Choosing Rationally and Choosing Correctly*

Ralph Wedgwood

1 Two views of practical reason

Suppose that you are faced with several different options (that is, several ways in which you might act in a given situation). Which option should you choose?

Let us take an example that Bernard Williams (1981: 102) made famous. Suppose that you want a gin and tonic, and you believe that the stuff in front of you is gin. In fact, however, the stuff is not gin but petrol. So if you drink the stuff (even mixed with tonic), it will be decidedly unpleasant, to say the least. Should you choose to drink the stuff or not?

It seems to me that there are at least two ways of interpreting this question. If we interpret the question in one way, “what you should choose” depends on what the available options are really like (not just on what you believe about what these options are like). For example, it may depend on the actual causal consequences of those options, or on other external facts that are quite independent of your state of mind. In this case, the option of drinking the stuff involves drinking petrol, and so giving yourself a decidedly unpleasant experience, while many of the other available options have no comparable drawbacks. So, when the question is interpreted in this way, you “shouldn’t choose” to drink the stuff. We could call this an “external” or “objective” ‘should’. I shall express this “external” ‘should’ by saying that in this case, choosing to mix the stuff with tonic and drink it is an incorrect choice for you to make. As we might say, in choosing to mix the stuff with tonic and drink it, you have got things wrong; your choice was a
If we interpret the question in another way, however, “what you should choose” depends only on your overall state of mind (not on external facts that could vary while your state of mind remained unchanged). Even though you mistakenly believe that the stuff in front of you is gin and not petrol, it could still be that there is an impeccable process of reasoning that leads from your current state of mind to your choosing to mix the stuff with tonic and drink it. In that case, the choice to mix the stuff with tonic and drink it fits perfectly with your current overall state of mind. So, in this case, when the question is interpreted in this way, it would be wrong to say that you “shouldn’t choose” to drink the stuff. We could call this an “internal” or “subjective” ‘should’. I shall express this “internal” ‘should’ by saying that in this case, choosing to mix the stuff with tonic and drink it is a perfectly rational choice for you to make.

The central topic of this essay is the relationship between these two kinds of ‘should’, the “internal” and the “external” ‘should’—or, in other words, between choosing “rationally” and choosing “correctly”. It would be odd if these two kinds of ‘should’ were completely independent of each other; it seems more likely that the truths involving one of these two kinds of ‘should’ are in some way explained by truths involving the other. But in which direction goes the order of explanation go? Are the truths about which choices are (internally or subjectively) rational ultimately explained by more fundamental truths about which choices are (externally or objectively) correct? Or does the order of explanation go in the other direction?

This issue marks a crucial disagreement between two views of practical reason. In effect, it is the issue that underlies the disagreement between what Garrett Cullity and Berys Gaut (1997: 1–6) have called the recognitional and constructivist views of practical reason.
As I shall understand them, “recognitional” views of practical reason take as fundamental some principle about when a choice is (externally or objectively) correct. For example, such recognitional views might take as fundamental the principle that a choice is correct if and only if the option chosen really is in a certain way a good thing to do—where whether or not an option is a good thing to do in this way may depend, at least in part, on external facts, such as the actual causal consequences of the available options, and the like.

These recognitional views rely on this fundamental principle in giving an account of what it is for a choice to count as (internally or subjectively) rational. Proponents of these views cannot say that a choice is rational just in case the option chosen really is a good thing to do. Whether or not a choice is rational is, I am assuming, an “internal” matter, determined by the agent’s overall state of mind alone, whereas whether or not an option is a good thing to do is an “external” matter, which may depend, at least in part, on external facts. Instead, proponents of these recognitional views could say something like this: a choice is rational just in case the agent believes that the option chosen is (in the relevant way) a good thing to do. But this would not be a very plausible thing to say: if the agent’s belief that the option chosen is a good thing to do is a grossly irrational belief, then surely the choice will be equally irrational. So it would be more plausible to say this: a choice is rational just in case it is rational for the agent to believe that the option chosen is (in the relevant way) a good thing to do. As many epistemologists agree, whether or not it is rational for the agent to hold a certain belief is also an “internal” matter in the relevant sense, determined by the believer’s overall state of mind, and not by facts that could vary while that state of mind remained unchanged. According to this conception of rational belief, what makes it rational for one to hold a certain belief are internal facts about one’s experiences,
memories, intuitions, background beliefs, and so on.\textsuperscript{4}

The first task for any version of the recognitional view, then, is to identify a certain concept that represents the property of being in the relevant way a “good thing to do”. Then this view will give an account of what it is for a choice to be rational in terms of the rationality of holding certain beliefs that involve this concept. For example, according to the version of the recognitional view under consideration, whenever it is rational for one to choose a certain option, what makes it rational for one to choose this option is the fact that it is rational for one to hold a certain belief involving this concept—namely, the belief that the option is in the relevant way a “good thing to do”.

Intuitively, if the fundamental principle applying to choices is that one “should” (in the external or objective sense of ‘should’) choose options that really are good things to do, then this seems to explain why there is also a subsidiary principle applying to choices, to the effect that one “should” (in the internal or subjective sense of ‘should’) choose options that it is rational for one to believe to be good things to do. But what exactly is the nature of the explanatory connection between these two principles? It may be that the explanatory connection is this.\textsuperscript{5} The fact that an option is a good thing for one to do is an external fact about that option; it is not determined by facts about one’s state of mind alone. So one cannot directly comply with the requirement that one should choose options that really are good things to do. One can only comply with this requirement indirectly, by means of complying with an internal requirement that one should adjust one’s choices to some internal fact about one’s mental states. The best internal requirement of this sort to comply with, in order to achieve the external result of choosing options that really are good things to do, is the requirement that one should choose
options that it is rational for one to believe to be good things to do. In this way, then, a
recognitional view may not only give an account of the feature that makes rational choices
rational; it may also help to explain why one “should” make choices that have that feature, rather
than choices that lack it.

A recognitional view of this sort can also give an explanation of why akrasia is irrational.
Let us assume that akrasia involves choosing to do something that one believes not to be a good
thing to do. Now, one’s belief that a certain option is not a good thing to do is either a rational
belief for one to hold or it is not. If it is not a rational belief for one to hold, then one is being
irrational in one’s beliefs. If, on the other hand, it is a rational belief for one to hold, then it
cannot simultaneously be rational for one to believe that the option is a good thing to do. So, one
is choosing the option even though it is not rational for one to believe the option to be a good
thing to do; hence—according to the recognitional view of rational choice—one’s choice is
irrational. So akrasia necessarily involves irrationality, either in one’s beliefs or one’s choices.

We might express the recognitional view, metaphorically, by saying that rational practical
reasoning “aims” at choosing options that really are good things to do. In this way, the
recognitional view of rational choice parallels a certain claim that is often made about belief.
According to this claim about belief, if one forms and revises one’s beliefs rationally, one’s
reasoning “aims” at believing the truth and nothing but the truth about the question at issue. In
this sense, the recognitional view holds that rational choice “aims at” options that are good things
do, just as rational belief “aims at” the truth.

I have only given a crude sketch of the recognitional view of practical reason here. The
recognitional view might be refined in many ways. For example, the recognitional view could be
generalized so that it applies not just to choices (which are mental events involving the formation of a new intention), but to all kinds of intention revision (including mental events in which one reaffirms or abandons an old intention). It could also be generalized so that it even applies to one’s failing to revise one’s intentions on a certain occasion, since intuitively failing to revise one’s intentions can also sometimes be a serious mistake.

