Options and Actions

Wolfgang Schwarz*

Draft, 7 August 2017

Abstract. Bayesian decision theory says that when faced with a range of alternative actions, rational agents generally choose an option that maximizes expected utility. But what are the available options? Whatever acts the agent could in fact perform? Or whatever acts she believes she could perform? Neither answer is satisfactory. After reviewing these and some other answers, I suggest that decision-theoretic models of rational agents should include a special domain of 'option propositions' to serve as objects of practical deliberation.

1 Introduction

According to Bayesian decision theory, rational agents choose actions that maximize expected utility relative to their subjective probabilities and utilities. But even if we know a rational agent's probabilities and utilities in a given situation, this does not allow us to predict what the agent will do, unless we also know what acts are available: what are the options from which she can choose?

The question is especially pressing if we understand decision theory—as I will throughout the present paper—as a psychological model: as 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 given as part of a well-defined decision problem.[1]

So what are the options available to an agent in a given decision situation? The naive answer is that the options are simply whatever acts the agent can perform. In sections 2 and 3, I will argue that this is at most correct on a special, and somewhat elusive, reading of 'can' (and of 'acts'). In particular, decision theory seems to require that the agent has perfect control over the relevant acts, so that she could be rationally certain

* Thanks to Romy Jaster and Theodore Korzukhin for helpful discussion.

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

that she will perform an act merely on the basis of deciding to perform it. Ordinary acts rarely meet this condition: when we decide to do something, there is usually some chance that the world will interfere and we will not end up doing what we decided to do.

In response, several authors have suggested that decision-theoretic options should be understood not as overt acts, but as mental events of intending, trying, or deciding to perform an act. Another response (due to Richard Jeffrey) is to model an option as a range of possible acts—all the acts that might result from the relevant decision—associated with different probabilities. I will review these proposals in sections 4 and 5 and argue that they do not provide a fully satisfactory answer.

With all hopes for an easy solution dashed, I will then turn to my positive proposal in sections 6 and 7. I will suggest that decision-theoretic models of rational agents should postulate a special domain of 'option propositions' to serve as objects of practical deliberation. The option propositions correspond to the 'basic acts' an agent can perform—essentially, the decisions she can make—, but they do not transparently describe or represent these acts.[2]

2 What we can do

Let's begin by having a closer look at the acts an agent can perform in a given decision situation.

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 little you can do about that. Not so when a hike leads you to a junction. Here, whether you end up turning right or left is sensitive to your psychological state: to your beliefs, desires, intentions, fears, hopes, and whims. In that sense, what you do is under your psychological control. You face a choice.

These are the kinds of choices studied in decision theory. They do not require a strong, libertarian kind of freedom. Decision theoretic algorithms are 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. If you build a robot, you don't need to specify that it should drop towards the Earth when it falls out of a helicopter; it will do that no matter what internal states and decision rules you build in. Not so when the robot reaches a junction. Here the outcome is under control of the robot's internal state: vary the state, and the robot will choose different paths. That is all the freedom we need.

As a first pass, we might therefore define the acts available to an agent (in the sense relevant to decision theory) as follows.

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

> Availability. An act is *available* to an agent iff there is a possible variation of the agent's motivational state that would cause her to perform the act.[3]

In section 7 I will have more to say on what should count as a 'possible variation' of a 'motivational state'. For now, we may vaguely understand a 'motivational state' as some kind of state that normally initiates action—something like a willing, a decision, an intention. 'Act' should be understood liberally, to include things like directing one's attention to one's breath, planning to go to the gym, or not pressing a button.

The available acts are the acts an agent *can* perform, on one salient reading of 'can'. There are many other readings though, and it is important to keep them apart. For example, there is a sense in which I can play the piano even if there is no piano around. That sense is not captured by the present definition, and it is irrelevant to decision theory: if there is no piano around, then playing the piano is not one of my options.

More importantly, when we talk about what an agent can or cannot do, we often restrict the relevant motivational states. When offered a meat pie, for example, I might say that I *can't* eat the pie on the grounds that I am vegetarian, or simply because I do not want to spoil my appetite for dinner. These considerations do not preclude eating the pie as an available act. Indeed, the reason why I don't eat the pie is that I don't *want* to eat it, not that I don't have a genuine choice to eat it in the first place. (By comparison, the reason why I don't turn the pie into gold is that this is not an available act; whether I would like to turn the pie into gold is therefore irrelevant.)

We sometimes also hold fixed other facts that restrict the available acts. For example, if a predictor has predicted what an agent will do and we are certain that her prediction is correct, we might say that the agent *can't* falsify the prediction (see [Horgan 1985: 231]). Similarly, we might say that a time traveller can't kill her own grandfather ([Lewis 1976: 75–80]), or that agents in a deterministic world can only do what they actually do ([van Inwagen 1983]). In each case, it does not follow that the agent did not have a genuine choice to perform the relevant act.

In other respects, what an agent can intuitively do (in a given situation) may outstrip the range of available acts. Consider Professor Procrastinate (from [Jackson and Pargetter 1986]), who is asked to write a review, but won't actually write it if he agrees. Assume the professor has no means of binding himself to write the review: there is nothing he

3 The counterfactual ('... would ...') must be read in a 'non-backtracking' way (compare [Lewis 1979]). That is, to evaluate what would be the case if the agent were in a certain motivational state, we should not consider how that state might have been caused and what else might have resulted from these causes. (To illustrate, imagine an agent who does not raise her left hand and narrowly escapes a stroke that would (a) make her want to raise her hand, and (b) paralyse her arm. Here, raising her hand is an available act even though there is a salient backtracking reading on which, if the agent had wanted to raise her hand, the intention would have been caused by the stroke and thus would not have caused her to actually raise her hand.)

can do at the point when he is asked that would ensure that his future self will write the review. *Agreeing and then writing the review* should then not count as an available act. In general, decision theory (as a psychological model) should not evaluate sequences of acts unless the agent has some way of binding herself to the sequence (see [Joyce 1999: 57–61]). Nonetheless, there is a sense in which the professor *can* agree and then write the review. If he were properly motivated now and in the future, he would succeed.

These ways in which judgements about what an agent 'can' do come apart from what acts are available to them also show up in normative judgements about people's actions. When we give advise or attribute blame and praise we often hold fixed deeply ingrained fears or hopes or convictions. We may even hold fixed simple intentions: since you intend to change lanes, you ought to accelerate (see [Goldman 1978]). Some intuit that determinism renders all praise and blame pointless. Again, in the other direction, there is a ('possibilist') sense of 'ought' in which Professor Procrastinate ought to agree and then write the review (see e.g. [Timmerman and Cohen 2016]). So we should not assume any simple and direct connection between our present topic—a decision-maker's options—and judgements about what an agent ought to do.

