# Options and the Subjective *Ought*

Brian Hedden

*Philosophical Studies* 158 (2)

## 1 Introduction

Determing what you ought to do can be broken down into two stages. The first stage is determining what your options are, and the second stage is ranking those options. While the second stage has been widely explored by philosophers of all stripes, from ethicists to decision theorists to epistemologists to action theorists, the first stage has gone largely unaddressed. And yet, without a theory of how to conceive of your options, the theory of practical rationality - of how you ought to act - will be incomplete.

I will argue that the fact that what you ought to do depends on your uncertainty about the world ultimately forces us to conceive of your options as consisting of all and only the *decisions* you are presently able to make. In this way, *ought*s apply only to decisions, and not to the non-mental acts that we ordinarily evaluate for moral and rational permissibility.

This change in our conception of your options is not of mere bookkeeping interest; it has substantive implications for first-order normative theorizing. First, it directly takes into account the potential costs and benefits of the decision itself in determining what you ought to do. Second, it provides a principled solution to Chisholm's Paradox, in which your doubts about your own self-control seem to give rise to conflicting claims about what you ought to do. These two cases provide direct support for my theory of what your options are, supplementing the more theoretical reasons considered earlier.

## 2 The Subjective *Ought*

In this paper, I will be focused on the sense of *ought* in which what you ought to do depends on your beliefs about how the world is. Consider: Your friend has a headache, and you have some pills that you justifiably believe to be pain relievers. But you're wrong. They are really rat poison. Ought you give the pills to your friend?

While there may be a sense in which the answer is 'no,' there is also a sense in which the answer is 'yes.' Call the sense of *ought* in which you ought to give your friend the pills the *subjective ought.* What you subjectively ought to do depends not on how the world actually is, but on how you believe the world to be. Since you believe the pills to be pain relievers, you subjectively ought to hand them over, even though your belief is false.[1]

The subjective *ought* has three important roles to play. First, the subjective *ought* is supposed to give you guidance about how to proceed (morally or prudentially speaking), given your uncertainty about the world. It is to be action-guiding, in at least the minimal sense that you are typically in a position to know what you subjectively ought to do.[2]

Second, the subjective *ought* plays an evaluative role. In ethics, it links up tightly

[1] The sense of *ought* in which you ought not give your friend the pills is often called the *objective ought.* In the case of prudential rationality, what you objectively ought to do is whatever would in fact maximize your utility, while what you subjectively ought to do is bring about whichever proposition has highest *expected* utility. In ethics, consequentialists will likely say that what you objectively ought to do is whatever would maximize moral value (total world happiness, say), while what you subjectively ought to do is bring about whichever proposition has highest *expected* moral value. The objective/subjective distinction can also be drawn in non-consequentialist moral theories, although there is less consensus on how exactly to do so.

[2] To emphasize - I am understanding the requirement that the subjective *ought* be 'action-guiding' as the requirement that you be *in a position to know* what you ought to do. Thus, for the subjective *ought* to be action-guiding, it is not required that you always in fact know what you ought to do (for you might make a mistake or fail to consider the question), nor is it required that you consciously employ the theory of the subjective *ought* in coming to a conclusion about what you ought to do. All that is required for the subjective *ought* to be action-guiding, in my sense, is for facts about what you ought to do to be in principle accessible to you.

with praise and blame. If you give your friend the pills, you are praiseworthy, or at least exempt from blame, for having done so, despite the disastrous consequences, since you did what you subjectively ought to have done. With respect to prudential rationality, you are subject to rational criticism unless you do what you prudentially subjectively ought to do.

Third, in the case of prudential rationality, the subjective *ought* plays a role in the prediction and explanation of behavior. Knowing what you believe and desire, we can predict what you will do, against the background assumption that you are rational. We predict that you will do that which you (prudentially) subjectively ought to do. Why is this? It is close to an analytic truth that fully rational agents successfully fulfill their rational requirements; they do what they rationally ought to do (that is, what they prudentially subjectively ought to do), believe what they rationally ought to believe, and so forth. So, given the assumption that you are rational, and given that this entails that you fulfill your rational requirements, it follows straightforwardly that you will perform the action that you subjectively ought to perform.[3] And we can explain why you did something by pointing out that you are rational and that, given your beliefs and desires, that was the thing you subjectively ought to have done.

So, the subjective *ought* should play an action-guiding, evaluative, and predictive/explanatory role in our normative theorizing. I favor this view, on which one sense of *ought* plays all three roles, both because it is parsimonious and because the three roles are interrelated. To take just one example, it is plausible that a sense of *ought* cannot play the evaluative role without also playing the action-guiding role (and vice-versa), since an agent cannot be criticized for performing some action or having some belief if she was in no way even

[3] Importantly, we only predict that you will do what you subjectively ought to do when we hold onto the background assumption that you are rational. But often, we have evidence that you fall short of ideal rationality in various respects, and in these cases we will not want to predict that you will do what you subjectively ought to do. For instance, we may have evidence from behavioral economics that you employ certain biases and heuristics that lead you to be irrational in certain systematic ways, and if such biases and heuristics are relevant in the case at hand, we will not want to predict that you will in fact do what you subjectively ought to do.

in a position to know that she oughtn't have performed that action or held that belief.[4] Importantly, however, my argument in this paper does not depend on this assumption that there is one sense of *ought* which can play all three roles. This is because each of the desiderata presented below that support my favored view of options is motivated by each of these three roles. Therefore, one will still be pushed toward my view of options even if one thinks that we will need multiple *ought*s, one of which will play the action-guiding role, another the evaluative role, and a third the predictive/explanatory role.[5]

## 3 The Problem of Options

Giving a theory of the subjective *ought* requires giving a theory of what your options are. Because the subjective *ought* is sensitive to your uncertainty about the world, this theory of options must also take into account this uncertainty if the subjective *ought* is to play its action-guiding, evaluative, and predictive/explanatory roles.