This view might also be revised to give a more refined account of when exactly a choice is rational. According to the simple account that we are currently considering, a choice is rational if and only if it is rational for the agent, at the time of choice, to believe the chosen option to be a good thing to do. But suppose that one has to make a choice in an emergency, in which one does not have enough time or information for it to be rational for one to hold an outright belief about whether any of the available options is a good thing to do. In cases of this sort, one’s choice can surely still be rational, even though it is not rational for one to believe the chosen option to be a good thing to do. This suggests that the account must be refined so that it requires one to adjust one’s choice to the evidence that one has in favour of beliefs about whether or not the available options are good things to do, even if that evidence is not good enough to make it rational for one to hold an outright belief to the effect that the chosen option is a good thing to do.

I shall return to the task of refining the recognitional view of practical reason in the final section of this essay. In most of this essay, however, I shall ignore all these refinements to the recognitional view. To simplify the discussion, I shall just focus on the simple version, according to which a choice is correct if and only if the chosen option is a good thing to do, and a choice is rational if and only if it is rational for the agent to believe the chosen option to be a good thing to do. All the arguments that I shall make in the next four sections could be adapted to apply to the
more refined versions just as much as to this simple version.  

This simple version makes it particularly clear why it is appropriate to call this a “recognitional” view of practical reason. According to this view, there are simply truths about which of the available options are good things to do and which are not; and the central or canonical method of practical reasoning is just to attempt to recognize, or to form rational beliefs about, these truths, and then to make one’s choice accordingly. Something like this conception of rational practical reasoning is suggested by Aristotle’s claim that practical wisdom involves both the “practical intellect”—that is, sound reasoning about action based on a true understanding of the human good—and choice in accordance with what the practical intellect asserts.  

Proponents of the constructivist view of practical reason, on the other hand, take as fundamental certain internal requirements that a choice must meet in order to be rational. They deny that these internal requirements of rational choice are explained by any principle about when a choice is (externally) “correct”, or when an option is (externally) a “good thing to do”. Thus, the constructivists deny that these internal requirements of rationality are explained by the good external results to which complying with these requirements either will actually lead, or may reasonably be expected to lead. According to the constructivists, these internal requirements either require no explanation at all, or else are explained in some other way. For this reason, according to the constructivists, there is no external concept of an option’s being “good” in some way such that it is a basic requirement of rationality that one must choose options that it is rational for one to believe to be good in that way. (If this were a basic requirement of rationality, it would be all but irresistible to conclude that the reason why one
should choose actions that it is rational for one to believe to be good in that way is because of a more fundamental principle that one should choose actions that really are good in that way.) Instead, constructivists typically propose that the requirements of practical rationality are either purely procedural requirements, or else pure requirements of formal coherence among one’s choices or preferences.

There are two main versions of constructivism that are defended by contemporary philosophers. One version is decision-theoretic constructivism, according to which one’s preferences are rational if and only if they satisfy certain conditions of coherence or consistency. Typically, the idea is that these preferences must satisfy the “axioms” (transitivity, monotonicity, independence, and so on) that are necessary to make it possible to represent those preferences by means of a “utility function” (Joyce 1999: 84–89). For an agent whose preferences are coherent in this way, it is rational to choose an option if and only if no alternative option is preferred.11

The other main version of constructivism is Kantian constructivism, according to which the fundamental principle of rational choice does not require that we choose options that we believe to be good in some way. Instead, this fundamental principle of rationality requires that in making choices, one should follow a procedure that meets certain formal conditions of consistency and universalizability. Specifically, according to Kant (1788: 62–63), to be rational, one must always make one’s choice through following some “maxim” or general rule, which it must also be consistent for one at the same time to will to be a universal law.12

In these constructivist theories, then, these internal requirements of rational choice are fundamental. Constructivists have made many claims about these requirements: some constructivists have compared them to the requirements of logical consistency among our
beliefs;¹³ and some constructivists have claimed that these requirements of rational choice are *a priori.*¹⁴ But they all deny that the external notions of a choice’s being “correct” or of an option’s being a “good thing to do” play any role in the explanation of the requirements of rational choice. The external notions are either denied to have any necessary connection to the notions of rational practical reasoning or of reasons for action, or else they are simply *defined* in terms of what it is “rational to choose”. For example, according to some of these philosophers, for an option to be a good thing for one to do just is for it to be an option that it *would* be rational for one to choose, if one were ideally well-informed about the relevant facts.¹⁵ According to this definition, all truths about which choices are correct, and about which options are good things to do, are “constructed” out of the internal requirements of rational choice.

Paraphrasing Christine Korsgaard (1996a: 36–37), we can express the difference between constructivism and the recognitional view as follows. Constructivists believe that there are truths about what is a good thing for one to do *because* there are rational procedures for making choices—whereas according the recognitional view, there are rational procedures for making choices *because* there are truths about what is a good thing for one to do, which it is rational for one to expect those procedures to track. For the constructivists, these rational procedures—or more generally, the internal requirements of rational choice—are fundamental. Everything else that has any necessary connection to reasons for action must be explained on the basis of these internal requirements of rationality. This is why both the Kantian constructivists, such as Korsgaard (1996a), and the decision-theoretic constructivists, such as David Gauthier (1986), seek to construct the whole of ethics on the basis of their account of these internal requirements of rational choice.
In this essay, I shall consider some of the objections that the constructivists have directed against the recognitional view. Ironically, as I shall argue, these objections apply just as much to constructivism as to certain versions of the recognitional view. Then I shall argue that there is a version of the recognitional view that is immune to constructivist objections. This, it seems to me, provides considerable support for this version of the recognitional view.

2 “Formal” and “substantive” versions of the recognitional view

As I explained in the previous section, the simple version of the recognitional view that I am focusing on here first identifies a certain concept that represents the property of being an option that is in a certain way a “good thing to do”, and then claims that a choice is rational if and only if it is rational for the agent to hold a certain belief involving this concept—namely, the belief that the chosen option is in this way a “good thing to do”.

But which way of being a good thing to do is “the relevant way”? As Judith Thomson (2001: 17–19) and others have pointed out, there are many different ways in which something can be good: it may be good for me, or good for Oxford University, or morally good, and so on. Perhaps there can also be more than one concept that represents the very same property (such as the concepts ‘good for Cicero’ and ‘good for Tully’, perhaps). If so, then even after the recognitional view has identified the relevant way of being good, it must still identify the relevant concept that represents the property of being in this way a good thing to do. So, exactly which, out of all the concepts that can be expressed by the term ‘a good thing to do’, is the one that the recognitional view is employing here?
According to David Velleman (2000: 176–77), there are two main types of concept that proponents of the recognitional view can employ here. They could employ a concept that gives a purely “formal” specification of the “object of practical reasoning”; or they could employ a concept that gives a “substantive” specification of the “object of practical reasoning”.

As Velleman explains, a concept gives a “formal” specification of “the object of an enterprise” if it is simply the concept of the object of that enterprise. For example, the concept ‘winning’ gives a formal specification of “the object of a competitive game”, since the concept “‘winning’ just is the concept of succeeding in competition” (176). Similarly, Velleman suggests, one concept that could be expressed by the term ‘a good thing to do’ is simply the concept of the “object” of practical reasoning.

Velleman believes that practical reasoning literally has an “object” or “aim”. I shall not assume this here. Instead, I shall suppose that a purely “formal” concept of a “good thing to do” is a concept such that it is a conceptual truth that an option is a “good thing to do” in this sense if and only if it is an option that it is correct to choose—in precisely the same sense of the term ‘correct’ that I explained in the previous section. One such concept, which can be expressed by describing an option as a “good thing to do”, is as I have argued elsewhere (Wedgwood 2001) a concept the content of which is determined by the special “conceptual role” that it plays in rational practical reasoning. Specifically, if a rational agent makes judgments using this concept, about which of the available options fall under this concept and which do not, then she will choose one of the options that she judges to fall under the concept, and not one of the options that she judges not to fall under it. I have argued that if the content of a concept is determined by its having a conceptual role of this kind, then it follows that an option falls under the concept if and
only if it is correct to choose it.