So far, I have assumed that an agent's decision-theoretic options are acts. But arguably one and the same act may be described or picked out in several ways, and the choice of presentation can make a difference to the act's expected utility. For example, suppose Bob is cooking risotto for a date with Alice, and Alice's least favourite dish happens to be risotto. Then Bob's cooking can be described both as *cooking risotto* and as *cooking Alice's least favourite dish*. But since Bob is not aware that Alice's least favourite is risotto, the expected utility he assigns to *cooking risotto* is a lot greater than the expected utility of *cooking Alice's least favourite dish*. For the purposes of decision theory, we should therefore not equate these options.

So we need a relatively fine-grained conception of acts. In decision theory, the two main approaches are due, respectively, to Leonard Savage [1954] and Richard Jeffrey [1965]. Savage construes acts as abstract functions from 'states of nature' to 'outcomes'. I prefer Jeffrey's proposal on which acts are simply modelled as propositions. In some sense, Bob's act of cooking risotto may be identical to his act of cooking Alice's least favourite dish, but the proposition *that he cooks risotto* is undoubtedly different from the proposition *that he cooks Alice's least favourite dish*.

Let me briefly mention some reasons in favour of Jeffrey's proposal. First, it meshes nicely with discussions in metaphysics, ethics, and action theory, where it has also proved useful to model acts as propositions (see e.g. [Bennett 1988] and [Bennett 1995]). Second, Jeffrey's proposal does not assume the logical independence between acts and 'states' and 'outcomes', which elegantly avoids many problems for Savage's theory (see e.g. [Broome 1991: ch.5]). On Jeffrey's view, there is no sharp distinction between acts on the one hand and states of nature on the other: "the human agent is taken to be part of nature and

his acts are thus ingredients in states of nature" [Jeffrey 1992: 226]. Third, construing options as propositions opens up the possibility of considering the agent's beliefs about what her options may cause or indicate, 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 beliefs, it is hard to give a satisfactory diagnosis of Newcomb Problems (see [Joyce 1999: 117ff.]) or to explain the way rational beliefs change through deliberation (see [Skyrms 1990], [Joyce 2012]).[4]

So there are good reasons to model decision-theoretic options as propositions. A proposition is the kind of thing that can serve as object of credence. For reasons that will become clear, I do not want to commit prematurely to any particular view on what these things are. If an agent chooses an option, and the option is modelled as a proposition, I will say that the agent *makes true* the propositions. The precise semantics of 'making true' won't be important; indeed, it will do no harm to read 'S makes true A' as equivalent to A: you make true that you raise your arm iff you raise your arm.

Now let's return Bob. When Bob decided to cook risotto, he decided to make true the proposition that he cooks risotto, and not that he cooks Alice's least favourite dish. Modelling acts as propositions allows us to distinguish these options. But was the second a genuine option? Could Bob have decided to cook Alice's least favourite dish? Arguably not.

If that isn't obvious, consider Rob, a counterpart of Bob with the same imperfect information about Alice who actually *wants* to cook Alice's least favourite dish. If cooking Alice's least favourite food were an option for Rob, then this is what he ought to choose; it would maximize expected utility. But if Rob falsely suspects that Alice's least favourite food is, say, Haggis, then it may well be rational for him to cook Haggis and thus *not* to cook Alice's least favourite dish (which is risotto). So cooking Alice's least favourite dish is not an option for Rob. Nor is it an option for Bob. It is not an option despite the fact that it is what Bob ended up doing!

If that sounds strange, it might help to distinguish two kinds of options. Cooking Alice's least favourite dish is an *objective option* for Bob insofar as it meets the above definition of an available act. Intuitively, an objective option is a proposition an agent can in fact make true by an act of her will. Bob can make true that he cooks Alice's least favourite dish. But he can't choose the act 'under that description'; he can't cook

4 Some (e.g. [Spohn 1977] and [Levi 2007]) have argued that deliberation is incompatible with having credences about one's own options. I am not convinced by these arguments, more or less for the reasons discussed in [Joyce 2002] and [Hájek 2017]. We will nonetheless come to reconsider Jeffrey's insight. On a certain understanding of 'credence'—namely, if 'credences' are restricted to consciously entertainable propositions, or to genuine ways the world could be—the model I will propose suggests that an agent's options are indeed not suitable objects of credence.

Alice's least favourite dish 'at will'. Cooking Alice's least favourite dish is not one of his *subjective options*.

According to decision theory, rational agents maximize expected utility among their subjective options. But what exactly determines whether a proposition is a subjective option?

3 Subjective options

Why does it seem that Bob can't cook Alice's least favourite dish at will? It is not because Bob would fail if, counterfactually, he had wanted to cook Alice's least favourite dish. For suppose the only thing Bob can cook is risotto. He might then well have decided to cook risotto even if he had wanted to cook Alice's least favourite dish, on the off chance that risotto is her least favourite dish. He would have succeeded. But that is still not enough to give Bob the desired power. Intuitively, if an agent can do something at will, then whether or not she does it should depend only on her well and not on further factors over which she has no control; it should be possible for her to become certain that she will perform the act merely by making the relevant choice. Whatever Bob decides to cook, whether or not it is Alice's least favourite dish depends on a further matter (Alice's tastes) of which he is ignorant; consequently, whatever he decides to cook, he can't be sure he will cook Alice's least favourite dish.

Unfortunately, appealing to 'factors' over which the agent has or lacks 'control' threatens to be circular. But the present reflections also suggest the following, non-circular condition, discussed in [Jeffrey 1968]:

> Anscombe's condition. A proposition A is a subjective option for an agent only if the agent could become rationally certain of A just by making a decision.

The condition is named after Anscombe's famous remark that intentional agency provides "knowledge without observation" [Anscombe 1957: sec.8]. When we decide to raise our hands, for example, we are usually not surprised to find our hands go up; making the decision already gave us good reason to think that that would happen.

In decision theory, the manner in which practical deliberation changes rational credence has been studied extensively by Brian Skyrms (see especially [Skyrms 1990]). In line with Anscombe's condition, Skyrms's models imply that if deliberation settles on a given option, then the agent becomes certain that the option will be realized.

So Anscombe's condition is fairly well motivated, and it seems to filter out unwanted options like cooking Alice's least favourite dish. Admittedly, the condition is not as direct and perspicuous as one might like. If a proposition meets Anscombe's condition, then this is clearly not a brute fact. There is a reason why Bob could not rationally

have become certain that he will cook Alice's least favourite dish merely on the basis of making a choice. The reason has something to do with his lack of information about Alice's tastes. There must be some other, plausibly non-normative, way to single out the propositions that satisfy Anscombe's condition, but it is not obvious how.

Is Anscombe's condition sufficient to identify an agent's subjective options? No. To begin, we presumably still want to say that an agent's options are propositions the agent can in fact make true. We don't want decision theory to predict or recommend an option that the agent can't actually realize. To be sure, one might argue that an agent can only become rationally certain of a proposition if the proposition is true, and hence that Anscombe's condition entails the Availability condition. But without further explanation, that looks questionable; we will actually meet a counterexample in a moment.