The problem of specifying what your options are, in a way that is appropriately sensitive to your uncertainty about the world, can be made precise by considering expected utility theory, the dominant account of the subjective *ought* of prudential rationality. (For clarity, I focus on prudential rationality in this paper, though the considerations I raise will apply equally to the case of ethics.) Expected utility theory provides a framework for assigning numbers to propositions, relative to a credence function P (representing the agent's doxastic, or belief-like, state) and a utility function U (representing the agent's conative, or desire-like, state). The expected utility of a proposition A is the sum of the utilities assigned to the possible outcomes $O_i$, weighted by the agent's credence that $O_i$

[4] This is just to express sympathy with internalism about practical and epistemic rationality. Externalists will predictably be unsympathetic, but in some sense this paper can be seen as exploring the viability of internalism about practical rationality and determining how internalists should conceive of a decision-maker's options.

[5] Thanks to Matthew Noah Smith for raising this worry.

will come about, conditional on A.[6] More formally:

**Expected Utility**: $EU(A) = \sum_i P(O_i|A)U(O_i)$

Expected utility theory, then, provides a way of ranking propositions. The connection between this ranking and prudential rationality is standardly expressed in the slogan, 'You ought to maximize expected utility.' That is, you ought to bring about the proposition with highest expected utility.

But now a problem arises. Expected utilities can be assigned to any proposition whatsoever. Consider, for example, the proposition that someone discovered a cure for cancer two weeks ago. This proposition has a very high expected utility. But even if this proposition has higher expected utility than any other proposition, there is no sense in which I ought to bring it about that someone discovered a cure for cancer two weeks ago.[7] Intuitively, the expected utility assigned to the proposition that someone cured cancer two weeks ago is irrelevant to the question of what I ought to do now because bringing about this proposition simply isn't one of my options!

Therefore, in order to have a theory of what I ought to do, we need some way of specifying a narrower set of propositions, such that what I ought to do is bring about the proposition with highest expected utility *in that narrower set.* Let us call such a narrower set of propositions a *set of options.* Our task, then, is to say what counts as a set of options.

In what follows, I argue that a theory of options must satisfy three desiderata: First, if something is an option for you, you must be able to do it. Second, if something is an option for you, you must *believe* that you are able to do it. Third, what your options are

[6] This is the definition of expected utility employed in *Evidential Decision Theory* and is sometimes called *evidential expected utility.* The definition of expected utility employed in *Causal Decision Theory* is slightly more complex, but the distinction between evidentialist and causalist views of expected utility will not matter in what follows.

[7] Of course, there would certainly be other propositions with higher expected utility, such as the proposition that someone discovered a cure for cancer *and* deposited $10,000 in my bank account. In fact, it may be that there is no proposition with highest expected utility.

must supervene on your beliefs and desires. I reject three initially tempting theories of options on the grounds that they each violates at least one of these desiderata. I then present my own theory of options, which I argue does satisfy the desiderata: your options are all and only the *decisions* you are presently able to make.

## 4 First Pass: Options as Actual Abilities

Consider again the example of the proposition that someone found a cure for cancer a couple weeks ago. One obvious reason for thinking that there's no sense in which I ought to bring it about that someone has found a cure a couple weeks ago is that *that's just not something I can do*! We must respect the principle that *ought implies can* if the subjective *ought* is to play the action-guiding, evaluative, and predictive roles that are supposed to be played by the subjective *ought*. First, the theory gives poor guidance to an agent if it tells her to do something that she cannot do. Second, an agent is not subject to any rational criticism if she fails to do something which in fact she couldn't have done. Third, we would not want to predict that a rational agent would do something which in fact she cannot do. This yields a first desideratum for a theory of options:

> **Desideratum 1**: If a proposition P is a member of a set of options for an agent S, then S is able to bring about P.

A minimal theory, on which *ought implies can* is the only restriction on what counts as a an option, would be the following:

> **Proposal 1**: A set of propositions is a set of options iff it is a maximal set of mutually exclusive propositions, each of which is such that the agent has the ability to bring it about.[8]

[8] The set must be maximal in the sense that there is no other proposition incompatible with the members of that set which is also such that the agent has the ability to bring it about. Note that this

Not only is Proposal 1 intuitively attractive, but it also has a formidable pedigree, having been defended by many prominent decisions theorists, including Richard Jeffrey and David Lewis.[9] But Proposal 1 is too permissive about what can count as an option for an agent.

First, Proposal 1 can yield unacceptable results in cases where the agent is uncertain or in error about her own abilities. Consider:

> **Raging Creek**: 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, since she would then be able to continue hiking with no detour. But she has serious doubts about whether she can in fact ford the creek. After all, the water is up to mid-thigh and flowing fast.

Ought Jane ford the creek? I suggest that the answer is 'no.' First, the subjective *ought* is supposed to be action-guiding, in that what an agent subjectively ought to do depends solely on facts that the agent is in a position to know. But since Jane is not in a position to know that she is able to ford the creek, she is not in a position to know that she ought to ford the creek. Second, Jane is not subject to any rational criticism for failing to ford the creek. It would be bizarre to call her irrational for failing to ford it, given that she had serious doubts about her ability to do so. Third, we would not predict that Jane will

proposal allows for the possibility of multiple sets of options for an agent, since we can cut up the things that she is able to bring about in more or less fine-grained ways and still have a maximal set of mutually exclusive propositions, each of which she is able to bring about.

[9] Jeffrey (1965, 84) regards options as acts, where 'An act is then a proposition which is within the agent's power to make true if he pleases.' And in 'Preference among Preferences,' he writes that 'To a first approximation, an option is a sentence which the agent can be sure is true, if he wishes it to be true' (Jeffrey (1992, 164)). In 'Causal Decision Theory,' Lewis writes, 'Suppose we have a partition of propositions that distinguish worlds where the agent acts differently...Further, he can act at will so as to make any one of these propositions hold; but he cannot act at will to make any proposition hold that implies but is not implied by (is properly included in) a proposition in the partition. The partition gives the most detailed specifications of his present action over which he has control. Then this is a partition of the agents' alternative *options*' (Lewis (1981, 7)).

ford the creek, given the information that she is rational and has these beliefs and desires. This case suggests that we must add a second desideratum for a theory of options which requires that something can count as an option for an agent only if she believes she is able to to it:

> **Desideratum 2**: If a proposition P is a member of a set of options for an agent S, then S believes that she is able to bring about P.[10]

Another way to see the need for Desideratum 2 is this: The calculation of the expected utility of an action $\phi$ fails to take into account the possibility that the agent might try to $\phi$ but fail to succeed in this effort. This is because the expected utility of $\phi$ is the sum the values of the possible outcome states, weighted by probability of that outcome state *given the assumption that the act $\phi$ is performed.* But the probability of an outcome in which the agent tries to $\phi$ but fails, conditional on her $\phi$-ing, is of course zero! So the disutility of any outcome in which the agent tries to $\phi$ but fails is multiplied by zero in the expected utility calculation, and hence not taken into account in the evaluation of the act of $\phi$-ing. But as **Raging Creek** illustrates, what would happen if the agent were to try but fail to $\phi$ is often very, very important to the agent's own deliberations and to our evaluation of her! Our response to this problem should be to think of an agent's options as including only actions that the agent is confident she can perform and thereby exclude from consideration as options any actions which the agent thinks she might try but fail to perform. Hence Desideratum 2.[11]

[10] One might prefer, here and elsewhere, to replace talk of what the agent *actually* believes and desires with talk of what the agent *ought* to believe and desire. In this way, what an agent subjectively ought to do would depend not on what she believes and desires, but on what she ought to believe and desire. Importantly, adopting this view will not affect the arguments in this paper; one will still be pushed to adopt my favored theory of options. I will continue to put things in terms of the agent's actual beliefs and desires for the sake of consistency, and also because I favor keeping epistemic and practical rationality distinct. In cases where an agent has beliefs (or desires) that she ought not have, but acts in a way that makes sense, given those misguided beliefs (or desires), we should criticize her for being epistemically irrational without also accusing her of practical irrationality.

[11] Heather Logue has pointed out to me that Desideratum 2 may not actually be necessary to motivate my favored theory of options, since my theory of options may also be the only one which can satisfy both

Second, since actual abilities do not supervene on beliefs and desires, Proposal 1 entails that what an agent subjectively ought to do will not supervene on beliefs and desires, either. If two agents have the same beliefs and desires but different abilities, then the sets of options relevant for assessing what they ought to do will also be different, with the result that they ought to do quite different things.[12] Consider:

> **Jane's Doppelgänger**: Jane, facing the raging creek, is in fact able to ford it, and among the things she is able to do, fording the creek is best (has highest expected utility). But her doppelgänger Twin Jane, who has the same beliefs and desires as Jane, faces her own raging creek, but Twin Jane is unable to ford it. Among the things that Twin Jane is able to do, turning back and heading home is best (has highest expected utility).

On Proposal 1, Jane ought to ford the creek, while Twin Jane ought to turn back and head home. But this is an implausible result. First, if Jane and Twin Jane are in exactly the same mental state but ought to do quite different things, neither is in a position to determine that she ought to ford the creek rather than turn back, or vice versa. So again, the subjective *ought* would be in an important sense insufficiently action-guiding.[13]

Desideratum 1 and Desideratum 3 (below). Still, I include Desideratum 2 since I think it is a genuine desideratum, even if it is not needed to motivate my view.

[12] Of course, if we think of *oughts* as attaching to act tokens, rather than act types, there is a harmless sense in which what an agent ought to do will not supervene on that agent's mental states. Perhaps my physically identical doppelgänger and I are in exactly the same mental states, but while I ought to bring it about that *I* donate money to charity, my doppelgänger ought to bring it about that *he* donate money to charity. There is nothing disconcerting about this. It would be more problematic if what an agent ought to do, put in terms of act types like *donating to charity* (perhaps modelled using sets of centered worlds instead of sets of worlds), failed to supervene on beliefs and desires. This more worrying type of failure of supervenience is entailed by Proposal 1.

[13] Of course, supervenience of what an agent ought to do on her beliefs and desires is not by itself sufficient for her to be in a position to know what she ought to do. She must also know her beliefs and desires. The important point is that self-knowledge and supervenience of *oughts* on beliefs and desires are individually necessary and jointly sufficient for the agent to be in a position to know what she ought to do. Therefore, if supervenience fails, then even a self-knowing agent would not be in a position to know what she ought to do. Importantly, knowledge of one's beliefs and desires is already required for one to be in a position to know what one ought to do, since even knowing what one's options are, one needs to know what one believes and desires in order to know how to rank those options. Provided that an agent's options supervene on her beliefs and desires, there are no obstacles to her being in a position

Second, it is implausible to evaluate Jane and Twin Jane differently (criticizing the one but praising the other) if they perform the same action (heading home, say) after starting out in the same mental state, but this is what would be required if we say that *ought*s fail to supervene on beliefs and desires and that Jane and Twin Jane differ in what they ought to do. Third, consider the role of the subjective *ought* in the prediction of behavior. We predict that an agent will perform the action that she ought to perform. So we should predict that Jane will ford the creek, while Twin Jane will immediately do an about-face and head straight home. But this is bizarre! They are in exactly the same mental state, after all! If they displayed such radically different behavior, it would appear that at least one of them wasn't fully in control of her actions (and hence not fully rational). In general, the failure of what an agent ought to do to supervene on her beliefs and desires would entail a sort of externalism about practical rationality. But externalism about practical rationality is a tough pill to swallow, as it would keep the subjective *ought* from adequately playing its action-guiding, evaluative, and predictive roles in our theorizing. This suggests adding a third desideratum for a theory of options:

> **Desideratum 3**: If something is an option for an agent S, then it is an option for any agent with the same beliefs and desires as S.

While Proposal 1 satisfies Desideratum 1, it allows too many things to count as options for an agent and violates two other compelling desiderata.

## 5 Second Pass: Options as Believed Abilities

The subjective *ought* is supposed to be sensitive to how the agent takes the world to be. Now, as noted earlier, there are two stages to determining what an agent ought to do. The first stage involves identifying the agent's options, and the second stage involves ranking

to know what her options are that are not already obstacles to her being in a position to know how those options are to be ranked.

those options. At the second stage, we have a compelling framework (expected utility theory) for ranking options in light of an agent's beliefs and desires. But this progress goes to waste if at the first stage we charactize an agent's options in a manner determined not by how the agent believes the world to be, but only by how the world actually is. Proposal 1 made this mistake by taking an agent's options to be determined by her actual abilities, regardless of what she believed about her abilities.

