features or consequences of the decision. If you are anxious about a certain choice, it may be reasonable to make a decision, just to stop worrying and calm your mind.

With respect to the problems from the previous section, we might hope that if options are decisions, then we have an automatic guarantee that every available decision corresponds to an option – which solves the cover problem. Moreover, while overt acts can always be thwarted by unknown external forces, we might hope that rational agents at least have full and certain control over what they decide or intend. This may seem to get around the ability and evidence problem.

But let's have a closer look. To begin, decisions, like overt acts, can be described in many ways, and the mode of presentation matters to their expected utility. (In the berry situation, your decision to try a berry is a *decision that causes you to eat a blueberry*, but that's not an adequate mode of presentation.) So again we have to decide how the available decisions should be represented or described by the agent's option propositions.

In ordinary language, decisions or intentions are commonly represented as having a subject and an object: an agent S decides or intends to ϕ, where ϕ is an infinitival clause characterising a type of act. On this way of individuating decisions, it is doubtful that we have escaped the three problems.

Return to the marksman. The marksman presumably intends to hit his target. Does he decide to hit it? Maybe. Or maybe he decides to shoot at it. But these propositions are far too unspecific. They don't single out the marksman's choice among the thousands of nearby alternatives. If the wind picks up, the marksman adjusts the orientation of his rifle. What is the option he thereby chooses? Does he *decide to adjust his joint torques in manner* $\vec{V}$, or *to hold his rifle in direction* α, β, γ?[20] Perhaps there is a sense in which he does. But in that sense, the marksman surely won't know what option he is choosing, nor what options he *can* choose.[21] If we don't want to make the available options completely opaque to the agent, it looks as though we're stuck with the cover problem.

We're also stuck with the evidence problem, even in simpler cases where there are only a few options. Suppose you face a choice between turning left and turning right. If you decide to turn left, why should you become absolutely certain that this is what you decided, merely on the basis of the decision? Couldn't you be unsure about the content of your decision? Couldn't you be unsure that you really made the decision? Similarly, couldn't you be unsure that you have the ability to decide to turn left, perhaps because you're unsure that you have the required willpower, or because you're unsure if you are a genuine decision-maker? If not, why not? What part of your evidence conclusively rules out the falsifying scenarios?

The Ability problem also remains open. Whether an agent can make a given decision generally depends on external matters of which the agent may be unsure. It may depend on whether the agent has enough energy, or on the absence or presence of a demon poised to intervene as soon as he detects that the agent is about to make a certain choice (compare [Frankfurt 1969]).[22]

[20] Orientations in 3D space are commonly specified by three angles, known as "Euler angles".

[21] Suppose we offered the marksman a prize if he decides to adjust his joint torques in the precise manner $\vec{V}$. Decision theory says that if A is one of your options, and you are certain that choosing A leads to a desirable outcome (and there are no better options, etc.), then you should choose A. So if we take the marksman's options to include a decision to adjust his torques in manner $\vec{V}$, we'd have to say that he would be irrational if he didn't win the prize. Clearly that's wrong. Rationality does not require knowing the precise torque configurations or angles determined by one's choices.

[22] In [Hedden 2012: 352f.], Brian Hedden argues that 'on any attractive theory of decisions', which decisions an agent can make does not depend on external facts of which the agent may be ignorant. By way of illustration, he considers a view on which an agent *can decide to* ϕ just in case the agent does not believe that if they decided to ϕ then they would fail to ϕ. This makes abilities to decide an

I believe that the options-as-decisions idea is on the right track. But it needs a further tweak to provide a fully satisfactory model. Before I describe that tweak, it will be useful to review a very different answer to the problems we have encountered.

6 Jeffrey's model

In [Jeffrey 1965: ch.11] and [Jeffrey 1968], Jeffrey observed that the problem of options resembles a well-known problem in Bayesian models of learning or perception. The classical Bayesian account of learning assumes that the agent becomes absolutely certain of some evidence proposition E, comprising the totality of what the agent learns. But what is that proposition, in ordinary situations? When I look at the cloudy sky, I do not become certain that it is cloudy. (Recall that I give some credence to the hypothesis that I'm a brain in a vat.) Here, too, it is tempting to retreat from external-world propositions to propositions about a more secure internal world: perhaps what I learn with certainty is that it *appears to be cloudy*. But couldn't we also be mistaken or unsure about how things appear? Moreover, ordinary propositions about how things appear are often too unspecific to explain the change in our beliefs prompted by a perceptual experience. My credence about the weather is sensitive to subtle details in my perceptual experience. We would need thousands of subtly different appearance propositions of which I could become absolute certain.