To judge that an option is a “good thing to do”, in this formal sense, then, is not to make a specific value-judgment, such as that the option is *morally* good, or good *for the agent*, or good *as a means to a certain end*. It is simply to judge that it is an option that it is correct to choose. This judgment could also be expressed in many other ways. For example, this judgment could also be expressed by saying that the option is “choiceworthy” or “OK”. The judgment that an option is in this formal sense *not* a good thing to do could be expressed by saying that the option is something that one had “better not” do, or that there is a “conclusive reason” for one not to do, or simply that one “should not” do.

One version of the recognitional view, then, would use the term ‘a good thing to do’ to express this purely formal concept. I shall call this the “formal” version of the recognitional view. The formal version of the recognitional view is distinct from the constructivist view, because whether or not something counts as a “good thing to do”, in this formal sense, is typically not determined by internal facts about the agent’s overall state of mind, but at least in part by external facts, which could vary while the agent’s overall state of mind remained unchanged.

Alternatively, the recognitional view might use the term ‘a good thing to do’ to express a more *substantive* concept of an option’s being good in some specific way. I shall call views of this sort “substantive” versions of the recognitional view. For example, one such substantive version of the recognitional view would use the term ‘a good thing to do’ to mean *optimal for the agent’s happiness*. This version of the recognitional view would in effect be what Derek Parfit (1984: 1–2) has called the “Self-interest Theory of rationality”. According to this theory, a choice
is correct if and only if the chosen option maximizes the agent’s happiness (that is, there is no alternative to the chosen option that will make a greater contribution to the agent’s happiness); and a choice is rational if and only if it is rational for the agent to believe that the chosen option maximizes her happiness.

Another “substantive” version of the recognitional view would use the term ‘a good thing to do’ to mean optimal for satisfying the totality of the agent’s present desires. This would in effect be what Parfit (1984: 92–94) has called the “the Instrumental version of the Present-aim Theory”. According to this theory, a choice is correct if and only if the chosen option optimally satisfies the totality of the agent’s present desires; and a choice is rational if and only if it is rational for the agent to believe that the chosen option optimally satisfies the totality of his present desires.\(^{17}\)

### 3 A problem for “substantive” versions of the recognitional view

Let us start by considering “substantive” versions of the recognitional view. For example, consider the egoistic version of the recognitional view (Parfit’s “Self-interest Theory”). This view interprets the notion of a “good thing to do” as optimal for the agent’s own happiness (where the term ‘happiness’ expresses a substantive concept of some sort—not simply the purely formal concept of a life of the sort that it is correct to choose). According to this view, if one makes a choice, one’s choice is correct if and only if the chosen option maximizes one’s happiness; and one’s choice is rational if and only if it is rational for one to believe that the chosen option maximizes one’s happiness.
As I shall argue here, this egoistic version of the recognitional view faces a problem that Christine Korsgaard (1996a: 9–21) has called “the normative question”. According to this egoistic view, it is irrational—indeed akratic—for one to choose any option if one rationally believes that that option will not maximize one’s own happiness. But we can imagine an agent (let us call her Alice) who knows that a certain option will not maximize her happiness, but is still uncertain about whether or not to choose that option—perhaps because the option has some other feature that she is tempted to regard as highly important. Alice need not doubt that the fact that the option will not maximize her happiness is some reason for her not to choose it. But she may still be uncertain whether to treat this fact as an overriding or decisive reason, as the egoistic view requires her to do. She might express her perplexity by asking, “Why should I always choose options that maximize my own happiness? Why shouldn’t I sometimes choose options that won’t maximize my happiness instead?”

In asking this “normative question”, Alice seems to be seeking some compelling further reason not to choose any option that does not maximize her happiness. That is, she is looking for some consideration that could rationally persuade her not to choose any option that she believes not to maximize her happiness—even if she did not yet have the disposition to avoid choosing any options that she rationally believes not to maximize her happiness. But according to the egoistic version of the recognitional view, there are no such further reasons. The only way in which any considerations can rationally persuade one not to choose any option, according to this egoistic view, is by making it irrational for one to believe that the option maximizes one’s own happiness. So according to this egoistic view, someone who lacked the disposition not to choose any option that it was not rational for her to believe to maximize her happiness could not be
rationally persuaded not to choose anything at all; such an agent would be beyond the power of rational persuasion altogether.

So, it seems, the egoistic version of the recognitional view must accept that if any agents lack the disposition to comply with this alleged rational requirement—the requirement that one should not choose any option that it is not rational for one to believe to maximize one’s own happiness—there is no way of rationally persuading them to do so. In that sense, the egoistic view must regard this rational requirement as basic. If one is not already disposed to conform to this requirement, there is no way of rationally persuading one to do so, since rational persuasion precisely consists in exploiting a person’s disposition to comply with this requirement.

There do seem to be some “basic” rational requirements of this sort. For example, if someone is not already disposed to accept instances of the basic laws of logic, there will be no way of rationally persuading them to do so, since all arguments that one might employ in order rationally to persuade them will themselves involve instances of those very laws of logic. So why shouldn’t the egoistic requirement also be a “basic” rational requirement of this sort?

However, it does not seem plausible that this egoistic requirement is a basic requirement of this sort. If you violate a rational requirement—that is, if you make an irrational choice or form an irrational belief—this reflects a cognitive defect in you. (This is a fundamental difference between the internal notion of a rational choice and the external notion of a correct choice. There need be no defect in you at all if you make an incorrect choice, by choosing an option that is not in fact a good thing to do; it may be sheer bad luck that the stuff was petrol and not gin, so that your choice to mix the stuff with tonic and drink it was in fact an incorrect choice.) Since irrationality is always a defect in you, then, so long as you are being sufficiently sane and
intelligent, you will tend to avoid such defects. At the very least, if it is a \textit{basic} requirement of rationality that one should not form a set of beliefs or choices that has a certain feature, then if you \textit{recognize} that a certain set of beliefs or choices has that feature, you will not remain uncertain about whether or not to form that set of beliefs and choices, unless you are being less than perfectly sane and intelligent. For example, suppose that it is a basic principle of rationality that it is irrational to believe any proposition that is logically self-contradictory. Then, if you recognize that a certain proposition is logically self-contradictory, you will not remain uncertain about whether or not to believe that proposition, unless you are being less than perfectly sane and intelligent.

Intuitively, however, it seems quite possible that Alice is being perfectly sane and intelligent, even if she recognizes that it is not rational for her to believe that the option in question will maximize her happiness, but still remains uncertain about whether to choose that option. Alice’s uncertainty about whether or not to choose this option hardly seems in the same category as a failure to be convinced by instances of the elementary laws of logic. Indeed, her perplexity seems eminently intelligible. So, it seems implausible to claim that it is a basic requirement of rationality that one should never choose any option that it is not rational for one to believe to maximize one’s own happiness. But as we have seen, the egoistic version of the recognitional view entails that it is a basic requirement. So this seems to be a serious problem for the egoistic version of the recognitional view.