With this in mind, we might conceive of an agent's options as all and only the things she believes she is able to do. More precisely:

> **Proposal 2**: A set of propositions is a set of options iff it is a maximal set of mutually exclusive propositions, each of which is such that the agent believes she has the ability to bring it about.[14]

This proposal satisfies Desiderata 2 and 3, which were Proposal 1's downfall. First, in cases where an agent can do something, but doubts whether she can do it, that thing will not count as an option for her. Hence, when Jane arrives at the raging creek and doubts whether she can ford it (even though she in fact can), fording the creek will not be one of her options, and so we will avoid the counterintuitive result that she ought to ford the creek. Second, because an agent's options are wholly determined by her beliefs (in particular her beliefs about her own abilities), what an agent ought to do will supervene on her beliefs and desires, as required.

But Proposal 2 is also a step back, for it fails to satisfy Desideratum 1 (*ought implies can*). If Jane believes that she can ford the creek, and fording the creek has highest expected utility among the things that Jane believes she can do, then we get the result that Jane ought to ford the creek, even if she is in fact unable to do so. And again, *ought implies can* is non-negotiable. First, our theory gives an agent poor guidance if it tells her to do something that she in fact cannot do. Second, an agent is subject to no rational

[14] Again, the set must be 'maximal' in the sense that there is no other proposition incompatible with the members of that set which is also such that the agent has the ability to bring it about.

criticism (or blame, in the ethical case) for failing to do something she was unable to do. And third, we would not want to predict that an agent will perform an action that she is unable to perform.[15]

## 6 Third Pass: Options as Known Abilities

Proposal 1 failed because it had an agent's options being determined solely by how the world actually is, irrespective of how the agent believes the world to be. Proposal 2 failed because it had an agent's options being determined by how she believes the world to be, irrespective of how it actually is. Perhaps these problems can be solved by taking an agent's options to be the things that are deemed to be options by both Proposal 1 and Proposal 2. That is, we might take an agent's options to be the set of propositions that she correctly believes she can bring about. At this point, we might even replace the mention of true belief with reference to knowledge, giving us:

> **Proposal 3**: A set of propositions is a set of options iff it is a maximal set of mutually exclusive propositions, each of which is such that the agent knows that she is able to bring it about.[16]

This conception of an agent's options satisfies Desiderata 1 and 2. It will yield the attractive result that if an agent ought to do something, then she is able to do it and believes that she is able to do it.

But this proposal violates Desideratum 3, that what an agent's options are should supervene on her beliefs and desires. This is because if one agent knows that she is able to perform some given action, while another falsely believes that she is able to perform

[15] A close cousin of Proposal 2 would characterize an agent's options in normative terms, so that an agent's options consist not of the things which she actually believes she can do, but rather of the things which she *ought* to believe she can do. This proposal, however, will likewise violate Desideratum 1 and is unacceptable on this account.

[16] Once again, the set must be 'maximal' in the sense that there is no other proposition incompatible with the members of that set which is also such that the agent knows that she is able to bring it about.

that action, then it will count as an option only for the first agent, even if the two of them have the same beliefs and desires. Recall **Jane's Doppelgänger**. Jane and Twin Jane have the same beliefs and desires. But Jane is able to ford the creek and knows that she is able to do so, whereas Twin Jane is unable to ford the creek and so falsely believes that she is able to do so. So on Proposal 3, fording the creek will count as an option for Jane but not for Twin Jane, and so it may be that Jane ought to ford the creek whereas Twin Jane ought to do something quite different. But as we saw earlier, this is an unacceptable result.

The root problem is this: In order for the subjective *ought* to play its theoretical roles, it is important that an agent is at least typically in a position to know what her options are. But Proposal 3 does not have this result. On Proposal 3, an agent will typically only be in a position to know, of the things that are options for her, that they are options for her[17]; she will not typically be able to know, of the things that are not options for her, that they are not options for her. For instance, if Twin Jane justifiably but falsely believes that she is able to ford the creek, then she will not be in a position to know that fording the creek is not an option for her. So, it is important that an agent typically be able to tell whether or not something is an option for her, but Proposal 3 fails to give this result because it violates Desideratum 3, that an agent's options should supervene on her beliefs and desires.

## 7 Options as Decisions

The proposals considered above each failed to satisfy one or more of our three desiderata on a theory of an agent's options. First, if P is an option for an agent, then she is actually

[17] Actually, this assumes a perhaps controversial application of the KK principle, which states that when an agent know that P, she is in a position to know that she knows P. This is because on Proposal 3, knowing that something is an option for you requires knowing that you know you are able to bring it about. If KK fails, then so much the worse for Proposal 3.

able to bring about P. Second, if P is an option for an agent, then she believes she is able to bring about P. Third, if P is an option for an agent, then it is also an option for any agent with the same beliefs and desires.

But is it even *possible* for an account to satisfy all three of these desiderata? I think so. The key is to think of an agent's options as consisting of the *decisions* open to her. Indeed, in cases where an agent is uncertain about whether she is able to do something like ford a creek, it is natural to think that the option we should really be evaluating is her *deciding* to ford the creek, or deciding to try to ford the creek. I suggest that the things that are in the first instance evaluated for rationality or irrationality are *always* the agent's decisions. This gives us a new proposal, which I will call **Options-as-Decisions**:

> **Options-as-Decisions**: A set of propositions is a set of options for agent S at time t iff it is a maximal set of mutually exclusive propositions of the form *S decides at t to* $\phi$, each of which S is able to bring about.[18]

Then, the highest-ranked such proposition (by expected utility) will be such that the agent ought to bring it about. That is, she ought to make that decision. (I am intending 'S decides to $\phi$' to be read in such a way that is incompatible with 'S decides to $\phi \wedge \psi$,' even though there is a sense in which if you decide to do two things, you thereby decide to do each one. If you have trouble getting the intended reading, just add in 'only' to the proposal, so that a set of options is a maximal set of propositions of the form *S only decides at t to* $\phi$, each of which the agent is able to bring about.)