In response, Jeffrey proposed a new model of learning. In Jeffrey's model, a learning event may directly confer non-trivial probabilities $x_1, \ldots, x_n$ to a partition of propositions $E_1, \ldots, E_n$, without rendering anything certain. This kind of belief update has come to be known as *Jeffrey conditionalization* and is now a staple in Bayesian epistemology.

Jeffrey made a parallel, though largely forgotten, proposal about an agent's options. Recall that on the classical model of deliberation, choosing an option goes along with becoming certain of that option. Jeffrey suggests that choices may instead go along with assigning non-extreme probabilities $x_1, \ldots, x_n$ to a partition of propositions $X_1, \ldots, X_n$. This probability assignment, rather than any particular proposition, represents the chosen option.

Informally, the idea is that our choices often amount to a kind of gamble. When you decide to eat a berry, you don't know what you will end up doing. You might end up eating a blueberry, or a tutsan berry; you might be struck by lightning, or remain a

internal matter. But the proposed interpretation of 'can decide' is highly counter-intuitive. In fact, Hedden himself acknowledges (on p. 354) that in cases where a demon would prevent the agent from deciding to ϕ, the agent *lacks* the ability to ϕ – even if the agent believes she can ϕ. Hedden argues that decision theory is not applicable to such cases. I'm inclined to agree (see p. 21 below); but the mere possibility of these cases proves the opposite of Hedden's claim: on any attractive theory of decisions, which decisions an agent can make generally depends on external facts of which the agent may be ignorant.

brain in a vat. In Jeffrey's model, what you choose is represented by an assignment of probabilities to these eventualities.

I think there is something deeply right about Jeffrey's model. But it also has serious costs. For one thing, if options aren't propositions, the expected utility of an option can no longer be defined (as in section 2) in terms of the agent's credences conditional on the option. We would need to revise our definition of expected utility. I won't dwell on how this might work, for there is a more fundamental problem. There is a more fundamental problem.

We want to know what options are available to an agent in a concrete decision situation. Let's assume there are a number of acts the agent can perform, in the sense I outlined in section 2. Our question is how these acts should be converted into decision-theoretic options. In Jeffrey's model, an option is a probability function over some partition of propositions. So the question becomes: how should the available acts be converted into probability functions over partitions?

The answer clearly depends on the agent's beliefs. Your credence that you will eat a blueberry depends on your prior beliefs about whether you have arms and whether the berry is a blueberry. So we need to explain how the acts an agent can perform, in combination with her prior beliefs, determine the relevant "Jeffrey options": the available probability functions over partitions of propositions. And that turns out to be hard.[23]

To see why, consider a simpler agent, of the kind studied in artificial intelligence: a robot with an internal representation of its environment, some goals, and a capacity to move around. If the robot's decision module figures out that moving to the left would be useful, a command is sent to its motor system that normally causes movements to the left. Whether the robot succeeds in moving to the left depends on various external factors such as the slipperiness of the floor or the presence of a glass wall blocking the way. Let's say the robot assigns probability 0.2 to such interfering factors. *Moving left* should then not count as one of its options (for the same reason for which *eating a blueberry* should not count as an option in the berry scenario.). Following Jeffrey, the robot may instead have an option of *moving left* with probability 0.8 and *staying in place* with probability 0.2. But how is the robot's decision module supposed to figure out that this is one of its options – the option it could realize by sending a specific motor command? If the robot receives information increasing the probability of a slippery floor, the robot's options will change. Perhaps it now only has an option of *moving left* with probability 0.5 and *staying in place* with probability 0.5. Where do these probabilities come from?

They can't be hard-coded into the robot's cognitive architecture. Suppose we programmed the robot so that the option that would be realized by sending the motor command for moving to the left is represented as *moving left* with probability

[23] The problem resembles the "input problem" for Jeffrey conditionalization, see e.g. [Field 1978], [Garber 1980], [Christensen 1992], [Weisberg 2009].

$Cr(\neg \mathit{Wall} \wedge \neg \mathit{Slippery}) + 0.5 \times Cr(\neg \mathit{Wall} \wedge \mathit{Slippery})$ and *staying in place* with probability $Cr(\mathit{Wall}) + 0.5 \times Cr(\neg \mathit{Wall} \wedge \mathit{Slippery})$. That would get the present example right, but it would be too inflexible. The robot should be able to figure out under what conditions issuing the relevant motor command is likely or unlikely to have the desired effect, and it should be able to revise these judgements in light of new information. How is that supposed to work?