The same problem also arises for other substantive versions of the recognitional view. For example, according to the \textit{instrumentalist} version of the recognitional view, a choice is rational if and only if it is rational for one to believe that the chosen option optimally satisfies the totality of
one’s present desires. So, according to this view, it is irrational (indeed *akratic*) for one to choose an option if one rationally believes that the option does not optimally satisfy the totality of one’s present desires. But we can imagine an agent (call him George) who starts to regard the majority of his desires with suspicion; perhaps he becomes attracted to the ideal of detaching himself from all self-centred desires. George could be convinced that a certain option will not optimally satisfy the totality of his present desires, but still wonder whether or not to choose that option. He might ask himself, “Why should I always choose options that optimally satisfy the totality of my present desires? Why shouldn’t I sometimes choose options that don’t optimally satisfy those desires instead?”

In asking this question, George appears to be seeking a compelling *further reason* not to choose any option that does not optimally satisfy the totality of his present desires. That is, he is seeking some consideration that could rationally persuade him not to choose any such option—even if he is not yet disposed to reject all options that it is not rational for him to regard as optimally satisfying the totality of his desires. But according to the instrumentalist view, there are no such further reasons. The only considerations that can rationally persuade one not to choose any option are considerations that make it irrational for one to believe that the option will optimally satisfy the totality of one’s desires. So, unless George is already disposed not to choose any option that it is not rational for him to regard as optimally satisfying the totality of his desires, then he cannot be rationally persuaded not to choose anything.

Thus, the instrumentalist must claim that it is simply a *basic* requirement of rationality that one should not choose any option that it is not rational for one to regard as optimally satisfying one’s total set of desires. But it seems doubtful whether the instrumentalist
requirement really can be a basic requirement of rationality. George could surely be perfectly sane and intelligent, even if he recognizes that it is not rational for him to regard a certain option as optimally satisfying his total set of desires, but still remains uncertain about whether to choose that option.

In general, this problem will arise for all substantive versions of the recognitional view. Whatever substantive concept of a “good thing to do” the recognitional view takes as its central concept, it will have to take the requirement that one should not choose any option that it is not rational for one to believe to be, in this substantive sense, a “good thing to do” as a basic requirement of rationality. But it seems doubtful whether there can be any basic requirement of rationality of this sort. For every such requirement, it seems that agents could be perfectly sane and intelligent, even if they recognize that it is not rational for them to believe a certain option to be (in this substantive sense) a “good thing to do”, and yet still feel uncertain about whether to choose the option. None of these alleged requirements resembles the clear examples of basic requirements—such as the requirement that one should not believe logical contradictions. For every substantive concept of a “good thing to do”, one could always ask, without revealing any insanity or lack of intelligence, “But why should I also choose options that are good things to do in that way? Why shouldn’t I sometimes choose options that aren’t in that way good things to do at all?” So, the substantive versions of the recognitional view all seem to face a serious problem.
4 An analogous problem for constructivist views

As I shall argue, a closely analogous problem also arises for constructivist views of practical reason. As I mentioned in the first section, there are two main varieties of constructivism that have many proponents today: decision-theoretic constructivism and Kantian constructivism.

According to decision-theoretic constructivism, rational choices are choices that cohere in a certain way with each other and with one’s preferences and beliefs. Specifically, one’s preferences should be transitive, monotonic, independent, and so on; and it is rational to choose an option only if no alternative option is preferred. According to this view, one’s choices “should” (in the internal use of that term) satisfy these conditions of coherence. But now it seems that we can raise the “normative question” again. Why should our choices satisfy these conditions of coherence? Why does it matter whether or not our choices are coherent in this way?

According to Kantian constructivism, rational choices are choices that are made by means of a procedure that satisfies certain formal conditions of consistency and universality. Specifically, one must make one’s choice by following a general rule or “maxim”, which it must also at the same time be consistent for one to will to be a universal law. According to this view, one “should” (in the internal sense of that term) always make one’s choices by following a universalizable maxim of this kind. But the “normative question” arises here too. Why should we always make our choices by following such universalizable maxims? Why does it matter whether or not we make our choices in this way?

According to the constructivists, the fact that one always “should” make choices that meet these internal conditions of coherence or universalizability is not explained by the good
external results to which such choices either will actually lead, or may rationally be expected to lead. But this seems to imply that it matters simply *in itself*, purely *for its own sake*, whether or not one’s choices meet these internal conditions of coherence, or whether they are made by following a suitably universalizable maxim.\(^\text{18}\)

On the face of it, however, this is a rather surprising idea. Why on earth should such a thing matter purely for its own sake? Perhaps choices that do not meet these internal conditions of coherence or universalizability are aesthetically unattractive in some way: they form a less pretty mental pattern than choices that do meet these conditions. But this hardly seems a sufficiently weighty consideration to explain why (in the internal use of the term) one “should” *never* make choices that do not satisfy these internal conditions. It would defy belief to claim that it matters purely for its own sake whether or not one’s choices meet these internal conditions in the relevant way, but absolutely no explanation can be given of why it matters. So constructivists must surely offer some further explanation of why it matters—that is, of why one’s choices should meet these internal conditions of coherence or universalizability.\(^\text{19}\)

Many constructivists try to offer such an explanation by arguing that it is “constitutive” of having the capacity for choices at all that one’s choices must tend to satisfy these internal conditions of coherence. For example, some of the decision-theoretic constructivists, such as David Lewis (1974), argue that we would not even be interpretable as having preferences at all unless our choices tended, by and large, to satisfy these conditions of coherence.\(^\text{20}\)

But is this claim, that it is constitutive of having preferences at all that one’s choices must tend, by and large, to satisfy these conditions of internal coherence, really enough all by itself to explain why it is *always irrational* to make choices that do not satisfy these conditions (that is,
that one should never make choices that do not satisfy these conditions)?

Even if it is impossible to have preferences at all unless your choices tend, *by and large*, to satisfy these conditions, how can this explain why your choices should *always* satisfy these conditions? Perhaps it does not matter at all if you *sometimes* make choices that do not satisfy these conditions.

Anyway, it seems doubtful whether the claim that a disposition not to make such internally incoherent choices is constitutive of having preferences at all can be the *basic* explanation of why it is irrational to make such choices. If this claim were the basic explanation of why it is irrational to make such internally incoherent choices, then it would in effect be a basic principle of rationality that it is irrational for anyone to go against those dispositions that are constitutive of having preferences at all. But it is doubtful whether there can be any such basic principle of rationality.

As I argued in the previous section, if you make an irrational choice, this choice reflects a cognitive defect *in you*. Since irrationality is always a defect in you, then, so long as you are being sufficiently sane and intelligent, you will tend to avoid such defects. At the very least, if it is a *basic* principle of rationality that it is irrational for anyone to form a set of beliefs or choices that has a certain feature, then, if you recognize that a certain set of beliefs or choices has that feature, you will not remain uncertain about whether or not to form that set of beliefs and choices, unless you are being less than perfectly sane and intelligent. But it seems intuitively quite possible that a perfectly sane and intelligent agent might recognize that a certain choice would involve going against a disposition that is constitutive of having preferences at all, and still remain uncertain about whether or not to make that choice. Such a person might ask herself,
“I recognize that a disposition not to make choices of this kind is constitutive of having preferences at all, but why shouldn’t I sometimes resist that disposition, and make a choice of this kind anyway?” So it seems doubtful whether the internal conditions of rational choice can be explained in this way.

The same problem also seems to arise for the Kantians’ explanation of why we should never make any choices except by following their supreme principle of practical reason (see Kant 1785: 446–47, and Korsgaard 1996a: 92–100). This Kantian explanation is based ultimately on the proposition that the will is free. Then the Kantians argue that it is “constitutive” of having free will at all that one must have the capacity to follow a law that one gives to oneself. Finally, they argue that the only possible law of this kind is their supreme principle of practical reason—the law that one ought always to make one’s choices by following a general maxim that one can consistently at the same time will to be a universal law.