Second, consider the following scenario. Imagine (for reasons that are none of our business) Alice, who has by now arrived, would like to unintentionally spill wine onto Bob, but she would not like to do so intentionally. Having no means of causing herself to unintentionally spill the wine, she (rationally) refrains from spilling it. In this case, it might well be that the most likely circumstances under which she would spill the wine are circumstances in which she does so unintentionally. As a consequence, spilling the wine has high expected utility.[5] It follows that spilling the wine should not count as an option: if it did, it would be rational of Alice to choose it, but it is not. Yet it may well meet Anscombe's condition: if Alice decided to spill the wine, she might well (rationally, and accurately) become certain that she would succeed.

Here it seems that not everything an agent can do 'at will' is a decision-theoretic option. Spilling the wine is plausibly something Alice can do at will; but it is not an option. So perhaps an agent's options should be restricted to the *logically strongest* propositions the agent can make true at will. *Intentionally spilling the wine* is logically stronger than *spilling the wine*; the present condition would therefore exclude the latter.[6]

Third, consider Newcomb's Problem with a perfect predictor: An agent faces a choice between taking an opaque box ('one-boxing') and taking the opaque box together with another box that is known to contain \$1000 ('two-boxing'); a mysterious predictor has put a million dollars into the opaque box iff the agent was predicted to one-box, otherwise

5 This is true both in Evidential and in Causal Decision Theory, for $Cr(\textit{spilling unintentionally} \,/\, \textit{spilling})$ and $Cr(\textit{spilling unintentionally} \setminus \textit{spilling})$ are both high; see [Joyce 1999] for the definition of $Cr(\cdot \setminus \cdot)$.

6 Lewis [1981: 308] defines options as the strongest propositions the agent can make true at will, but he does not explain or motivate his definition. Extant discussions of whether options should be maximally strong (as in [Weirich 1983] and [Sobel 1983]) mostly focus on cases where a logically weaker option is an (exclusive) disjunction of more specific options. This makes the issue rather subtle, as it may seem that a disjunctive option has maximal expected utility only if all its disjuncts do, in which case including the unspecific option would at least do no harm. This is certainly true in evidential decision theory. On the other hand, Sobel [1983: 203–208] presents a complicated scenario in which a disjunctive option seems to have greater causal expected utility than all its disjuncts.

the opaque box is empty. The decision-maker is rationally certain that the predictor has made the right prediction. Under these conditions, one-boxing could make the agent rationally certain that she will *one-box and get a million*, and two-boxing that she will *two-box and get (only) a thousand*. Note however that only one of these is something the agent can in fact make true. If, for example, the predictor has left the opaque box empty, then nothing the agent can do would result in the money suddenly appearing in the box; so if the opaque box is empty then *one-boxing and getting a million* is not an available act. (Here is the counterexample to the claim that Anscombe's condition entails the Availability condition.) Even if the opaque box does contain the million, as long as the agent is not sure that it does, *one-boxing and getting a million* should arguably not count as something the agent can make true at will, nor should it count as one of her decision-theoretic options.[7]

To rule out options like *one-boxing and getting a million*, we might impose the further condition that each option should be logically compatible with every 'state', where a 'state' is something like a "maximally specific proposition about how the things [the agent] cares about do and do not depend causally on his present action" [Lewis 1981: 313]. The state in which the opaque box is empty and thus one-boxing would (causally) yield \$0 and two-boxing \$1000 is logically incompatible with *one-boxing and getting a million*.[8] Unfortunately, like appealing to what is or isn't under an agent's 'control', the quoted definition of states refers back to the agent's options (as 'present actions'), so it might be circular to use it in the definition of options.

Various other conditions have been proposed to distinguish subjective options. J.H. Sobel has suggested that a proposition is an option for an agent only if the agent is *certain that she can make it true* (see [Sobel 1980: 178–80], [Sobel 1983: 199f.], [Sobel 1986: 155f.], [Sobel 1988: 8]). Sobel's condition is motivated by the idea that decision theory is supposed to be a theory of subjective rationality—of what an agent should do *from her own perspective*, relative to her own beliefs and desires. This suggests that the available options should not turn on external facts inaccessible to the agent.

However, the accessibility requirement seems to clash with the requirement that an agent's options should be genuinely available. What if an agent is certain that she can

7 Here one's intuitions might turn on which choice one thinks is rational. [Seidenfeld 1984], who endorses one-boxing, suggests that the agent really does face a choice between a million and a thousand; see [Sobel 1988] for discussion.

One might also think that Newcomb's Problem with a perfect predictor is not a genuine decision problem, either because the scenario is incoherent or because the agent could not 'regard herself as free'. But the present problem also arises in other scenarios, e.g. those discussed in [Ahmed 2013] and [Schwarz 2015], where this response looks less appealing.

8 The compatibility has to be *logical* in some suitable sense, rather than doxastic or epistemic: in the perfect Newcomb Problem, *one-boxing* is epistemically incompatible with the described state, but it should count as an option.

fly while in fact she can't? We don't want decision theory to predict or recommend that the agent will fly. So even with Sobel's condition in place, an agent's options still seem to depend on external facts about what the agent can in fact do.[9]

Moreover, Sobel's condition arguably goes beyond what the demand for 'accessibility' requires. Compare an agent's beliefs and desires. These are surely accessible in the relevant sense; nonetheless, a decision-maker need not be certain of her own beliefs and desires.[10]

Note also that Sobel's condition cannot do the work of Anscombe's. If Bob is a versatile cook, he might be certain that he is able cook Alice's least favourite dish; but if he doesn't know what that dish is, cooking Alice's least favourite dish should (still) not count as an option.[11] Conversely, Anscombe's condition can do most of the work of Sobel's. Consider a typical scenario where an agent has doubts about what she can do. Brian Hedden [2012: 347] presents the following case in support of Sobel's condition.

> Jane is hiking along the trail when she comes to a raging creek. She is in fact able to ford the creek (where this entails getting across), head upstream to look for an easier crossing, or turn back. Among these three things that she is able to do, fording the creek has highest expected utility [...] But she has

9 Sobel ([1980], [1986]) stipulates that a rational decision-maker is never mistaken when she is certain about what she can do, but why should we accept that? Hedden [2012] offers a possible explanation. He argues that (1) agents who are certain of their own beliefs can never be wrong about *what intentions they can have*; and that (2) an agent's options are always intentions. Let's grant (2); we will return to it in section 4. Hedden's argument for (1) consists in a tendentious analysis of the conditions under which an agent *can have an intention*: according to Hedden, an agent can intend to X just in case she is not certain that if she intended to X she would fail to X. But that threatens to change the topic. Intuitively, it seems perfectly possible for an agent to be unsure whether a certain intention—say, an intention to cry, or to multiply 941,083,981 by 67,280,421,310,721—would result in success while as a matter of fact it is biologically impossible for them to have the intention. And if an act is biologically impossible, then arguably decision theory should not predict or recommend it. Independently of this problem, Hedden's explanation obviously falls silent for agents who are not certain of their own beliefs; and certainty about one's beliefs is hardly a norm of rationality (see [Skyrms 1980]).