7 Inventing options

I will now present a new model of decision-theoretic options. The model combines elements of Jeffrey's model with a tweaked version of the options-as-decisions model from section 5.

Let's stay with artificial agents for a moment. The task of a robot's decision module is to select an appropriate motor command. It is tempting to use the available motor commands as the relevant options. That is, suppose each available motor command X is represented by a special element X^* in the domain of the robot's probability function. (Motor commands are electrical signals, not propositions, so I assume they aren't themselves objects of probability.) These elements stand in probabilistic relations to propositions describing overt acts. For example, if L is a particular motor command, the robot might assign high probability to the proposition that it will move left conditional on L^*. But this probability can change. The robot can learn under which conditions choosing L^* leads to movements to the left and under which it does not.

In rough outline, this is indeed how options are commonly represented in artificial intelligence (see e.g. [Russell and Norvig 2010: ch.s 2 and 16–17]). Somewhat more precisely, the agent is assumed to have a "transition model" defining conditional probabilities over new states of the world given a present state and a given motor command. For example, if S is a slippery-floor state, S' is a similar state in which the robot is a little further to the left, and L is the motor command for moving to the left, the transition model might assign middling probability to S' conditional on S and L (or rather, L^*, although the distinction between L and L^* is rarely made in theoretical computer science). These conditional probabilities can be adjusted through learning: given partial observation of a post-action state S', the robot can update its beliefs about the likely effect of a given motor command in a given pre-action state S.

I want to suggest that all rational decision-making should be understood along these lines. To be sure, human decision-making is more complex. When we choose to move our limbs, a whole hierarchy of control systems appears to be in play. This presents an independent challenge to classical decision theory, which assumes a single interface between beliefs and desires on the one hand and actions on the other. The challenge could be met in different ways; I'll adopt the lazy response of idealising away the hierarchy, treating it as a lower-level mechanism for selecting the final motor commands. Humans can also make decisions whose output isn't a motor command, as when we decide to travel to Paris in the summer, to visualize a scene, or to focus on our breath. To accommodate these decisions, we need to generalise the notion of a motor command.

The generalized model assumes that whenever we make a choice, our beliefs and desires combine to determine a *control state* which may in turn cause relevant body movements, changes of attention, new commitments, or whatever. I have already appealed to these control states in section 2, when I defined the acts an agent can perform. There I called them 'decisions'.

On this usage, decisions are not assumed to be conscious or introspectible, and they are not assumed to be individuated by a subject and an intentional object. A decision is a state that plays a certain functional role, namely to control the agent's behaviour in accordance with her beliefs and desires.

A complete model of a rational decision-maker would have to include a specification of the available control states. (For simple agents, the available control states might be the same in all decision situations.) Each control state X should, I suggest, be represented by a special element X^* in the domain of the agent's subjective probability function. These elements are the agent's decision-theoretic options. They are what the agent deliberates over. If the process of deliberation selects an element X^*, the agent becomes certain of X^* and the corresponding control state X is instantiated.

X^* is not an ordinary proposition. It is not expressed by a sentence of the form 'I decide [or intend] to ϕ'. Nor does it characterise the associated control state X in lower-level physical terms. If you decide to raise your arm, you normally don't become certain of the neurophysiological events that trigger the action, and that's not a failure of Bayesian rationality.

There are no ordinary propositions of which we could become rationally certain just through making a choice – especially if these propositions are supposed to be sufficiently fine-grained to distinguish all available options. So we have to introduce new propositions. That is, we have to extend the domain over which an agent's probabilities are defined, introducing new elements to play the role of options.[24] The propositions X^* associated with control states are primitive propositions in the extended domain, logically independent of the old propositions. The sense in which X^* *represents* X is merely causal: there is a reliable mechanism by which X is instantiated whenever the agent chooses the option X^*.

More concretely, suppose we model an agent's credence function as defined over sentences in a suitable language. Option propositions are then simply new sentences that are not equivalent to sentences in the old vocabulary used to represent ordinary

[24] Formally, the extended domain will be the algebraic product of the original domain and the algebra generated by the atoms $\{X^* : X \text{ is a control state}\}$.

hypotheses about the world. In principle, it doesn't matter what these new sentences look like; they could be atomic tags: 'X', 'Y', 'Z', etc. (although from a computational perspective that would be rather inefficient).