Now, the Kantians cannot argue that it is constitutive of having free will that one actually makes all one’s choices by following this fundamental principle; then it would be impossible to violate this principle of rationality—in which case it would surely not be a genuine principle of rationality at all. So, instead, the Kantians typically argue only that it is constitutive of free will that one has the capacity to follow this law. But then the normative question arises yet again. A perfectly sane and intelligent person might ask himself: “I recognize that having the capacity to follow this law is constitutive of having free will at all, but why should I always make my choices by exercising this capacity? Why shouldn’t I sometimes make choices without exercising my capacity to follow this law?” So it seems most doubtful whether this approach can give a satisfactory explanation of why we should follow the Kantians’ supreme principle of practical
reason.

In general, it seems doubtful whether the constructivists can explain why we should comply with the internal requirements of rational choice, in a way that deals adequately with the “normative question”. Indeed, it seems that constructivist views are just as vulnerable to this problem as the “substantive” versions of the recognitional view.

5 The “formal” version of the recognitional view

So far as I can see, there is only one approach to practical reason that avoids this problem—namely, the version of the recognitional view that is based on a purely formal concept of a “good thing for one to do”. As I explained above, it is a conceptual truth, built into the nature of this formal concept of a “good thing to do” that an option is a “good thing to do” in this sense if and only if it is an option that it is “correct” to choose. According to the “formal” version of the recognitional view, a choice is rational if and only if it is rational for the agent to believe that the option chosen is, in this purely formal sense, a good thing to do.

Suppose that someone raises the normative question with respect to this notion of what is a “good thing to do”: “Why should I always choose options that are good things to do? Why shouldn’t I sometimes choose options that are not good things to do?” When the question is understood in this way, it is equivalent to the question: “Why shouldn’t I sometimes choose options that it is not correct for me to choose?” But as I explained in the first section, to say that it is “not correct” to choose an option is just to say that one “shouldn’t” choose it. So asking this question is also equivalent to asking: “Why shouldn’t I sometimes choose options that I
shouldn’t choose?” But that question will hardly perplex any sane and intelligent person! If you shouldn’t choose it, you shouldn’t choose it. That is an utterly trivial truth—not a truth that requires any further explanation. So it may well be a basic requirement of rationality that one should not choose any option that it is not rational for one to believe to be, in this formal sense, a good thing to do.

As I suggested in the second section, this formal concept of a “good thing to do” may, more specifically, be a concept the very content of which is given by its special “conceptual role” in practical reasoning. If so, then it is by definition a concept such that if one is rational, and judges that a certain option is in this sense not a good thing to do, then one cannot need any further reasons to persuade one not to choose that option. Any rational thinker who masters this concept will treat the judgment that an option is not in this sense a good thing to do as simply settling the practical question of whether to choose the option. In effect, it would be built into the very nature of this concept that it is a basic requirement of rationality that one should not choose options that one rationally believes not to be good things to do.

The objection that this view must face is not that it imposes substantive requirements that it cannot adequately explain, but rather that it is empty or trivial—an objection that has been pressed forcefully by Velleman (2000: 174–78). When the term is used in this formal sense, to say that an option is a “good thing to do” is just to say that it is correct to choose it. So the fundamental principle of the recognitional view—that it is correct to choose an option if and only if the option is a good thing to do—amounts to nothing more than the trivial claim that it is correct to choose an option if and only if it is correct to choose it. So how can this principle explain anything at all?
Proponents of the recognitional view could reply as follows. This specification of the fundamental principle, using this formal concept of a “good thing to do”, is indeed trivial. But perhaps the property of being in this formal sense a good thing to do can be also specified in other ways—not just by means of this purely formal concept. For example, perhaps this property is identical to the property of being a morally permissible option that, out of all such morally permissible options, best satisfies the agent’s total set of desires. Then the fundamental principle could also be specified in a non-trivial way—as the principle that a choice is correct if and only if the option chosen is a morally permissible option that, out of all such morally permissible options, best satisfies the agent’s desires.

However, the formal version of the recognitional view itself tells us nothing about which such non-trivial specifications are true. It just tells us to form rational beliefs, involving this purely formal concept, about which of the available options are good things to do and which are not, and then to choose accordingly. But all that it tells us about what it is for an option to be a good thing to do is simply that it is an option that it would be correct to choose. Velleman (2000: 175–76) objects that this “would be … like asking [the agent] to hunt for something described only as ‘the quarry’, or to play a game with an eye to something described only as ‘winning’.” According to this objection, the formal version of the recognitional view is too empty to tell us how to set about making rational choices about what to do. True, it tells us to form rational beliefs about which of the available options are good things to do and which are not. But according to this complaint, this view tells us so little about what it is for an option to be a good thing to do that it cannot tell us how to set about forming such rational beliefs.

In fact, it is not obvious that this objection is correct. It is true that the formal version of
the recognitional view does not by itself give us any substantive specification of what it is for an option to be a good thing to do (although as I have argued, it is quite compatible with the existence of true substantive specifications of this sort). Nonetheless, this view may tell us enough, about the nature of beliefs about what is and what is not in this formal sense a good thing to do, to explain how we are to set about forming such beliefs in a rational way. That is, this view of practical reasoning, together with other independently acceptable truths, may entail an informative *epistemology* for these beliefs.

As I explained in the first section, the recognitional view implies that beliefs about what is and what is not, in this formal sense, a good thing to do play a certain crucial role in practical reasoning. According to the recognitional view, if one is rational, and forms beliefs about which of the available options are good things to do and which are not, then one will choose an option that one believes to be a good thing to do, and not one of the options that one believes not to be a good thing to do. (Indeed, as I suggested in the second section, this may be precisely what determines the content of the formal concept of a “good thing to do”.) So, for example, if one is rational and believes ‘A is the only available option that is a good thing to do’, then one will choose A.

In this sense, being rational and having a belief of this kind is sufficient to *motivate* one to make a choice. Nothing like that is true of most beliefs. For example, one might be perfectly rational, and believe ‘A is the only available option that will lead to my winning this game of Scrabble’, and yet not be motivated in any way to choose A (perhaps because one knows that one has compelling reason not to win this particular game of Scrabble). Thus, it is a quite special feature of beliefs about what is and what is not a good thing to do that they are sufficient to
motivate choice in this way. But according to a plausible thesis, known as the “Humean Theory of Motivation”, all motivation of choice involves desire.25 And according to another thesis, which Jay Wallace (1990: 370) has called the “desire–out, desire–in” principle, “processes of thought which give rise to a desire (as ‘output’) can always be traced back to a further desire (as ‘input’)”. Together, these two theses imply that no belief can be sufficient to motivate choice in this way unless that belief was itself originally produced, at least in part, by some of one’s desires. So, assuming these two theses about motivation, beliefs about what is a good thing to do must also be based, in some way, on some of one’s desires. (This is not to say that these beliefs can be based on any of one’s desires; it may be only certain special desires that can play this role as part of the ultimate rational basis for these beliefs.) Moreover, since one’s practical reasoning must ultimately be based on ordinary empirical information about the available options and what their non-normative properties would be, the same must be true of these beliefs. Finally, the general principles of rational belief apply to these beliefs as much as to any others: one should try to keep these beliefs consistent, unified by relations of explanatory coherence, and so on.