But will **Options-as-Decisions** satisfy the three desiderata? This all depends on what it takes for an agent to be able to make a given decision. I do not want to commit myself to any particular account of what it takes to be able to make a decision, but I

[18] Once again, 'maximal' means that there is no proposition of the form *S decides at t to* $\phi$ which is not a member of the set but which is incompatible with each member of the set. Note that maximality and mutual exclusivity apply not to the *contents* of decisions, but to propositions about which decision was made. Hence the set {*S decides at t to* $\phi$,*S decides at t not to* $\phi$} will not count as a set of options, since it does not include propositions about other decisions that S might have made (e.g. the proposition that S decides at t to $\psi$.

will argue that on any attractive theory of decisions, which decisions an agent is able to make will supervene on her beliefs and desires and thereby allow **Options-as-Decisions** to satisfy the three desiderata.

As one example of a prominent theory of decisions, Bratman (1987) holds that you can make any decision that you do not believe would be ineffective: You are able to decide to $\phi$ if and only if you do not believe that, were you to decide to $\phi$, you would fail to $\phi$.[19]

If Bratman is right, then **Options-as-Decisions** satisfies Desiderata 1-3. First, it satisfies Desideratum 1 (that if something is an option for an agent, then she is able to bring it about), since it is part of **Options-as-Decisions** that a proposition of the form *S decides at t to* $\phi$ can count as an option for S only if S is able to bring about that proposition. Second, it satisfies Desideratum 2 (that if something is an option for an agent, then she believes she is able to bring it about), at least insofar as the agent knows what she believes. If an agent knows what her beliefs are, then she will know whether she believes that, were she to decide to $\phi$, she would $\phi$. And hence, she will know whether she is able to decide to $\phi$. Third, it satisfies Desideratum 3 (that what an agent's options are supervenes on her beliefs and desires), since on Bratman's view, which decisions an agent is able to make, and so what her options are, is entirely determined by her beliefs.

Even if Bratman's view is incorrect, other attractive views of abilities to make decisions still yield the result that **Options-as-Decisions** satisfies Desiderata 1-3. One might hold that in order to be able to decide to $\phi$, you must not only lack the belief that your decision to $\phi$ would be ineffective (as Bratman holds); you must also have the belief that your

[19] Actually, Bratman is discussing intentions, but I think that the relevant considerations apply equally to decisions, insofar as there is any difference between decisions and intentions. This theory of abilities to make decisions gains support from Kavka's Toxin Puzzle (Kavka (1983)). Suppose that in one hour, you will be offered a drink containing a toxin which will make you temporarily ill. Now, you are offered a large sum of money if you make the decision to drink the beverage. You will receive the money even if you do not then go ahead and drink the beverage; the payment depends only on your now making the decision to drink it. It seems that you cannot win the money in this case; you cannot decide to drink the beverage. Why not? Because you believe that, if you were to make the decision to drink the beverage, you would later reconsider and refuse to drink it. You cannot make a decision if you believe you will not carry it out. Supposing that this is the only restriction on agents' abilities to make decisions, we get Bratman's theory of abilities to make decisions.

decision to $\phi$ would be effective.[20] Or one might hold that whether or not you are able to decide to $\phi$ depends not just on your beliefs, but also on your desires. So, it might be that you are unable to decide to commit some horrible murder, even though you believe that, were you to manage to make this decision, you would indeed carry it out. You just have an incredibly deep aversion to making this decision, and this aversion prevents you from being able to do so.

But even if Bratman is wrong and one of these alternative accounts of abilities to make decisions is correct, **Options-as-Decisions** would still satisfy Desiderata 1-3, since even on these alternative accounts, which decisions you are able to make is wholly determined by your beliefs and desires. First, a decision would count as an option for an agent only if she is able to make it (by the wording of **Options-as-Decisions**), thereby satisfying Desideratum 1. Second, insofar as the agent knows what her beliefs and desires are, she will know which decisions she is able to make, thereby satisfying Desideratum 2. Third, since which decisions an agent is able to make will be fully determined by her beliefs and desires, which options she faces will supervene on her beliefs and desires, thereby satisfying Desideratum 3. So, in general, if which decisions you are able to make depends solely on your mental states, **Options-as-Decisions** will satisfy Desiderata 1-3.

But what if an agent's abilities to make decisions are restricted not just by her own mental states, but also by external forces? Frankfurt (1969) considers the possibility of a demon who can detect what's going on in your brain and will strike you down if he finds out that you are about to make the decision to $\phi$. Plausibly, you lack the ability to decide to $\phi$, even if you believe that, were you to decide to $\phi$, you would $\phi$. The possibility of

[20] Anscombe (1957) famously holds that in order to be able to decide to $\phi$, you do not even need to lack the belief that your decision to $\phi$ would be ineffective. You can make decisions that you believe you will not carry out. For instance, as you are being led to the interrogation room, you can decide not to give up your comrades, even though you know you will crack under the torture. Some have interpreted Anscombe as holding that there are *no* restrictions on which decisions you are able to make. If this (admittedly somewhat implausible) view is true, **Options-as-Decisions** will still satisfy Desiderata 1-3, for trivially an agent will always be able to know which decisions she make make, and which decisions she can make will supervene on her beliefs and desires.

such demons threatens the claim that which decisions you are able to make supervenes on your mental states, since which decisions you can make depends also on whether or not such a demon is monitoring you. It also threatens the claim that you are always in a position to know which decisions you are able to make, since you are not always in a position to know whether such a demon is watching you.

In response to this worry, I find it plausible that if a Frankfurtian demon is monitoring you with an eye toward preventing you from deciding to $\phi$, then you lack the capacity to exercise your rational capacities which is necessary in order for you to be subject to the demands of prudential rationality in the first place. Suppose that the decision to $\phi$ looks best out of all the decisions you believe you are able to make, but a demon will strike you down if it detects that you are about to $\phi$. What ought you to do in this case? Certainly, it is not that you ought to make some decision other than the decision to $\phi$, since all such decisions look inferior. And it is not the case that you ought to decide to $\phi$, since *ought* implies *can*. Instead, there simply isn't anything that you ought to *do*; rather, you ought to be in a state of being about to decide to $\phi$, where this will lead to your being struck down before you are actually able to *do* anything at all. The rational *ought* thus only applies to agents who are not being disrupted by Frankfurtian demons in this way, and so once we restrict our attention to agents to whom the rational *ought* applies, which options an agent has will both supervene on her beliefs and desires and be knowable by her.