While the new propositions and the old propositions are logically independent, they should not be probabilitistically independent. If an agent has no idea which ordinary propositions are likely to be true if she chooses X^* rather than Y^*, she will have no basis for choosing one over the other. (I assume the agent does not assign basic value to the option propositions themselves.) Instead, an agent might assign 90% probability to *moving left* conditional on X^*, and 10% to *staying in place*.

These conditional probabilities can be learned and adjusted, by observing the effect of particular choices. Recall that when an agent makes a choice, she becomes certain of the chosen option proposition X^*. Soon afterwards, she may observe that she is moving to the left. That's how she can learn the association between X^* and that type of movement. In the same way, she can learn under what conditions the association breaks down.[25]

You might object that propositions are not the kinds of thing we can simply invent according to our needs. Fair enough. I can give you the word 'proposition'. In section 2, I reviewed some arguments for construing options as propositions. What the arguments really show is that options are best construed as *objects of probability*: elements in the domain of the agent's subjective probability function. My primitive option propositions are "propositions" in this sense, even if they aren't "propositions" in some other sense.

I do assume that we can simply use an extended probability function in our model of rational agents, provided that we explain how the extended function behaves: how it relates to the agent's behaviour, how it is affected by new information, and so on. I have outlined such an explanation, although I have not spelled out all the details.

We may want to reserve the word 'credence' for the old, unextended function. Credence is partial belief, and the point of belief, one might argue, is to represent the world as being one way or another. Any object of credence should therefore be a genuine way the world might be (or might not be). From this perspective, an agent's credence function is only a part of her extended probability function: the part defined over genuine ways the world might be (or not be).

In the agent's credence function, her options will then correspond to Jeffrey-style "gambles": probabilities defined over relevant act-describing propositions. For example, if a given control state X is associated with an element X^* in the agent's extended probability function, and conditional on X^* the agent is 80% confident that she will move to the left and 20% confident that she will stay in place, then this option might

[25] A full description of this learning process would probably have to involve temporally indexed option propositions: the fact that X^*-*now* goes along with *moving to the left now* is evidence that X^*-*later* might go along with *moving to the left later*.

correspond to the gamble of moving to the left with probability 0.8 and staying in place with probability 0.2.[26]

We need the extended probability function to explain how the available gambles are determined – why the agent has an option of moving left with probability 0.8 and staying in place with probability 0.2 – and how the probabilities involved in the gamble are sensitive to the agent's information.

8 Options and actions revisited

I will now explain how the model I have presented fits the picture I outlined in sections 2–4. In section 2, I tentatively suggested that an agent *can* perform an act, in the sense relevant to decision theory, iff there is some available decision that would bring about the act, where a "decision" is an internal control state that initiates action or inaction, and where a decision is "available" if (roughly) it would be realised by some variation of the agent's beliefs or desires. The problem of options then emerged as the problem of how the acts an agent can perform should be represented for the purpose of computing expected utilities.

In section 3, I described three intuitive constraints on option propositions: Ability, Cover, and Maximality. Let's revisit them in reverse order.

The Maximality condition says that if A is an option then there is no other option that entails A. This is easily ensured in the model I have presented, by stipulating that the new option propositions are logically independent of one another (in the agent's probability space).

The Cover condition demands that for any proposition B which the agent can make true there is some option A such that if the agent were to choose A then B would be the case. This gave rise to the "cover problem": we needed thousands of distinct options for Jeffrey's marksman. More generally, when we were looking for option propositions among genuine ways the world could be, we were stuck between the competing demands of (a) making the propositions sufficiently weak so that they don't entail external facts of which the agent is unsure, and (b) making them sufficiently strong so that they exclude one another. With primitive option propositions both demands are easy to satisfy. Option propositions are disjoint without settling any external facts about the world.

To see how the Cover condition is satisfied, suppose B is some proposition the agent can make true, in the sense that there is some available decision X that would lead to the truth of B. On the model I have presented, X is represented in the agent's probability space by an option proposition X^*. To choose X^* is to bring about X. By assumption,

[26] I'm assuming that *moving left* and *staying in place* "screen off" X^*, so that the effect of conditionalizing on X^* is identical to the effect of Jeffrey conditionalizing on the two propositions. If not, *moving left* and *staying in place* is not the right partition: X^* will correspond to a different gamble.

X would lead to the truth of B. So for any proposition B the agent can make true, the option proposition corresponding to the decision that would bring about B is such that if the agent were to choose that option then B would be the case.

The Ability condition says that a proposition is an option only if the agent can make it true. This gave rise to the "ability problem": the range of options now seems to depend on external facts about what the agent can do.