10 Some authors portray decision theory as the absurd doctrine that rational decision making is a matter of consciously computing the expected utilities of one's options. This might require that the agent knows her options, as well as her probabilities and utilities. But the proposed interpretation of decision theory makes the theory utterly implausible as a descriptive, normative, or constitutive model. Real people don't consciously compute expected utilities whenever they face a choice, nor is there any reason to think that they should, or that ideally rational people would. (Here I agree with [Pettit 1991: 167–169], [Jackson 1991: 468–471], [Maher 1993: 5–8], [Joyce 1999: 80], and many others.)

11 A different type of case: In the perfect Newcomb Problem, if the opaque box contains a million, *one-boxing and getting a million* and *two-boxing and getting a million and a thousand* are both available; they are things the agent can do. Now suppose the agent is certain that she will one-box and hence that the opaque box contains a million. *Two-boxing and getting a million and a thousand* then satisfies Sobel's condition but not Anscombe's. In this case, however, it is not obvious who is right.

> serious doubts about whether she can in fact ford the creek. After all, the water is up to mid-thigh and flowing fast.

Here, fording the creek should not count as a (subjective) option for Jane. Hedden infers that an act should only count as an option if the agent believes that she can perform it. But Anscombe's condition already deals with that kind of case: since Jane is not certain that she has the required strength to ford the creek, it would not be rational for her to become certain that she will ford the creek merely on the basis of making a decision. In general, if an agent is unsure about whether she can make true X because that depends on external factors outside her control about which she is not certain, then Anscombe's condition typically implies that X is not an option: the dynamics of rational deliberation alone could not make the agent certain that the external factors are favourable.

But there are exceptions (hence 'typically'). In the perfect Newcomb Problem *one-boxing and getting a million* passes Anscombe's condition even though whether the agent can perform the act depends on what is in the opaque box. Here Sobel's condition might be added to rule out unwanted options.

Hedden also endorses another condition: that *options should supervene on mental states*. That is, a proposition is an option for an agent only if it is an option for every possible psychological duplicate of the agent. Supervenience is another way of explicating the idea that options should be accessible. But here again there is a tension between accessibility and actual availability. Indeed, supervenience combined with availability suggests that an agent never has any options besides the one she actually chooses. For suppose there is some option A an agent does not in fact make true. Then (for almost any A), one can imagine a psychological duplicate of the agent who cannot in fact make true A, perhaps because someone monitors their psychological state, ready to intervene as soon as the agent sets out to make true A (compare [Frankfurt 1969]). By the availability condition (that A is an option only if the agent can in fact make true A), A is not an option for the duplicate. By the supervenience condition, it follows that A is also not an option for the original agent.

Instead of reviewing further conditions or refinements of the conditions I have discussed, let me take stock of where we stand. We have seen that not every proposition an agent can in fact make true should count as a decision-theoretic option. Intuitively, for a proposition to be a genuine option, the agent must be able to make it true 'at will'; she must have 'full control' over the proposition; but it is not clear how these notions should be spelled out. Anscombe's condition seems to point in the right direction, but it is neither sufficient nor optimally perspicuous. Conditions of accessibility are more perspicuous, but they cannot replace Anscombe's condition and seem to clash with the requirement that an agent's option should be actually available.

To make progress on these matters, let me now turn to another worry about the direction we have taken. All the conditions I have reviewed are rather strong. Right

now, for example, I am not absolutely certain that I am able to raise my arm: I cannot rule out that my arm just got paralysed, or that I am a brain in a vat. Consequently, merely deciding to raise my arm would not make me certain that I *will* raise my arm. By Anscombe's (or Sobel's) condition, raising my arm is therefore not one of my options.

Now, from a strict and philosophical point of view, this is arguably correct. For suppose I attribute extremely low utility to the scenarios in which I can't raise my arm. If raising my arm were an option, it would therefore have greater expected utility than not raising my arm, simply because the proposition that I raise my arm rules out the undesirable states of being paralysed or envatted. But clearly my desire not to be paralysed or envatted gives me no reason to raise my arm.

So raising my arm is not one of my options. But then what is? Is there any suitable proposition of which I could rationally become certain as a result of deciding to raise my arm?

Well, how about the proposition *that I intend to raise my arm*?

4 Decisions

Whether an agent would succeed in performing an overt act, like raising an arm, generally depends on factors outside her control about which she may reasonably be uncertain. This makes overt acts (or act propositions) ill-suited as decision-theoretic options. So perhaps we should retreat to more immediate and secure consequences of decisions: episodes of *trying*, *intending*, *willing*, or *deciding*. Ideas along this line have been defended by several authors, including Sobel [1971], [1983], Weirich [1983], Joyce [1999], and Hedden [2012] (see also [Pollock 2002] for criticism).

There are independent reasons to allow for intentions or decisions as options. Suppose you are in Damascus at noon and consider going to Aleppo in the evening. If *going to Aleppo in the evening* is one of your options, the expected utility of that option turns on what is likely to happen if you go (or were to go) to Aleppo. But suppose you're inclined to stay in Damascus, so that the most likely scenario in which you go to Aleppo is one in which you get drunk in the afternoon and then decide to go on a whim, without packing enough food and clothes and without informing your family. This scenario, we may assume, has very low subjective utility. But clearly that does not give you a reason against deciding (now) to go to Aleppo in the evening. If you now decided to go, you would make the necessary arrangements. Thus what matters are not the probable scenarios in which you *go to Aleppo in the evening*, but the probable scenarios in which you *now decide to go to Aleppo in the evening*. For the purposes of calculating expected utilities, this is the relevant option to consider.

Indeed, sometimes the main reasons for deciding to perform an act are not features or consequences of the act, but features or consequences of the decision. An anxious

person may reasonably make a decision just to stop worrying and calm her mind, even if other considerations would support delaying the decision until more information becomes available. In Gregory Kavka's [1983] toxin puzzle, the reason for deciding to drink the toxin is that the resulting state of decision or intention will be rewarded. Similarly in the more realistic scenarios on which the puzzle is modelled: a firm decision to leave your partner if she betrays your trust can have high expected utility even though acting on the decision does not.