I conclude that **Options-as-Decisions** will satisfy the three desiderata to which a theory of options is subject. Indeed, I think that it is the *only* theory of options which can satisfy these desiderata.

# 8 Costs of Decisions

The question of what your options are is not a mere terminological issue. In this section and the next, I consider two sorts of cases in which **Options-as-Decisions** yields attractive treatments of particular sorts of decision situations.

First, **Options-as-Decisions**, unlike other theories of options, directly takes into account the potential costs and benefits of the decision itself. Consider:

> **God and Church**: You are deliberating about whether or not to skip Church tomorrow. You would prefer to stay home, but you believe that you will incur God's wrath if you decide to do so. However, you believe that God only punishes people who *decide* to avoid church; He does not punish forgetfulness or sleeping through alarms.

On **Options-as-Decisions**, God's punishment will be directly factored into the advisability of deciding to stay home. It is the decision itself which is evaluated for expected utility. Since you believe that this option (deciding to stay home) will put you at the mercy of God's wrath, it has very low expected utility, and so you ought not decide to stay home.

But on Proposals 1-3, we can directly evaluate the options of staying home and attending church (rather than just decisions to do such things). Most likely, you have some high credence that if you stay home, it will be the result of a decision to stay home. But you also have some credence that if you stay home, it will be the result of forgetfulness or an ineffective alarm clock. In calculating the expected utility of staying home, the disutility of God's wrath is weighted by your credence that your staying home would follow a decision to do so. Then, if your credence that your staying home would be the result of a decision to do so (rather than forgetfulness or sleepiness) is sufficiently low, Proposals 1-3 will say that you ought to stay home, even in a case where **Options-as-Decisions** says that you ought to decide to attend church. On this point, I think that **Options-as-**

**Decisions** has things right. If some action only looks good on the assumption that it will not be the result of a decision to perform that action, it seems mistaken to go on to say that you ought to perform that action. Proposals 1-3 sometimes have this unappealing result, while **Options-as-Decisions** never does.

# 9 Chisholm's Paradox

My account of options provides an attractive and well-motivated response to Chisholm's Paradox. Consider:

> **Professor Procrastinate**:[21] You have been invited to write an article for a volume on rational decision-making. The best thing would be for you to accept the invitation and then write the article. The volume is very prestigious, so having a paper published there will help you to get tenure and receive invitations to conferences. But you believe that if you accept the invitation, you'll wind up procrastinating and never getting around to writing the article. This would be very bad - the editors would be frustrated and you would often be passed over for future volumes. It would be better to reject the invitation than to accept it but not write.

Ought you accept the invitation or decline? It would seem that you ought to decline. After all, you believe that if you were to accept the invitation, you would wind up with the worst possible outcome - accepting the invitation and not writing, which will result in angry editors and decreased opportunities in the future.

But now there is a puzzle, for it also seems that you ought to accept the invitation and write the article. After all, you correctly believe that you are able to accept the invitation and then write the article. And if you were to accept and write, you would wind up with

[21] This example is a slightly modified version of a case presented in Jackson and Pargetter (1986).

the best possible outcome - far better than what would happen if you were to accept and not write, or if you were to decline and not write.

We have concluded both that you ought not accept and that you ought to accept and write. This is puzzling. To begin with, standard deontic logic (deriving from the work of von Wright (1951)) has the result that $Ought(A \wedge B)$ entails $Ought(A) \wedge Ought(B)$ and that $Ought(A) \wedge Ought(\neg A)$ entails a contradiction. Therefore, if we accept this description of the case, we must give up standard deontic logic (since $Ought(Accept \wedge Write) \wedge Ought(\neg Accept)$ would entail a contradiction). This is the conclusion Chisholm (1963) draws from this sort of case. Much effort must then be expended in trying to modify standard deontic logic so as to be compatible with this description of your obligations in this case.[22]

Worse, if given your beliefs, it is the case both that you ought to accept and write and that you ought to decline, then you cannot do everything that you ought to do, since obviously you cannot both accept and write and also decline.[23] This is problematic for three reasons. First, it means that our theory is giving you conflicting guidance. Second, it means that you cannot avoid being subject to rational criticism, no matter what you do. Third, it throws a wrench into our use of the subjective *ought* to predict agents' behavior. In this case, we obviously would not want to predict both that you will accept

[22] Chisholm puts his case in terms of conditionals, but I prefer for the sake of simplicity to put express it using conjunctions. See footnote 24, below, for the version of the paradox based on conditionals, along with the dissolution of that version of the paradox based on **Options-as-Decisions**.

[23] Jackson and Pargetter (1986) argue that you can in fact do all that you ought to do in this case. But they are considering what you *objectively* ought to do. In considering what you objectively ought to do, whether you ought to accept the invitation depends not on what you believe you will later do, but on what you will in fact later do. Then, in a case where if you accept the invitation, you in fact won't write the review, it appears both that you ought to accept and write, and that you ought to decline. But in this case, it is still possible for you to fulfill all of your requirements, since if you were to accept and write, it would no longer be the case that you ought to have declined. It is only given the truth of the conditional *if you accept, then you won't write* that you ought to decline. But you are able to affect the truth value of that conditional.

But when we are considering the subjective *ought*, things are different. Whether you ought to decline depends not on the actual truth value of the conditional *if you accept, then you won't write*, but on whether you believe that that conditional is true. And while your actions can affect the truth of that conditional, they cannot affect whether you presently believe that conditional to be true.

and write and that you will decline.

**Options-as-Decisions** yields an attractive dissolution and diagnosis of this paradox. In arriving at this problem, we noted that the best course of action for you in **Professor Procrastinate** is to accept the invitation and write the paper. Because you are also capable of accepting the invitation and then writing the paper (and believe you are capable of so doing), we inferred that you ought to accept the invitation and write the paper.