Is is not clear in what sense a primitive option proposition 'can be made true'. In the model I have presented, to choose an option is to bring about the corresponding control state. So we might understand the Ability condition as saying that an agent's doxastic space must only contain option propositions that correspond to control states the agent could realize.

To some extent, this can be satisfied by stipulation. We can take it to be part of the model that every control state must be represented by a unique option proposition, and every option proposition by a unique control state. An agent who deliberates over propositions that don't correspond to any control state doesn't conform to the model.

But the issue is more complex, because control states can become temporarily or contingently unavailable. An agent may be incapable of choosing an option not because the relevant control state doesn't exist, but because something prevents them from realizing it. Remember the case of the demon. Suppose you face a choice between turning left and turning right; a demon monitors your brain, ready to strike you down as soon as he detects that you're about to decide to turn left. Intuitively, the only thing you *can* decide to do is turn right; control states for turning left are unavailable. However, the presence or absence of the demon presumably doesn't affect your probability space. So your probability space will contain an option proposition for turning left, even though you can't choose to make it true. We still seem to have a violation of the Ability condition.

But what should we say about that kind of case? Suppose you have good reason for turning left, and you are rational. We should predict that you will be struck down by the demon. We shouldn't predict that you will (decide to) turn right, even though this is the only objectively available choice. So it seems that an adequate representation of your decision problem really should include the option of (deciding to) turn left – in contradiction to the Ability condition. But then we would also have to give up the idea that rational agents always choose options that maximize expected utility. After all, you don't choose to turn left. (You can't.) For an even clearer illustration, suppose the demon causes a decision to turn left if he notices that you are forming a decision to turn right, and vice versa. Since you have good reason to turn left, we should expect you to (decide to) turn right; you will choose the option with lower expected utility.

Instead of dropping core parts of classical decision theory, I'm inclined simply to set these trouble cases aside and restrict decision theory to situations in which the agent has full control over her decisions. That is, I'm inclined to say that in a genuine decision problem, the agent must be capable of bringing about every control state corresponding to her option propositions.[27] The Ability condition thereby comes out as satisfied – though admittedly only by setting aside possible counterexamples.[28]

I discussed two further conditions in section 4: Modal Certainty and Deliberational Certainty. I will concentrate on the latter, because it is simpler, better motivated, and arguably stronger.

Classical models of decision-making imply that choosing an option goes along with becoming certain of the option. Accordingly, Deliberational Certainty says that A is an option only if the agent could rationally become certain of A through deliberation. This gave rise to the "evidence problem": on what basis should the agent be able to conclusively rule out situations where A is false?

The model I have proposed vindicates the idea that decision-making does not provide infallible, conclusive evidence about contingent reality. The classical models of decision-making apply to an agent's extended probability function, but the "propositions" of which the agent becomes certain are not genuine ways the world might be.

9 Ramifications

The model I have presented may look strange and unfamiliar. But how revisionary is it really? Do we have to start writing meaningless symbols like 'X' and 'Y' in the rows of decision tables? Can we still use decision theory to give practical advice? Can we still regard it as a formal development of folk psychology (as [Lewis 1974: 338] suggested)? Can it still serve as the foundation of microeconomics? I will argue that there is no need to worry.

When it comes to explicit uses of decision-theoretic reasoning, it should be clear that we almost always work with a simplified representation of the true choice situation. When we draw a decision matrix for an agent at a juncture, we pretend that the options are *turning left* and *turning right*, when in reality there are countless other (and more specific) things the agent could do – for example, sitting down and trying to prove the Riemann hypothesis. We also ignore far-fetched states. We ignore brain-in-a-vat scenarios, and scenarios in which the agent is capable of proving the Riemann hypothesis.

[27] More worryingly, we might also have to set aside cases in which the agent merely suspects that she is incapable of choosing a certain option. For example, suppose you give some credence to the possibility that a demon will strike you down if he detects that you will decide to turn left. Intuitively, you now have a reason against deciding to turn left. But how should we model this? The problem is that deciding to turn left is incompatible with the presence of the demon (because the demon would prevent that decision): $Cr(\mathit{Strike}/\mathit{Decide\text{-}Left}) = Cr(\mathit{Strike}\backslash \mathit{Decide\text{-}Left}) = 0$. So how can the possible presence of the demon provide a reason against deciding to turn left?

[28] [Hedden 2012: 354] reaches a similar verdict.