What is highlighted by these cases is that agents face decisions not only when they physically reach a junction in a road. We can make decisions not only concerning our immediate behaviour but also concerning our behaviour in the distant future. This is useful for a variety of reasons. Among other things, it can provide valuable information about the future and thus allow for more informed decisions. Whether or not you will go to Aleppo in the evening makes a big difference to how you should spend the afternoon. Conversely, what you do in the afternoon affects the expected utility of then leaving to Aleppo. Deciding ahead of the time to go to Aleppo simplifies the problem and allows for more optimized courses of action, by reducing the uncertainty about your future behaviour.

To allow for planning, it would not be enough to move from overt acts to tryings. The most likely scenarios in which you *try to go to Aleppo in the evening* might still be scenarios in which you do so on a drunken whim. What gets rewarded in the toxin puzzle is not *trying to drink the toxin*. The relevant options here are not tryings, but intentions, understood as a kind of binding or commitment. When an agent decides ahead of the time to pursue some course of action, she adopts a commitment to pursue that course. At least in part, this is a change in her beliefs and desires: she becomes more confident that she will act as she intends, and she attaches subjective cost to scenarios in which she acts otherwise. Perhaps other aspects of her mental state change as well; perhaps her commitment even involves a special propositional attitude (see e.g. [Bratman 1987]).

In any case, there are good reasons to think that in many decision contexts, the relevant options are (propositions about) intentions rather than overt acts. It is less obvious that *all* decisions necessarily involve intentions. When we spontaneously raise a hand to greet a friend, navigate around other people on the footpath, or take a sip of coffee from a cup, no special commitment or binding may be in play. Yet we are still making choices. These choices don't involve conscious deliberation, but they should in principle be captured by decision theory as a psychological model of decision-making.

To capture these kinds of cases, we might focus on a weak reading of 'intention' on which an intention to A is nothing more than a *decision* to A. It is safe to assume that whenever an agent makes a decision (consciously or subconsciously), then they make a decision. We can leave open whether the decision involves an element of commitment or not.

The present proposal, then, is that an agent's subjective options are the *decisions* she can make. The idea is attractive in the light of the discussion from the previous section because one might hope that decisions are automatically under the agent's complete control. When we decide to raise an arm, cook risotto, or go to Aleppo, there is always a chance that the world interferes and we will fail to do what we decided to do. But one thing is guaranteed: we will at the very least have decided to perform the relevant act. There is no causal gap between the decision and the decision, so nothing could possibly interfere; we can be absolutely certain that if we were to make the relevant decision, then we would make the decision; making the decision gives us conclusive evidence that we make the decision.

Or so one might think. But let's have a closer look. First of all, what are the decisions an agent can make? By the analysis from section 2, a decision is available just in case some possible variation of the agent's motivational state would cause the decision. A good candidate for the role of a 'motivational states' in this analysis are an agent's decisions. But arguably we don't want to say that a decision is available if some (available?) other decision would cause it.[12] So perhaps we should say that a decision is available just in case some possible variation of the agent's motivational state would *be* the decision. But now availability of an intention reduces to the intention being a 'possible variation' of the agent's motivational state. So we urgently need a story about how these variations are determined.

Second, on reflection, it is not at all clear that decisions satisfy Anscombe's condition. Can you really become rationally certain that you decide to A merely by deciding to A? If decisions are functionally individuated, this would mean that merely by making a decision, an agent may become rationally certain that she is in a state with a specific functional role. That is far from trivial. The idea also seems vulnerable to the 'anti-luminosity' arguments in [Williamson 2000: ch.4], and it might be challenged on the grounds that the relevant decisions can have 'wide content' so that knowledge of one's decision would imply knowledge of various features of the environment of which the agent may in fact be unaware.

Third, in many decision situations, the available decisions seem too coarse-grained to do the work of options. Jeffrey [1968: 37] considers the case of an experienced marksman aiming at a distant target, taking into account the wind conditions, the quality of the rifle, and various other such factors. The marksman presumably intends to hit the target. Does he decide to hit it? Maybe. Or maybe he decides to shoot at it. In any case, these

12 Compare Moore's [Moore 1912: ch.7] implausible claim that an agent can intend to A just in case if she intended to intend to A, she would succeed. The claim is implausible because even if we can have intentions to form intentions, most intentions are certainly not formed as a result of higher-order intentions; why then should one's ability to form an intention be measured by whether it can be formed in this particular and unusual way? Similarly for decisions.

propositions are too unspecific. There is a reason why the marksman holds the rifle the way he does, pointing in that particular direction. The reason has something to do with his desire to hit the target and his knowledge about the wind etc. If his only options were deciding to hit (or shoot at) the target and deciding something else, he would obviously choose the former; there would be no need to pay attention to the wind. The point is that the marksman has fine-grained control over the precise torque in his joints, but the available torque configurations do not seem to correspond uniquely to relevant intentions or decisions—especially if the marksman is supposed to become certain of these events merely on the basis of his choice. When the marksman holds his rifle at a particular height h and a particular orientation, described by three angles α, β, γ, he is unlikely to be aware of these parameters; he will be even less aware of the torque configurations in his joints. He may be reasonably confident that he is pointing the rifle in some general direction, or that he intends to point in that direction, but that strongly underspecifies his choice.[13]

I will suggest a response to these worries in section 6. First, I want to briefly discuss the lesson Jeffrey himself drew from cases like the marksman (in [Jeffrey 1965: sec. 11.9] and [Jeffrey 1968]).

5 Mixed options

The move from overt acts to internal acts of intending or deciding was motivated by the hope that decision-makers have perfect control and infallible knowledge about these internal acts. Jeffrey realized that this hope is an illusion. An adequate model of decision-making, he suggested, must account for the fact that certainty of contingent matters is generally unattainable; there is always a possibility of error. The 'knowledge without observation' we get from our choices should therefore be understood as partial, probabilistic knowledge: we come to assign new credences to various propositions; these propositions are ordinary propositions about overt acts (or plans), not about a special, luminous realm of internal options. In the case of the marksman, for example, his choice of a stance might make the marksman 30% confident that *he will shoot and hit*, and 70% that *he will shoot and miss*. If the wind picks up, his credences will change and he might choose a different stance which promises a better chance of hitting.

So what is the (subjective) option the marksman chooses? According to Jeffrey, it is not any particular proposition. Rather, it is given by a whole range of (mutually exclusive and

13 In the analysis of perception, the analogous move to postulating intentions as options is to postulate *seemings* as perceptual inputs: what we learn through perception, on this account, is not that the external world is a certain way, but only that it seems to be a certain way. Here, too, the main problem is that propositions about seemings are far too unspecific to explain the change in our beliefs prompted by a perceptual experience.

jointly exhaustive) propositions, with associated probabilities. The marksman chooses *shooting and hitting with probability 0.3; shooting and missing with probability 0.7; not shooting with probability 0.* If he is unsure whether the rifle works or whether his muscles will fail, the range of propositions will include further, low-probability elements such as *failing to move one's finger.*

Anscombe's condition is generalized as follows.