Roughly then, the epistemology of these beliefs is as follows. A belief about a particular option A, to the effect that A is, in the “formal” sense, a “good thing to do”, must be based on (i) ordinary empirical beliefs about A, to the effect that A has such-and-such non-normative properties, and (ii) some at least rough general principle, to the effect that, at least typically and for the most part, options that have those non-normative properties are good things to do.26 This rough general principle must in turn be based on a search for what John Rawls (1972: 46–53) has called “reflective equilibrium”. This is a process that starts out by treating one’s disposition to have a certain sort of desire for certain options as weak, prima facie evidence that those options
are in the formal sense good things to do; then this process then aims to form, on the basis of this evidence, a maximally coherent set of rough general principles about which options are (in this sense) good things to do.27

This is only a rough sketch of this epistemological account. Much more would have to be said to make it plausible that this is an adequate account of the epistemology of the relevant beliefs. My main point here is simply to argue that it is far from obvious that no such account can be developed. Once such an account is developed, the formal version of the recognitional view will clearly be able to give an informative account of how rational practical reasoning proceeds: one first follows the rules and procedures that are specified in this epistemology to arrive at rational beliefs about what is and what is not a good thing to do; and then one chooses accordingly, by choosing an option that one rationally believes to be a good thing to do.

6 Instrumental reasoning and choosing in uncertainty

The formal version of the recognitional view appears, then, to have the resources to deal both with Velleman’s “emptiness objection” and with the demand for explanation that is expressed by Korsgaard’s “normative question”. But can it really give a satisfactory account of practical rationality? In this final section, I shall sketch a refined version of this recognitional view. I shall try to make it plausible that this refined recognitional view gives an adequate account of two crucial aspects of rational practical reasoning: the importance of instrumental reasoning, and the conditions for rational choice in uncertainty.

A great deal of practical reasoning is instrumental. One starts by considering an end that
one intends to achieve, and then chooses which *means* to use in order to achieve that end. Indeed, such instrumental reasoning is involved in almost all cases when practical reasoning leads to action. (The only exception is the special case where the option that one judges to be a good thing to do and then chooses is what Arthur Danto (1968) has called a “basic action”—that is, something that one can just do, without having to choose any further means in order to do it.) But how can the recognitional view account for the crucial role of such instrumental reasoning?

To account for instrumental reasoning, the recognitional view has to be refined, in two ways. First, the fundamental principle must be refined so that applies not only to choices but also to *intentions*. (A choice is a conscious mental event involving the formation of an intention, while an intention is an enduring mental state that typically lasts until it either leads to action or is abandoned.) Second, this fundamental principle must be refined so that, for an intention to be correct, it is not enough that one’s intended course of action should be a good thing to do; it must also be a course of action that *one will actually carry out* if one intends to do so. The underlying idea here is that an intention that one simply will not execute is pointless, and to that extent a mistaken or incorrect intention. So I propose the following revised version of this fundamental principle: one’s intentions about how to act in a certain situation are correct if and only if one will in fact execute those intentions and thereby do something that really is a good thing to do in that situation.

The account of practical rationality must also be revised accordingly. I shall use the phrase “the way in which one revises one’s intentions” broadly. Thus, one “way in which one may revise one’s intentions” is to form a new intention; but another way is to abandon (or to reaffirm) an old intention; and a third way is not to form any intention at all. Then we can say
that the way in which, on a given occasion, one revises one’s intentions about how to act in a
given situation is rational if and only if it is rational for one to believe that one will execute all
one’s intentions about how to act in that situation and thereby do something that is a good thing
to do in that situation.

This account of rational intention revision can explain the role of instrumental reasoning.
Suppose that you intend to achieve a certain end; but suppose that it is not rational for you to
believe that you will execute this intention and successfully achieve this end, unless you also
form an intention about which means you will use in order to achieve the end. According to this
account, the way in which you revise your intentions about how to act in this situation will be
rational only if it is rational for you to believe that you will execute all the intentions (if any) that
you have about how to act in this situation. In this case, there are only two ways of revising
your intentions that will meet this condition: you can either abandon your intention to achieve the
end, or you can form an intention about which means you will use in order to achieve the end.
For example, suppose that you intend to be in Montreal tomorrow; but suppose that, unless you
form some intention now about how you will travel to Montreal—whether to fly or drive or take
the train, for instance—it will not be rational for you to believe that you will actually be in
Montreal tomorrow. Then, to be rational, you must either abandon your intention to be in
Montreal tomorrow, or form an intention about how you will travel to Montreal.

In a case of this sort, which of these two ways of revising your intentions is
rational—forming an intention to take the means, or abandoning your intention to achieve the
end? The answer depends on which of the two it is rational for you to believe to be a good thing
to do. For example, if it is rational for you to believe that abandoning the intention to achieve the
end is a good thing to do, and not rational to believe that taking any of the available means to that end is a good thing to do, then the only rational response is to abandon your intention to achieve the end. On the other hand, if it is rational for you to believe both that abandoning your intention to achieve the end is a good thing to do, and that taking the means to that end is also a good thing to do, then both ways of revising your intentions are equally rational.

Suppose that you do form an intention about which means to use to achieve your end. How is it rational to choose which means to use? Clearly, you must choose means that it is rational for you to regard as effective at achieving the end. But often there are many means that are equally effective at achieving the end, even though many of these means have grave drawbacks of other kinds (for example, some of these means might be excessively expensive, or boring, or painful, or morally impermissible). In these cases, the refined version of the recognitional view that we are currently considering implies that it is rational to choose a course of action as a means to your end only if it is rational for you to believe that course of action to be, not only effective at achieving the relevant end, but also (in the purely formal sense) a “good thing to do” overall.

Finally, how should one choose what to do if one simply does not have enough time or information for it to be rational for one to hold any outright belief about which of the available options is a good thing to do? How should one choose what to do, if one cannot be rationally certain which of the available options are good things to do? I shall suppose here that even if it is not rational to have an outright belief in a proposition, it may still be rational to have a partial belief in that proposition; that is, one may put some credence in the proposition, but also put some credence in the proposition’s negation. Specifically, I shall suppose that it is rational for
one to proportion one’s degree of belief in the proposition to the proposition’s probability given one’s evidence.

To take advantage of the fact that it is rational to have partial degrees of belief in some propositions, however, we must recognize that the goodness of options also comes in degrees. Some options are better (in a purely formal sense) than others. Just as the goodness of options comes in degrees, then, so too does the correctness of choices. The better the option that one chooses, the closer one’s choice is to being correct. If one’s choice is completely correct, the option chosen must be optimal—that is, there must be no available alternative that is better.

Moreover, options can not only be ranked according to which are better and which are worse. In some cases, an option might be only slightly less good than optimal, while in other cases, an option might be much less good than optimal. (For example, one might not know whether A or B is optimal, but know that if A is optimal, then B is much less good than A, while if B is optimal, A is only slightly less good than B.) At least to some extent, then, we can make sense of the notion of the degree to which an option falls short of optimality. Let us suppose that these “degrees of shortfall from optimality” can, at least in principle, be measured by means of real numbers. (An option that is optimal can be regarded as falling short of optimality to degree 0—that is, not falling short of optimality at all.)