But if **Options-as-Decisions** is correct, the mistake in this reasoning was to suppose that because you have the ability to perform some action (and believe you have this ability), that action constitutes an option for you and so is potentially something that you ought to do. But we have already seen that the view that everything you are able to do (or believe you are able to do) is an option for you is untenable.

According to **Options-as-Decisions**, your options are all and only the decisions presently open to you - things like (i) *deciding to accept the invitation but not write the paper*, (ii) *deciding to accept the invitation (leaving open the issue of whether to write the paper)*, and (iii) *deciding not to accept the invitation.* (According to the Bratman's account of abilities to make decisions, you cannot make the decision to accept and then write the paper, since you believe that if you were so to decide, you would not carry out the decision. Therefore, this decision is not an option for you.) You are confronted with just this one, unique set of options. And given your belief that if you make a decision that involves accepting the invitation, you won't write the paper, the option with the highest expected utility is *deciding not to accept the invitation.* So you ought to decide not to accept the invitation, *and that's all*! There is no sense in which you ought to decide to accept the invitation and write the paper (if you are even able to make this decisions). (Note that even if making the decision to accept and then write counted as an option for you, it would have low expected utility and hence not be the decision that you ought to make.)

Chisholm's Paradox arises for Proposals 1-3 precisely because they allow for different ways of chopping up the space of options available to an agent. In **Professor Procrastinate**, they allow us to characterize your options in a coarse-grained way, as consisting of the two options of (i) accepting and (ii) declining, or in a more fine-grained way, as consisting of the three options of (i) accepting and writing, (ii) accepting and not writing, and (iii) declining (and not writing). And so it could be true both that declining has highest expected utility when we characterize your options in a coarse-grained way, and that accepting and writing has highest expected utility when we characterize your options in a more fine-grained way, yielding the result that you both ought to decline and also ought to accept and write. And in general, when a theory of options allows for multiple ways of characterizing your options, there is a possibility of winding up with conflicting *ought*s, since the option with highest expected utility on one way of chopping up your options might be incompatible with the option with highest expected utility on another way of chopping up your options.

**Options-as-Decisions** resolves Chisholm's Paradox by yielding a unique set of options for each agent. There is one set of options for each agent S and time t, namely the set of propositions of the form *S (only) decides at t to* $\phi$ that you are able to bring about. As a result, there is no way to wind up with conflicting *ought* claims (and hence no way to wind up with a case that poses a counterexample to standard deontic logic). There is only one set of options, and so whichever proposition has highest expected utility in that one set will be the one that you ought to bring about, period.[24]

[24] Chisholm originally presented the paradox using conditionals. The following statements are supposed to all be true descriptions of the case, but they are jointly incompatible with standard deontic logic: (i) You ought to write the paper; (ii) It ought to be that if you write the paper, you accept the invitation; (iii) If you believe you won't write the paper, you ought not accept the invitation; and (iv) You believe you won't write the paper. From (i) and (ii) it follows, by standard deontic logic, that you ought to accept the invitation, while from (iii) and (iv) it follows, by modus ponens, that you ought not accept the invitation. But on standard deontic logic, it cannot be the case both that you ought to accept and that you ought not accept.

But while these statements all *sound* compelling, **Options-as-Decisions** entails that (i) and (ii) are simply false. (i) is false because writing the paper is not an option for you, and (ii) is false because making this conditional true is not an option for you. Chisholm's Paradox shows that an intuitive description

This is a desirable result. But still, you might worry that my account is too lenient. In **Professor Procrastinate**, it says that you ought to decide not to accept the invitation, as a result of your believing yourself to be weak-willed. Does **Options-as-Decisions** therefore have the problematic implication that you are excused from any criticism if you fail to achieve the best possible outcome by accepting and then writing the article?

No. My account allows that you may be still subject to rational criticism if you decide not to accept the invitation, but this criticism will not stem from your having failed to do what you ought to have done. Such criticism can arise in a variety of ways.

First, you might be in fact weak-willed (so that if you were to accept the invitation, you would procrastinate) and take this fact into account in deciding not to accept the invitation. Then, if weakness of will is a failure of rationality, then you are irrational not for having *done* something you rationally ought not have done, but instead for lacking a certain ability, namely the ability to control your future selves through your present decisions.

Second, even if you are not now weak-willed, you might have been weak-willed in the past, and this past weakness of will may have given you the justified but false belief that you are now weak-willed. You responded appropriately to this justified belief in deciding not to accept the invitation, and so your present self is in no way irrational. But if weakness of will is a failure of rationality, you are still subject to rational criticism, although this criticism is directed at your past self rather than your present self.

Third, you might have lacked good evidence for your belief that if you were to accept, you would procrastinate. If this is so, you were not *practically* irrational in deciding not to accept. Rather, you were *epistemically* irrational, since the belief that recommended

of this case (expressed in the statements (i)-(iv)), including a description of your obligations therein, is incompatible with standard deontic logic, given a standard interpretation of the conditionals in (ii) and (iii). One response is to modify standard deontic logic. Another response is to try to reinterpret the conditionals in (ii) and (iii) to avoid incompatibility with standard deontic logic. A third response, which falls out of **Options-as-Decisions**, is to simply deny the description of your obligations in this case. If Chisholm's description of your obligations is incorrect, then the paradox dissolves even without any modification of standard deontic logic or non-standard interpretation of the conditionals.

deciding not to accept was not responsive to your evidence.

However, if you had good but misleading evidence that you are weak-willed, and this evidence was not the result of any past irrationality on your part, then you are in no way subject to rational criticism if you decide not to accept the invitation. Your past self was perfectly rational, you formed beliefs that were supported by your evidence, and you responded appropriately to these rational beliefs by deciding to decline the invitation. So, sometimes you take into account predicted weakness of will on your part and therefore wind up with a sub-optimal outcome (among those that you could have achieved if you had performed certain sequences of actions) without being in any way irrational.