These simplifications are usually harmless, since the missing details would rarely affect the resulting verdict.[29]

On the model I have outlined, an agent's true options are primitive option propositions. However, it usually does little harm to use ordinary act descriptions. To illustrate, return once more to the berry scenario from section 2. I assumed that your (objective) choice is between eating a blueberry and not eating a blueberry, although in any real situation you would have to choose between countless ways of eating a blueberry and countless ways of not eating a blueberry. I also assumed that you're not sure if the berry is a blueberry. As we saw, it would therefore be wrong to count *eating a blueberry* as one of your options. However, *eating a berry from the shrub* will usually do the job. That's because (a) the utility you expect to get from eating a berry doesn't depend much on how exactly you pick the berry (within reason), so we can ignore the difference between these more fine-grained options. Moreover, (b) although the hypothesis that you are about to eat a berry rules out various possibilities of which you may be unsure – such as the possibility that you are a brain in a vat or that you just got paralyzed –, including these states in the decision matrix, and adjusting the act descriptions so as to allow for these states, plausibly wouldn't affect the final recommendation.

In other cases, where the possibility of failure is more salient, or where it matters not just what overt act the agent performs, but also what decision or intention they form, we can retreat from act-describing propositions like *eating a berry* to intention- or decision-describing propositions like *deciding to eat a berry.* In some cases we also have to be more specific about how you pick the berry. But rarely will there be a need to be fully specific, or to consider the outcome in every possible scenario, including scenarios in which you are not an intentional agent capable of making decisions.

The upshot is that my proposal does not call for a revision to practical applications of decision-theoretic reasoning (either formal or informal), as these almost always rely on simplified, coarse-grained representations of decision problems to make the applications tractable.

What about the idea that decision theory is "the very core of our common-sense theory of persons, dissected out and elegantly systematized" [Lewis 1974: 338]? If you want your decision theory to systematize folk psychology, does that give you a reason to reject my proposal? I don't think so.

Let's go through some respects in which my model differs from traditional models. One difference is that my model appeals to "control states" mediating between an agent's attitudes and her actions. These states are well-known to common sense. Folk psychology does not assume that (standing) beliefs and desires directly move our limbs. When we move our limbs so as to further our goals in light of our beliefs, the goals and beliefs feed

[29] We also often make simplified assumptions about what counts as a relevant outcome; these assumptions are less innocent, but that's a different topic (see e.g. [Joyce 1999: 52–56], [Greaves 2016]).

into a decision, which in turn causes the movement. These decisions are (roughly) my control states.

I also assume that decision-makers don't have infallible access to the (physical or functional) nature of their control states. More generally, I assume there is often no suitably detailed description of the available choices for which we can be absolutely sure that the so-described acts are things we can do. And if we're not sure we can do something, then the possibility of failure must be taken into account. So when we evaluate a choice under an ordinary description, we should really evaluate a "gamble" that may or may not lead to the described choice.

These are philosophical points, no doubt, but I don't think they clash with common sense. The remaining details of my model, involving an extended probability space with extra option propositions, are of course alien to the folk. But that's true for every formal model of decision-making. The folk do not talk about sigma algebras.

Finally, what about the role of decision theory in the foundations of microeconomics? Here things are a little more complicated. There is good news and bad news.

Economic models of decision-making generally assume a close connection between choice behaviour, preference relations, and utility functions. On the orthodox "revealed preference" approach, an agent's utility function is seen as an abstract representation of the agent's choice dispositions. Decision theory then boils down to certain formal constraints ("axioms") on choice dispositions (see e.g. [Gul and Pesendorfer 2008], [Chambers and Echenique 2016]).

In this context, the objects of choice can't be construed as things like *turning left* or *saying 'yes'*, since most people don't have well-defined, stable choice dispositions involving so-described options. Are you disposed to say 'yes' when given a choice between saying 'yes' and saying 'no'? It depends. You may be disposed to say 'yes' when asked if you'd like a piece of cake, but 'no' when asked if you'd like a saucer of mud. Choice functions in economics are therefore not defined over ordinary act descriptions, but over something like Savage's functions from "states" to "outcomes".[30]

Can we replace these functions by my primitive option propositions? No. Choice functions and preference relations cannot be defined over my option propositions, for the same reason for which they can't be defined over act descriptions like *turning left* or *saying 'yes'* (or *deciding to say 'yes'*): whether you are disposed to choose the option corresponding to the control state that normally triggers an utterance of 'yes' varies

[30] In some parts of economics, it is assumed that the agent has no uncertainty or ignorance about the outcomes (here, typically, commodity bundles) of the choices they face; the objects of choice can be identified directly with these outcomes. In other parts of economics, it is assumed that the agent knows the objective probability of getting a particular outcome as the result of a particular choice; the objects of choice my then be modelled as probability functions over outcomes, following [von Neumann and Morgenstern 1944]. I concentrate on the hardest case, because this is where the full machinery of Bayesian decision theory enters the picture.

from choice situation to choice situation, even if your underlying desires remain the same. That's the bad news.