We can now define the notion of an option’s “expected shortfall from optimality” as follows. Consider a set of propositions each of which describes each of the available options as falling short of optimality to a certain degree. Let us call these propositions the “relevant uncertain propositions”. One may not know for certain which of these relevant uncertain propositions is true. But suppose that one does know for certain that exactly one of these
propositions is true. Moreover, suppose that one is also rationally certain that these uncertain propositions are all beyond one’s control; nothing that one can do can determine which one of these propositions is true. Finally, suppose that each of these relevant uncertain propositions has a certain probability given one’s evidence. Then, we can say that an option’s “expected shortfall from optimality” is the sum of the degrees to which the option falls short of optimality according to each of these relevant uncertain propositions, weighted by the probability of that proposition’s being true. I propose that it is rational to choose an option if and only if that option has minimal expected shortfall from optimality. (In effect, this proposal about how to make choices in uncertainty is directly modelled on expected utility theory, except that it uses an objective evaluative notion—“shortfall from optimality”—instead of the more familiar notion of a “utility function”.)

This proposal—that one should always choose an option with minimal expected shortfall from optimality—is a plausible generalization of the simple version of the recognitional view that we considered above. This proposal entails that if there are any options such that it is rational for you to have the outright belief that those options are optimal, then it is rational to choose any of those options. If it is rational for you to have the outright belief that a certain option is optimal, then, given your evidence, the probability that that option is optimal is 1; so the expected shortfall from optimality of such an option will always be 0. On the other hand, this proposal seems a plausible extension of the simple version of the recognitional view to cases in which it is not rational to have any outright belief about which of the available options are optimal. In such cases, you cannot choose an option such that it is rational for you to have an outright belief in the proposition that that option is optimal, since ex hypothesi it is not rational for you to have any
such outright belief. So it seems plausible that the best thing to do in this case is to choose an option with minimal *expected* degree of shortfall from optimality. If you cannot aim at what you firmly believe to be optimal, you can at least aim to *minimize your expectation of falling short* of optimality. In this way, then, the recognitional view may be able to give an adequate account of how it is rational to make choices if one is uncertain about which of those options are good things to do and which are not.

Many more refinements may be necessary before the formal version of the recognitional view can give a fully satisfactory account of rational choice. But in this section I hope to have made it plausible that the view has the resources to provide such a satisfactory account. Given that this view also seems to have a plausible answer both to Korsgaard’s “normative question” and to Velleman’s “emptiness objection”, it seems to me that this view is a uniquely promising approach to understanding practical reason.

**Notes**

1 This essay is an extensively revised restatement of the arguments of an earlier paper (Wedgwood 2002a), which was based on the talk that I gave at the conference on Weakness of Will and Varieties of Practical Irrationality at the Université de Montréal in May 2001. This restatement of those arguments was presented to audiences at Oxford and at the University of Arizona in February 2002. I am grateful to all those audiences (and especially to my commentator in Montreal, Josée Brunet), and also to Carla Bagnoli, John Broome, Alex Byrne, Philip Clark, Ned Hall, Elizabeth Harman, Sally Haslanger, Richard Holton, Leonard Katz, Kathrin Koslicki, Jim Pryor, Susanna Siegel, David Velleman, and the editors of this volume, for helpful comments.

1 I need not claim that the term ‘should’ is simply *ambiguous*. It may be that ‘should’ is *context-sensitive*. Perhaps, for example, ‘should’ is always implicitly relativized to a contextually
determined parameter of some sort. Then we could say that the “external” or “objective” ‘should’ is relative to the parameter of this sort that is determined by certain contexts, while the “internal” or “subjective” ‘should’ is relative to the parameter of this sort that is determined by other contexts. Unfortunately, I cannot go further into these semantic issues here.

That is, I am assuming that something like the thesis that epistemologists—such as Fumerton (1995: 60–69)—call “internalism” holds of rational choice just as much as of rational or “justified” belief. I should emphasize that in using the terms ‘internal’ and ‘external’ to indicate this contrast between two ways of evaluating choices, I am following the usage that is standard among epistemologists. I am not following the usage that is most common among meta-ethicists. So it should not be assumed that what I am here calling the “internal ‘should’” has any relation to the so-called “internal ‘ought’” that has been discussed by Stephen Darwall (1995: 9–12) and others.

Christine Korsgaard (1996a: 35) calls these two views of practical reason “realism” and “constructivism” respectively.

So, if one has sufficiently misleading evidence, it might be rational for one believe an option to be a good thing to do, even if it is not in fact a good thing to do. In this case, according to this account of rational choice, it is rational to choose the option, even though the option is not a good thing to do. Conversely, misleading evidence might prevent it from being rational to believe an option to be a good thing to do, even though in fact it is a good thing to do. In this case, it is irrational to choose the option, even though the option is a good thing to do.

I have developed this approach to understanding the relation between internal and external uses of ‘should’ in other work (Wedgwood 2002b).

In other work (Wedgwood 2002c), I have suggested that the slogan that “belief aims at the truth” is best interpreted as making a claim about belief that is analogous to the claim that the recognitional view makes about choice. The fundamental principle applying to belief is that a belief is correct if and only if the proposition believed is true. This fundamental principle of correct belief explains the principles of rational belief. Roughly, the formation of a belief is rational if and only if, in forming that belief, the believer was following the appropriate rules and procedures—namely, those rules and procedures that it is rational for the believer to regard as reliable ways of reaching the truth.

Proponents of the recognitional view need not maintain that the good is the “constitutive aim” of all desire or of all action as such. They can agree with David Velleman (2000: 117–122, 190–91) that it is quite possible for one to desire states of affairs, or to perform actions, that one believes not to be good in any way. Proponents of the recognitional view only insist that if one performs such actions, one is akratic and therefore irrational.

John Broome has objected to me that in presenting the issues in this way, I am conflating two separate questions: the question of the relative priority of the “internal” and the “external”
‘should’, and the question of the relation between the terms ‘should’ and ‘good’. But as I explain in §2, there is a “formal” way of using the term ‘good’ such that to say that a certain option is “not a good thing to do” in this way just is to say that it is something that one “should not do”. There are also many more “substantive” ways in which something can be good: for example, something can be good for me, or good for Oxford University, or morally good. The question of how these substantive ways of being good are related to what one “should” do is one that I shall not be addressing here.

9 I describe this as the “central or canonical” method of practical reasoning because strictly speaking, according to this view as I formulated it, it may also be rational to take certain shortcuts in one’s practical reasoning. So long as it is rationally permissible for one to believe that the chosen option is a good thing to do, and one’s choice is suitably sensitive to whatever considerations make it the case that it is rationally permissible for one to believe this, then one’s choice may be rational even if one has not actually formed any belief at all about whether or not the option in question is a good thing to do.

10 See Aristotle’s definition of “practical wisdom” (phronêsis) in Nicomachean Ethics VI.5, 1140b5 (repeated at 1140b22); and compare also VI.2, 1139a24, and VI.9, 1142b34.

11 I tentatively suggest that the broadly procedural account of rational practical reasoning sketched by Bernard Williams (1982: 104–5) is also a version of constructivism. Williams seems to think that it is just intuitively clear that certain procedures of deliberation are rational. He certainly does not try to explain why these procedures are rational by appealing to the good external results that complying with these procedures may be expected to have.

12 This interpretation of Kant—which is basically due to Korsgaard (1996b: 43–76)—is controversial, and rejected some philosophers, such as Wood (1999), who regard themselves as Kantians. But for present purposes, it does not matter whether this interpretation of Kant is correct. All that matters here is that the view that I am calling “Kantian constructivism” is worthy of serious examination.

13 For example, Richard Jeffrey (1983) talks of the principles of rational choice as the “logic of decision”; and Korsgaard (1996a: 235) claims that, just as “if I am going to think I must think in accordance with the principle of non-contradiction”, so too, in essentially the same way, “if I am going to will at all I must do so universally… The requirement of universality is in this way constitutive of willing.”