> Jeffrey's condition. An assignment of probabilities $x_1, x_2, \ldots$ over some partition $A_1, A_2, \ldots$ of propositions is an option for an agent only if the agent could rationally come to assign credence $x_1, x_2, \ldots$ to $A_1, A_2, \ldots$ respectively just by making a decision.

Jeffrey's proposal is elegant and undeservedly neglected. But it suffers from two main defects. First, it retracts Jeffrey's fruitful construction of options as propositions. Now the options look more like 'mixed strategies' in game theory, which are also probability measures over acts rather than acts or act propositions themselves. In section 2 I mentioned several advantages of modelling options as propositions; the new proposal threatens to give most of them up. For example, how are we to define the causal expected utility of an option, if the option is not a proposition?[14]

Second, and more seriously, Jeffrey's proposal makes it very hard to give an informative answer to our original question: what options are available to an agent in a concrete decision situation. That is because the available mixed options are highly sensitive to the agent's beliefs, in a way that is hard to make precise. Suppose that one of the marksman's options would rationally make him 30 percent confident that he will hit the target, and none of his options would rationally make him 50 percent confident. Why is that? As Jeffrey points out, "[t]he basis for his belief may be his impressions of wind conditions, quality of the rifle, etc." [Jeffrey 1968: 37]. If the wind increases, the marksman may no longer have an option that corresponds to probabilities 0.3 and 0.7 over $\{$*shooting and hitting*, *shooting and missing*$\}$. We would like to know the mechanics behind this update: why does the information about the wind affect the available options in just that way?

To see the difficulty, 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

14 To see the problem, consider Newcomb's problem with a predictor who is taken to be 90 percent reliable, so that choosing to one-box should make the decision-maker 90 percent confident that she will get a million dollars (and 10 percent that she will get nothing), while choosing to two-box should make her 90 percent confident that she will get a thousand (and 10 percent that she will get a million and a thousand). If we replace the options of one-boxing and two-boxing by these probability assignments over $\{\$M, \$MT, \$T, \$0\}$, it is hard to see how two-boxing could come out as superior.

the left would be useful, a certain signal—a 'motor command'—is sent to the motor system which ideally causes the robot to move to the left. Whether the robot succeeds in moving to the left, however, depends on various factors outside its control: whether there is a glass wall in the way, whether the robot's wheels are blocked, and so forth. If the robot assigns some probability to these possibilities, then *moving left* is not among its decision-theoretic options. Concretely, let's say the robot assigns probability 0.2 to the presence of a wall that would block movements to the left. On Jeffrey's model, the robot then has 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 that increases the probability of the wall to 0.5, by what rule does the relevant option change to *moving left with probability 0.5 and staying in place with probability 0.5*? Any plausible answer I can think of requires that the robot has some other, information-independent way of picking out the option, for example as *issuing such-and-such motor command*, or *moving to the left if there is no obstacle in the way and otherwise staying in place*. But the availability of such information-independent option propositions is precisely what Jeffrey's model denies.[15]

6 Decisions revisited

Let's return to the idea that an agent's options are the decisions she can make. What is a decision? We might understand a decision as some kind of attitude—an intention perhaps—that arises as a result of practical deliberation. That's what I assumed in section 4, and it led to the problems I raised: that practical deliberation arguably can't make us certain about our decisions, and that decisions are often more coarse-grained than the things we choose. But there is another way of understanding decisions, namely as whatever physiological event constitutes the end point of practical deliberation. In the example of our robot, the end point is the issuing of a particular motor command. Human decision-making is more complex. Not only do our decisions often not concern immediate actions, even where they do the relevant process seems to involve a whole hierarchy of control systems. But perhaps the whole hierarchy can be regarded as an efficient mechanism for selecting a motor command. In any case, it is safe to say that practical deliberation always results in some characteristic physiological state that either is the reached decision or realizes that decision. If we identify the agent's options with these

15 A parallel question arises in Jeffrey's account of perception. There, the problem is known as the 'input problem for Jeffrey conditioning', and widely thought to be unsolvable (see e.g. [Field 1978], [Garber 1980], [Christensen 1992], [Weisberg 2009]). My own view is that in either case, the problem is indeed unsolvable within the strict confines of Jeffrey's 'radical probabilism'. It can only be solved by reverting to the traditional picture on which there are propositions to which perception and action confer absolute certainty.

states, the options are guaranteed to have just the right granularity. If the marksman has control over whether he holds the rifle at height $h \pm \epsilon$ rather than $(h + \delta) \pm \epsilon$, but not over whether he holds the rifle at height h rather than $h \pm \epsilon$, then this means that (at some relevant level) his cognitive system can issue a motor command fixing the height to within $h \pm \epsilon$ but no command fixing it to a more narrow range.

Other advantages of identifying options with physiologically individuated decisions are perhaps better to see if we return to our robot. On the present proposal, the robot's immediate objects of choice are motor commands. Let's say α is the command that would normally cause the robot to move left. We may then program the robot to give high prior probability to the assumption that α will causes movements to the left. We might also build in knowledge of exceptions: if there's a wall to the left, α typically doesn't result in the expected movement. Or we could let the robot discover these regularities and exceptions from experience. Either way, it is clear how information about the absence or presence of a wall will affect the robot's credence in whether it will move to the left by issuing the relevant command. The update is a straightforward instance of Bayesian learning.

In general, given the robot's information about the world, every available motor command determines a Jeffrey-type probability measure over possible movements. But unlike in Jeffrey's model, these measures are not taken as unexplained primitives. If the robot has a mixed option represented by an 80 percent probability of moving left and a 20 percent probability of staying in place, that is because it is 80 percent confident that issuing α will move it to the left and 20 percent confident that it won't.

There is one obvious problem with this picture. I have argued that options should be modelled not as physiological states but as propositions. Moreover, we have seen that these propositions should be accessible to the agent at least in the Anscombian sense that the agent could become certain of the proposition merely on the basis of making a decision. Yet when we make a decision, we do not become certain of fine-grained physiological events in our brain, and it does not seem right to classify that as a failure of ideal rationality. (On the contrary, one might argue that it would be irrational to become certain about the specifics of one's neurophysiology on the basis of everyday decisions.)

Fortunately, there is a simple way out. The above considerations suggest that an agent's options should be construed as propositions that stand in a one-one correspondence to motor commands (or more generally, to physiologically individuated decisions) insofar as choosing one of the propositions physiologically corresponds to issuing the corresponding command. There is no reason to assume that the propositions should accurately describe these commands. For example, it does not matter in the least whether our robot represents commands α and β as the propositions *that I issue command* α and *that I issue command* β respectively, rather than the other way round.