The good news is that my proposal may help to fill a gap in the revealed preference account. If we want to determine an agent's utility function from her choice dispositions, we are not *given* a description of the choice dispositions in terms of functions from states to outcomes. An agent's choice dispositions are, in the first place, dispositions to choose between concrete acts. To get the revealed preference account off the ground, these acts must be converted into functions from states to outcomes. And that conversion is far from trivial. (This is the problem of options as it presents itself in Savage's framework.)

Consider the berry scenario. Here the function associated with your choice of eating the berry presumably maps ordinary *blueberry* states to *eating a blueberry* outcomes, *tutsan berry* states to *eating a tutsan berry* outcomes, and *brain in a vat* states to *still a brain in a vat* outcomes. Why is the chosen act associated with this particular function? What's the general principle?

This is where my model may help. When you decide to eat the berry, you choose a primitive option proposition X^*. But you have some idea of how this proposition relates to other propositions. Under some idealisation, we may assume that you are certain that choosing X^* would[31] lead to a blueberry outcome *if* a normal blueberry state obtains, to a tutsan berry outcome *if* a tutsan berry state obtains, and to a brain in a vat outcome *if* a brain in a vat state obtains. That's how we get the above mapping from states to outcomes.

In general, if there is a partition of states for which the agent is certain which outcome any available option proposition would bring about in any given state, we can use this information to construct the available mappings from states to outcomes (for a particular choice situation).

This is evidently not a full answer. How do we find the relevant states and outcomes? What if the agent isn't sure which outcome would result from a given option in a given state? We might make further progress by adapting an idea from [Lewis 1981: 11]. Lewis suggests that a state (a "dependency hypothesis", in his terminology) is "a maximally specific proposition about how the things [the agent] cares about do and do not depend causally on his present actions". Let's replace the "present actions" with my option propositions. In easy (deterministic) cases, a Lewisian state then effectively specifies for each option proposition which outcome would result from its choice. This not only gives us the relevant states, it also ensures that the agent can be certain which outcome would result from any given option in any given state.

More needs to be said to turn these impressionistic remarks into a fleshed-out theory. But I hope I have said enough to convey that although my account of options does not

[31] Friends of evidential decision theory might want to say 'will' instead of 'would'.

fit the standard take on options in economics, it may nonetheless prove useful to fill an important (but largely ignored) gap in the foundations of microeconomics.

To conclude, let me briefly return to Jeffrey's insight that our choices are part of the natural world and should therefore be represented as ordinary propositions. The model I have proposed suggests that this is not quite right. Option propositions are not ordinary ways things might be. In the decision-maker's mind, they are only contingently and probabilistically related to propositions about the natural world. This might explain or even vindicate the popular intuition that rational decision-makers must treat their choices as 'interventions', not governed by ordinary physical laws, and not constrained by degrees of belief pertaining to natural propositions. Whether my model really vindicates these intuitions is another question I have to leave for another occasion.

References

G.E.M. Anscombe [1957]: *Intention.* Ithaca, NY: Harvard University Press

Campbell Brown [2018]: "Maximalism and the Structure of Acts". *Noûs.* Forthcoming

Christopher P. Chambers and Federico Echenique [2016]: *Revealed Preference Theory.* Cambridge: Cambridge University Press

David Christensen [1992]: "Confirmational Holism and Bayesian Epistemology". *Philosophy of Science*, 59(4): 540–557

Hartry Field [1978]: "A Note on Jeffrey Conditionalization". *Philosophy of Science*, 45(3): 361–367

Harry Frankfurt [1969]: "Alternate possibilities and moral responsibility". *The Journal of Philosophy*, 66(23): 829–839

Andre Norman Gallois [2009]: "The Fixity of Reasons". *Philosophical Studies*, 146(2): 233–248

Daniel Garber [1980]: "Field and Jeffrey Conditionalization". *Philosophy of Science*, 47(1): 142–145

Allan Gibbard and William Harper [1978]: "Counterfactuals and Two Kinds of Expected Utility". In C.A. Hooker, J.J. Leach and E.F. McClennen (Eds.) *Foundations and Applications of Decision Theory,* Dordrecht: D. Reidel, 125–162