14 Kant repeatedly insists that the fundamental principle of practical reason must be a priori (1785: 388–90, 406–12, 425–27). Some decision-theoretic constructivists, such as Gauthier (1985) and Dreier (1997), also seem to suggest that the fundamental requirements of rational preference are a priori.

15 Compare Christine Korsgaard (1996b: 122): “what makes the object of your rational choice good is that it is the object of a rational choice”. Here Korsgaard is broadly speaking following
Kant, who appears to believe that an option is a good thing to do (not merely good in relation to some end, but good simpliciter) just in case it is an option that no well-informed agent can reject without violating the fundamental internal requirement of rationality.

To say that \( X \) is a “good thing to do” (in this “formal” sense) does not imply that one should do \( X \). One might be in a “Buridan’s ass” case, in which both \( X \) and \( Y \) are good things to do, but it is impossible to do both. In this case, it is not true that one should do \( X \) or that one should do \( Y \). Still, they are both good things to do.

This instrumentalist version of the recognitional view must be distinguished from the decision-theoretic version of constructivism, since according to this instrumentalist view, the internal requirements of rational choice are explained by a more fundamental principle that defines when a choice counts as (externally) correct.

This point is more or less explicit in Kant (1785: 396–400), who claims that the good will—i.e., the will that complies with the fundamental principle of rational choice, rejecting all non-universalizable maxims precisely because they are not universalizable in the relevant way—is valuable purely in itself, not merely because of its actual or expected results.

Decision-theoretic constructivists often appeal to the so-called “money pump” argument. For some criticism of the “money pump” argument, see Broome (1999: 74–75) and Maher (1993: 36–38).

Alternatively, this argument could be based on the idea that unless one conforms to these axioms, one will not count as a unified agent with a genuine will, as opposed to a bundle of disparate desires and needs; see Gauthier (1985).

I need not claim that such constitutive claims are simply irrelevant to explaining why we should comply with the basic requirements of rationality. Indeed, I suggested that such constitutive claims provide part of the explanation in some of my earlier work (Wedgwood 1999: §4). All that I am claiming here is that these constitutive claims are not sufficient all by themselves to provide such an explanation. In the case of the basic requirement that one should not believe logical contradictions, for example, it also seems crucial that it is rational for one to believe that one must conform to this requirement if one is to reach the truth about the question at issue.

One philosopher whom it may be correct to interpret as accepting this formal version of the recognitional view is Lawrence (1995).

For Velleman (2000: 190), this “emptiness charge” is just one stage in his argument for a conception of practical reason that seeks “to avoid the twin pitfalls of internalism and externalism”. Specifically, Velleman attacks the recognitional view of practical reason because it appears to provide one way of defending “externalism”—by which he means the view that there are some reasons for action that are entirely independent of the agent’s contingent desires. Even if the recognitional view would provide one way of defending “externalism”, however, we
should not assume without further discussion that the recognitional view is committed to “externalism”. (Perhaps it is not rational to believe that an action is a good thing to do unless that action satisfies one of one’s contingent desires.) At all events, I shall focus purely on Velleman’s objection to the recognitional view, and ignore the other aspects of his arguments here.

As Velleman (2000: 176–78) claims, there cannot be any “object of practical reasoning” unless there is some substantive specification of what achieving this object consists in. As he puts it: “A game whose object was specifiable only as ‘winning’ wouldn’t have an object—that is, wouldn’t have any object in particular” (176).

I have given a more careful statement of the argument for the conclusion that the Humean Theory of Motivation entails a broadly “sentimentalist” epistemology for beliefs of this sort in other work (Wedgwood 2002d). But why should we accept the Humean Theory of Motivation? Very roughly, I suspect that the reason is that, as Dummett (1991: 287) has often maintained, there must be a sort of “harmony” between the grounds for holding a certain belief, and what that belief commits one to. Beliefs about what is a “good thing to do” have the very special feature that they commit one to making certain choices. So there must be something equally special about the grounds on the basis of which those beliefs are held; for example, perhaps these grounds must include desires—which are states that essentially involve a disposition to make the corresponding choices.

Suppose that one of these “rough general principles” is the principle that (at least typically and for the most part) no option that has property P is a good thing to do. Why can’t someone raise the normative question again, “Why should I (at least for the most part) avoid choosing options that have property P?” But this principle is not an internal requirement of rationality. It is an external fact about which options are good things to do. So, even if it is just a basic principle that cannot be any further explained, there is no reason to expect that it must be convincing to all agents who being sufficiently “sane and intelligent”.

We should perhaps also require that it must be rational for one to expect that a process of further refining this rough general principle, and further empirical investigation of A and the other available options, would not finally defeat the conclusion that A is a good thing to do.

Objection: Sometimes, making a choice that one will not execute will have extremely good effects. For example, making a choice that one spectacularly fails to execute might shock one into mending one’s ways, with the result that one does things in the future that are much better than one otherwise would have done. So why isn’t the initial choice correct after all? Reply: It is true that in such cases, the higher-order choice to manipulate oneself into making a choice that one will fail to execute may itself be a correct choice. But this does not mean that the lower-order choice that one manipulates oneself into making is a correct choice after all. (Similarly, sometimes the choice to manipulate oneself into holding a false belief is a correct choice; this does not make the false belief a correct belief after all!)
Ideally, this principle should also be refined so that it applies to such states of affairs as one’s failing to hold an intention about how to act in a situation in which one will in fact find oneself. This could be done by revising the fundamental principle so that it reads as follows: one’s holding of whatever intentions one holds (if any), about how to act in a certain situation, is correct if and only if (i) one will in fact execute all those intentions (if any), and (ii) if one executes all those intentions (if any), one will do something that really is a good thing to do in that situation. This second clause can explain when failing to hold any intention at all is itself a mistake. (1) Not holding any intention about how to act in a given situation is itself a mistake if it is true that if one executes all one’s intentions about how to act in that situation (i.e., no intentions at all), one will do something that is not a good thing to do (e.g., one will dither ineffectually while disaster strikes). (2) Often, however, there is nothing mistaken about not holding any intentions at all. If one will do something that is a good thing to do in the situation in question even if one holds no intentions at all about how to act in that situation, then it is still true that if one executes all one’s intentions about how to act in that situation (i.e., no intentions at all), one will do something that is a good thing to do.

This point—that rationality requires that one revise one’s intentions so that it is rational for one to believe that one will in fact execute one’s intentions—explains the two demands that Bratman (1987: 31) puts on intentions: first, that one’s intentions must be “strongly consistent relative to one’s beliefs”; and second, that they must be “means-end coherent”.

This uncertainty may partly be due to uncertainty about what the causal effects of these options will be. But even if one is quite certain what all the causal effects of each of two options will be, one may still be uncertain whether it is better for one to perform the first option, with all of its causal effects, or to perform the second option, with all of its causal effects. This is why we need to focus on uncertainty about normative or evaluative propositions of this sort.

I have here followed so-called “causal decision theory” (Joyce 1999: 115–18) in relying on the unconditional probability of the relevant uncertain propositions, rather than on their probability conditional on one’s performing the option in question. (This is why I have also had to follow the causal decision theorists in assuming that the “relevant uncertain propositions” are all “beyond one’s control”.) If the notion of an option’s “shortfall from optimality” according to a certain “relevant uncertain proposition” is identified with the difference between its utility and the utility of the option that is optimal according to that proposition, then this account of rational choice will actually coincide with the “causal” version of expected utility theory. It would be a worthwhile project to investigate what the consequences would be of revising this account by relying on conditional probabilities (instead of unconditional probabilities), and on other definitions of “shortfall from optimality” as well. But unfortunately I cannot carry out this investigation here.
References


   *Topoi*, 21: 139–152.