Imagine, for the sake of concreteness, that an agent's degrees of belief are defined over

sentences in some language of thought. Let's say the language has sentences describing ordinary acts, states of the environment, and so on. There may also be sentences describing torque positions, beliefs, desires, intentions, etc. None of these are adequate as option propositions. When designing an ideal decision-maker, we therefore need to extend the agent's language of thought by a new range of sentences, for the specific purpose of representing options. In principle, it doesn't matter what these sentences look like; they could be atomic tags: 'X', 'Y', 'Z', etc. Each of the sentences is paired with a decision output—say, a motor command—in such a way that when the agent decides to make true one of the sentences, then the corresponding output is produced. Due to this causal pairing between sentences and outputs, the sentences might be regarded as 'expressing' or 'denoting' corresponding outputs. But from the internal perspective of the agent's cognitive system, they do not carry information about physiology. If an agent gives positive credence to the hypothesis that there is no physical world, then making a decision does not require her to reduce that probability to zero, even though she will rationally become certain of some option proposition.

On the other hand, some aspects of the pairing must be reflected in the decision-makers beliefs. If the agent has no idea about 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. How could the agent acquire the information that choosing 'X' is likely to result in a certain movement, if 'X' doesn't express any genuine proposition about the world? Well, suppose 'X' is paired with a certain command for moving to the left. When this output is produced, then the agent usually ends up moving to the left. So the agent can easily learn that choosing 'X' tends to cause those movements; she can also learn how the movement depends on the inclination of the ground, the presence of walls etc. As before, it also makes sense (from an evolutionary or engineering perspective) to endow the agent with prior probabilities linking the truth of 'X' with movements to the left.

I don't really think that the objects of belief are best modelled as sentences, and my proposal does not assume so. My own view is that an agent's subjective credences and utilities represent a certain causal-functional profile of her cognitive state—chiefly, the way the state would typically manifest itself in behaviour under various conditions, and the way these manifestations change through perceptual input. On a simple version of this picture, to have such-and-such beliefs and desires *is* to be in a state that, among other things, systematically causes acts which maximize expected utility relative to these beliefs and desires. What I want to suggest is that this is not quite right: intentional agents don't maximize the expected utility of their *acts*, on any ordinary conception of acts. Rather, they maximize the expected utility of certain 'option propositions' that are systematically correlated with acts, and even more closely correlated with internal physiological events which normally initiate acts. Propositions, on that view, work somewhat like numbers in the representation of mass or height or temperature: they are abstract objects useful to

pick out, classify, and reason about properties with a characteristic functional profile. (See e.g. [Ramsey 1931], [Lewis 1974], [Loar 1981], [Stalnaker 1984], [Braddon-Mitchell and Jackson 1996], [Beckermann 1996] for this approach to intentional states.) Just as an integer representation might be unsuitable to systematize temperature profiles, so a representation of credences in terms of genuine metaphysical possibilities is, on my proposal, not adequate to fully capture the connection between beliefs and desires on the one hand and behaviour on the other: the space of propositions must be extended by further elements—the 'option propositions'—that do not correspond to genuine ways the world could have been.

There are other conceptions of belief and desire on which the move I have suggested is harder to make. Some authors think of belief as a potentially conscious attitude towards sentences in the agent's public or private language. It is doubtful that we have 'beliefs', in this sense, towards option propositions. If so, the connection between our beliefs, desires, and our actions cannot be fully spelled out on this conception of attitudes. All we can see is something like Jeffrey's picture: when people make choices, their credences towards sentences change without anything becoming absolutely certain; the marksman may become 70 percent confident that he will hit and 30 percent confident that he will miss. What we can't explain is where the available 'mixed options' come from and how they are affected by an agent's evidence.

7 Options and Actions

Most actions involve a complex chain of events. When I turn on the light, a series of motor commands is sent to the muscles in my arm where they cause a series of movements which cause a flipping of the light switch which causes electricity to flow through a circuit which causes the light to go on. At each point in the chain, something could go wrong: the muscles could fail to contract, the arm could fail to move, the switch could fail to flip, the electricity could fail to flow, the light could fail to go on. Since the interfering possibilities are not under my control, I can't guarantee that the light will go on merely by an act of my will. Accordingly, *turning on the light* is not one of my decision-theoretic options—at least not if I give positive credence to the presence of interfering factors. Genuine options should be 'basic acts' over which the agent has full and immediate control: the starting point of the causal chain. That starting point is a decision.[16]

Decisions, like all acts and all events, can be presented in different ways, and the difference often matters for the comparison of expected utilities. The right presentation should, on the one hand, be sufficiently fine-grained to distinguish it from all other decisions the agent could have made, but it should also be accessible to the agent at

16 The decision itself will have causes, but these causes are not part of relevant act; they are (generally) not the result of practical deliberation.

least insofar as she can rationally become certain that she has made that decision merely on the basis of actually making it. The way we normally conceptualize decisions meets neither of these conditions. For example, I have *decided to turn on the light*. But that is not sufficiently fine-grained: I could have decided to turn on the light with my elbow or by throwing a ball at the light switch, which would have resulted in completely different chains of events. It is also doubtful that I can rationally become certain that I decided to turn on the light merely on the basis of making the decision.

The way out, I suggested, is to simply invent new presentations (or quasi-presentations) of the available decisions that satisfy the desired conditions by *fiat*: for every decision an agent can make, add a new proposition to the agent's doxastic space—the algebra over which her credences are defined—to serve as a subjective option so that the option is chosen iff the corresponding decision is made.

I therefore suggest the following update to the decision-theoretic model of rational agents. The model should include a range of 'basic acts' over which the agent is taken to have full and immediate control. In the agent's doxastic space the basic acts are represented one-to-one by special 'option propositions' that are logically independent of (but probabilistically related to) genuine ways the world could be. Practical deliberation is deliberation between these propositions. If in a given situation all option propositions have a well-defined expected utility, deliberation will settle on some proposition with maximal expected utility, and the corresponding basic act will be performed.

There is generally no need to specify the physiological nature of the basic acts. It is more useful to know what other events are normally caused by these acts, and perhaps what the acts would cause under relevant non-normal circumstances. For example, suppose unbeknownst to an agent, her arms are tied down, and it would further her goals to raise her arms. All else equal, we may then predict that the agent will decide in favour of an option proposition of which she is confident that it would bring about a raising of her arms; accordingly, she will perform a basic act that would normally lead to a raising of her arms, but fail to bring about that result in the agent's actual predicament. If the agent's arms are not tied down and no other unusual obstructions are in play, things are even simpler; we can then predict that the agent will indeed raise her arms.

On the model I have proposed, the options do not depend on the agent's information, nor on the concrete situation in which she finds herself. In practice, of course, a lot of options can usually be set aside; one of the features that make decision-making hard for cognitively limited agents is pruning the space of considered options. But from the idealized perspective of classical decision theory, where cognitive costs are ignored, it does no harm to include options that would be pointless in a given context. For example, if an agent knows that her arms are tied down, the option that normally corresponds to raising her arms is unlikely to maximize expected utility, so it makes little difference whether we treat it as a genuine option or not.