## 10 Conclusion

The project of coming up with a theory of what an agent's options are is part of a broader project in the theory of rationality. Formal Bayesian tools have proven incredibly fruitful for theorizing about epistemic and practical rationality. On the epistemic side, Bayesian models have illuminated questions about how agents ought to modify their beliefs (or degrees of belief) in response to various sorts of evidence. And on the practical side, Bayesian expected utility theory has shed light on how agents ought to behave, given their uncertainty about the world.

But these formal models embody serious idealizations and are most at home in somewhat artificial casino cases. Suppose you are at the roulette table deliberating about how to bet. In this situation, all of the information needed to employ expected utility theory is right in front of you. Expected utility theory requires three things as inputs in order to yield a recommendation about how to act: probabilities, utilities, and a set of options. At the roulette table, all of these three elements are straightforwardly determined by the setup of the case. The probabilities are precise and given by the physical symmetries of the roulette wheel. The utilities can be thought of as the monetary gains or losses that would be incurred in different outcomes of the spin. And the options are all of the different bets you might place, given the rules of roulette. Expected utility theory is so easy to apply in the casino precisely because it is so clear what the relevant probabilities, utilities, and options are.

But out in the wilds of everyday life, things are not so neat and tidy. Suppose that instead of standing at the roulette table deliberating about to bet, you are standing at the raging creek deliberating about how to act in light of the fast-flowing current. Instead of precise probabilities determined by the physics of the roulette wheel, you might have only hazy degrees of belief about which outcomes would result from various actions you might take. And instead of precise utilities linked to the possible monetary payoffs determined by the casino's rules, you might only have rough preferences between possible outcomes (fording the creek safely is better than heading home, which in turn is much, much better than being swept away into the rapids downstream) without being able to assign precise numbers representing their desirabilities. And finally, instead of the options being specified by the rules of roulette, it is *prima facie* less clear which things you should be evaluating as options.

In order for expected utility theory to be relevant to your situation, each of these three disanalogies between the creek case and the roulette case needs to be addressed. How should the formal framework be employed (or extended) in cases where you lack precise probabilities, precise utilities, and a set of options stipulated by the rules of a game? The first of these problems - employing the formal machinery in the absence of precise probabilities - has received considerable attention since the 1970s, and much progress has been made.[25] The second - employing the machinery in the absence of a precise utility function - has only recently be addressed in any detail.[26] But the third - what should count as your options in a case where this isn't specified by something like the rules of

[25] See especially Levi (1974), Joyce (2005), White (2009), and Elga (2010) for discussion.

[26] See Hare (2010) for compelling discussion of this issue.

the casino - has hardly been dealt with at all.[27]

My aim in this paper has been to confront head-on the problem of coming up with a theory of an agent's options. This task proved surprisingly tricky. On the one hand, your options must be things that you believe you can do and must supervene on your mental states, in order for the subjective *ought* to be appropriately sensitive to your uncertainty about the world. On the other hand, your options must also be things that you can actually do. In a sense, these desiderata require your options to consist of a kind of action such that your actual abilities to perform actions of that kind always match up with your beliefs about your abilities to perform actions of that kind. I argued that only decisions can plausibly play this role. Hence **Options-as-Decisions**, according to which your options consist of all and only the decisions you are presently able to make. Focusing on an agent's decisions, as opposed to the non-mental actions she might be able to perform, impacts on our first-order normative theorizing in a variety of ways. In this paper, I showed how it directly takes into account the costs and benefits of decisions themselves and also leads to a principled and attractive response to Chisholm's Paradox.[28]

# References

Anscombe, G.E.M. *Intention.* Oxford University Press, 1957.

Bermudez, Jose Luiz. *Decision Theory and Rationality.* Oxford University Press, 2009.

Bratman, Michael. *Intentions, Plans, and Practical Reason.* CSLI, 1987.

[27] Aside from the aforementioned brief discussions in Jeffrey (1965) and Lewis (1981), this issue is discussed in Jackson and Pargetter (1986), Joyce (1999), Pollock (2002), and Smith (2010).

[28] I would like to thank Dan Greco, Caspar Hare, Richard Holton, Heather Logue, Tyler Paytas, Agustín Rayo, Miriam Schoenfield, Paulina Sliwa, Matthew Noah Smith, Robert Stalnaker, Roger White, and Steve Yablo, as well as audiences at the 2011 MITing of the Minds Conference, the 2011 Bellingham Summer Philosophy, and the 2011 Rocky Mountain Ethics Congress, for very helpful comments.

Buchak, Lara. "Review of José Luis Bermúdez, _Decision Theory and Rationality_." *Notre Dame Philosophical Reviews* 2009 (2009).

Chisholm, Roderick. "Contrary-to-Duty Imperatives and Deontic Logic." *Analysis* 24 (1963): 33–36.

Elga, Adam. "Subjective Probabilities Should be Sharp." *Philosopher's Imprint* 10 (2010).

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66 (1969): 829–839.

Hare, Caspar. "Take the Sugar." *Analysis* 70 (2010): 237–247.

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

Jeffrey, Richard. *The Logic of Decision.* University of Chicago Press, 1965.

Jeffrey, Richard. "Preference among Preferences." *Probability and the Art of Judgment.* . Cambridge University Press, 1992.

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

Joyce, James. "How Probabilities Reflect Evidence." *Philosophical Perspectives* 19 (2005): 153–178.

Kavka, Gregory. "The Toxin Puzzle." *Analysis* 43 (1983): 33–36.

Levi, Isaac. "On Indeterminate Probabilities." *Journal of Philosophy* 71 (1974): 391–418.

Lewis, David. "Causal Decision Theory." *Australasian Journal of Philosophy* 59 (1981): 5–30.

Pollock, John. “Rational Choice and Action Omnipotence.” *Philosophical Review* 111 (2002): 1–23.

Smith, Matthew Noah. “Practical Imagination and its Limits.” *Philosophers' Imprint* 10 (2010).

Wright, G.H.von . “Deontic Logic.” *Mind* 60 (1951): 1–15.

White, Roger. “Evidential Symmetry and Mushy Credence.” *Oxford Studies in Epistemology, vol 3.* . Oxford University Press, 2009.