Hilary Greaves [2016]: "Cluelessness". *Proceedings of the Aristotelian Society*, 116(3): 311–339

Daniel Greco [2017]: "Cognitive Mobile Homes". *Mind*, 126(501): 93–121

Faruk Gul and Wolfgang Pesendorfer [2008]: "The Case for Mindless Economics". In Andrew Caplin and Andrew Schotter (Eds.) *The Foundations of Positive and Normative economics: A Handbook,* vol 1. Oxford: Oxford University Press, 3–42

Alan Hájek [2012]: "Is Strict Coherence Coherent?" *dialectica*, 66(3): 411–424

Brian Hedden [2012]: “Options and the subjective ought”. *Philosophical Studies*, 158(2): 343–360

Frank Jackson [1991]: “Decision-Theoretic Consequentialism and the Nearest and Dearest Objection”. *Ethics*, 101(3): 461–482

Richard Jeffrey [1965]: *The Logic of Decision*. New York: McGraw-Hill

— [1968]: “Probable knowledge”. *Studies in Logic and the Foundations of Mathematics*, 51: 166–190. Reprinted with minor revisions in [Jeffrey 1992]

— [1992]: *Probability and the Art of Judgment*. Cambridge: Cambridge University Press

James Joyce [1999]: *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press

— [2012]: “Regret and instability in causal decision theory”. *Synthese*, 187(1): 123–145

Angelika Kratzer [1981]: “The notional category of modality”. *Words, Worlds, and Contexts*: 38–74

David Lewis [1974]: “Radical Interpretation”. *Synthese*, 23: 331–344

— [1979]: “Attitudes *De Dicto* and *De Se*”. *The Philosophical Review*, 88: 513–543

— [1981]: “Causal Decision Theory”. *Australasian Journal of Philosophy*, 59: 5–30

Patrick Maher [1993]: *Betting on theories*. Cambridge University Press

John Maier [2018]: “Abilities”. In Edward N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy,* Metaphysics Research Lab, Stanford University, spring 2018 edition

John Perry [1979]: “The problem of the essential indexical”. *Noûs*, 13: 3–21

Philip Pettit [1991]: “Decision Theory and Folk Psychology”. In M. Bacharach and S. Hurely (Eds.) *Foundations of Decision Theory: Issues and Advances,* Cambridge (MA): Blackwell, 147–175

John L. Pollock [2002]: “Rational Choice and Action Omnipotence”. *The Philosophical Review*, 111: 1–23

Stuart J. Russell and Peter Norvig [2010]: *Artificial Intelligence: A Modern Approach*. Cambridge (MA): MIT Press, 3rd edition

Leonard Savage [1954]: *The Foundations of Statistics*. New York. Wiley

Wolfgang Schwarz [2018]: “Imaginary Foundations”. *Ergo*, 29: 764–789

Brian Skyrms [1980]: “Higher Order Degrees of Belief”. In D.H. Mellor (Ed.) *Prospects for Pragmatism,* Cambridge: Cambridge University Press

— [1984]: *Pragmatics and Empiricism*. Yale: Yale University Press

— [1990]: *The Dynamics of Rational Deliberation*. Cambridge (Mass.): Harvard University Press

Jordan Howard Sobel [1971]: “Value, Alternatives, and Utilitarianism”. *Noûs*, 5(4): 373–384

— [1980]: *Probability, Chance and Choice*. Unpublished book manuscript

— [1983]: “Expected utilities and rational actions and choices”. *Theoria*, 49: 159–183. Reprinted with revisions in [Sobel 1994: 197–217]

— [1986]: “Notes on decision theory: Old wine in new bottles”. *Australasian Journal of Philosophy*, 64: 407–437. Reprinted with revisions in [Sobel 1994: 141–173]

— [1988]: “Infallible Predictors”. *The Philosophical Review*, 97(1): 3–24

— [1994]: *Taking Chances*. Cambridge: Cambridge University Press

Jean-Paul Vessel [2003]: “Counterfactuals for consequentialists”. *Philosophical Studies*, 112(2): 103–125

John von Neumann and Oskar Morgenstern [1944]: *Theory of games and economic behavior*. Priceton: Princeton University Press

Paul Weirich [1983]: “A decision maker’s options”. *Philosophical Studies*, 44(2): 175–186

Jonathan Weisberg [2009]: “Commutativity or holism? A dilemma for conditionalizers”. *The British Journal for the Philosophy of Science*, 60(4): 793–812