Of course, if we turn to an actual agent in an actual decision context, there is still a non-trivial question about her options. There is also a non-trivial question about her subjective probabilities and utilities. I want to treat these questions alike. There are no hard and fast rules for applying abstract, high-level models to concrete situations in the world.

I have suggested that we identify an agent's subjective options with her option propositions. But suppose in a certain decision situation, an agent is certain that if she were to choose option O, then some other proposition A would be true. Then the (causal) expected utility of O is plausibly the same as that of $O\&A$,[17] so we may just as well treat the latter as the relevant option. Moreover, suppose the agent is certain that each of the options $O_1, O_2, \ldots$ would bring about A, and that no other option might do so. Suppose also she assigns the same expected utility to all of $O_1, O_2, \ldots$. Then this expected utility plausibly coincides with the (causal) expected utility of A, and so we can effectively ignore the option propositions and treat A as the relevant option.[18]

To illustrate, imagine a Newcomb Problem. Presumably the decision-maker could issue a variety of motor commands all of which are certain to result in taking both boxes; similarly for taking just the opaque box. All that matters to the expected utility of these commands is whether they lead to one-boxing or two-boxing. Thus we can simplify the problem and identify the options as *one-boxing* and *two-boxing*. The agent plausibly has further options that would not result in taking any box, but these can be ignored because they have lower expected utility.

In general, let's say that a proposition A is *guaranteed* by an option proposition O if (1) the agent is certain that O would bring about A, and (2) the basic act corresponding to O would in fact bring about A. We may then say that A is a *subjective option* for an agent iff it is the logically strongest proposition guaranteed by one of her option propositions.

Option propositions trivially satisfy Anscombe's condition: if the agent makes a decision, she rationally becomes certain of the corresponding option proposition. Subjective options on the present, more liberal, definition also satisfy Anscombe's condition except in the unusual case where an agent's certainty that O would bring about A is undermined by choosing to make true O. I do not know what to do about such cases because I cannot think of a clear example.

The present definition correctly rules out that in the perfect Newcomb Problem the agent's options settle what's in the opaque box: none of the options entails that she will get a million.[19]

17 If $Cr(A\backslash O) = 1$, then plausibly for any X, $Cr(X\backslash A\&O) = Cr(X\backslash O)$.

18 Friends of evidential decision theory might want to swap all requirements of subjunctive conditional certainty by indicative conditional certainty.

19 If condition (1) is swapped by a condition of indicative conditional certainty (see the previous footnote),

The external condition (2) prevents subjective options from being 'accessible' in a strong sense as captured for example by Hedden's supervenience requirement. But note that the more narrow option propositions may satisfy even that requirement: any intrinsic duplicate of a rational agent plausibly has the same option propositions.

Condition (2) alone is essentially the availability condition from section 2. The agent's basic acts are what I previously referred to as the 'possible variations of her motivational state'. (I promised that I would eventually explain what that means.)

To conclude, let me briefly return to Jeffrey's insight that our choices are part of the natural world and thus should be represented as ordinary propositions. The model I have proposed suggests that this is not quite right. From the decision-maker's perspective, her choices select propositions that (at least in part) stand outside the realm of ordinary physical propositions, being only contingently and probabilistically related to physical propositions about the world. This might be taken to explain or even vindicate the 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 physical propositions. I myself am sceptical of these conclusions, but I will not pursue the matter any further.

References

Arif Ahmed [2013]: "Causal Decision Theory: A Counterexample". *The Philosophical Review*, 122: 289–306

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

Ansgar Beckermann [1996]: "Is there a problem about intentionality?" *Erkenntnis*, 45: 1–23

Jonathan Bennett [1988]: *Events and Their Names*. Oxford: Clarendon Press

— [1995]: *The Act Itself*. Oxford: Clarendon Press

David Braddon-Mitchell and Frank Jackson [1996]: *Philosophy of Mind and Cognition*. Oxford: Blackwell

Michael Bratman [1987]: *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press

John Broome [1991]: *Weighing Goods*. Oxford: Blackwell

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

and the opaque box in fact contains a million, then *one-boxing and getting a million* does come out as a subjective option (or entailed by a subjective option). *Two-boxing and getting a thousand* still doesn't, because it violates the 'actual availability' condition (2).

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

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

Holly S. Goldman [1978]: "Doing the Best One Can". In A. Goldman and J. Kim (Eds.) *Values and Morals,* Dordrecht: Reidel, 185–214

Alan Hájek [2017]: "Deliberation Welcomes Prediction". Manusccript

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

Terry Horgan [1985]: "Newcomb's problem: A stalemate". In R. Campbell and L. Sowden (Eds.) *Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem,* Vancouver: University of British Columbia Press

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

Frank Jackson and Robert Pargetter [1986]: "Oughts, Options, and Actualism". *The Philosophical Review*, 95: 233–255

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

— [2002]: "Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions". *Philosophical Studies*, 110: 69–102

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

Gregory S. Kavka [1983]: "The toxin puzzle". *Analysis*, 43: 33–36

Isaac Levi [2007]: "Deliberation Does Crowd Out Prediction". In T. Rønnow-Rasmussen, B. Petersson, J. Josefsson and D. Egonsson (Eds.) *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz,* www.fil.lu.se/hommageawlodek

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

— [1976]: "The Paradoxes of Time Travel". *American Philosophical Quarterly*, 13: 145–152

— [1979]: "Counterfactual Dependence and Time's Arrow". *Noûs*, 13: 455–476

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

Brian Loar [1981]: *Mind and Meaning*. Cambridge: Cambridge University Press

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

George E. Moore [1912]: *Ethics*. London: Williams and Norgate

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

Frank Ramsey [1931]: "Truth and Probability (1926)". In R.B. Braithwaite (Ed.) *Foundations of Mathematics and other Essays,* London: Routledge & P. Kegan, 156–198

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

Wolfgang Schwarz [2015]: "Lost memories and useless coins: Revisiting the absendminded driver". *Synthese*, 192: 3011–3036

— [Unpublished]: "Imaginary Foundations". Unpublished manuscript, 2014

Teddy Seidenfeld [1984]: "Comments on causal decision theory". In *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association,* 2. Philosophy of Science Association, 201–212

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

Wolfgang Spohn [1977]: "Where Luce and Krantz do really generalize Savage's decision model". *Erkenntnis*, 11(1): 113–134

Robert Stalnaker [1984]: *Inquiry*. Cambridge (Mass.): MIT Press

Travis Timmerman and Yishai Cohen [2016]: "Moral Obligations: Actualist, Possibilist, or Hybridist?" *Australasian Journal of Philosophy*, 94(4): 672–686

Peter van Inwagen [1983]: *An Essay on Free Will*. Oxford: Clarendon 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

Timothy Williamson [2000]: *Knowledge and its Limits*. Oxford: Oxford University Press