

# Decisions, Reasons, and Rationality\*

*Garrett Cullity*

What difference do our decisions make to our reasons for action and the rationality of our actions? There are two questions here and good grounds for answering them differently. However, it still makes sense to discuss them together. By thinking about the relationships that reasons and rationality bear to decisions, we may be able to cast light on the relationship that reasons and rationality bear to each other.

Two forceful thoughts have set the agenda for recent discussion of this pair of questions and suggest that the answers must be different. On the one hand, your decisions surely do make a difference to what it is rational for you to do. If you decide to do something (and do not rescind the decision), not taking suitable steps toward implementing your decision can be irrational. And this seems true even if you ought not to have decided to do it. On the other hand, though, it is hard to see how decisions that you ought not to have made can be reasons in favor of the things you ought not to have decided to do.

These two thoughts set the agenda here, too, and explain the structure of what follows. The second thought is often expressed as the complaint that treating decisions as reasons amounts to an objectionable kind of “bootstrapping.” After explaining the force of that complaint in Section II, I ask what exactly it shows. It does not show that our decisions never make a difference to our reasons for action: Section III explains three kinds of difference they can make. Moreover, in the third kind of case—one that is overlooked by most discussions of these issues—decisions that we ought not to have made can indeed provide us

\* John Broome’s written comments on a previous draft of this article were invaluable. I thank the organizers of the conferences of the Australasian Association of Philosophy and the British Society for Ethical Theory and the ANU conference on the work of John Broome for the opportunity to present my ideas. For helpful comments on those and other occasions, I would like to thank Joseph Raz, Andrew McGonigal, Jennie Louise, Jordi Fernandez, Michael Smith, Wlodek Rabinowicz, Jamie Dreier, Geoff Brennan, Dave Chalmers, George Couvalis, and—someone I cannot now thank as I would like—the late Maurice Goldsmith.

*Ethics* 119 (October 2008): 57–95

© 2008 by The University of Chicago. All rights reserved. 0014-1704/2008/11901-0005\$10.00

with reasons. However, when this happens, there are special circumstances that explain it. Outside those circumstances, decisions cannot bootstrap themselves into reasons. The rest of the article examines the question this leads to: When decisions are not reasons, how can we explain their impact on rationality? Section IV examines John Broome's important work on this question and points out some of its limitations. Sections V and VI offer an alternative description of the bearing of decisions on rationality—one that seeks to overcome those limitations. Finally, in Sections VII and VIII, I turn from describing the bearing of decisions on rationality to trying to explain it: this leads me to an account of the relationship between reasons and rationality.

### I. PRELIMINARIES

To focus the questions I am asking, I need to begin by saying something about each of the three terms in my title. A 'decision', I shall take it, is the conscious formation of an intention to do something. This identifies decisions by placing one restriction on the class of intention formations: they must be conscious. Other writers sometimes place different restrictions on that class.<sup>1</sup> However, the decisions I shall be discussing are ones which I think would satisfy any plausible definition.

An intention to do something is a commitment to doing it. On the approach championed by Michael Bratman, the sense in which intentions are commitments can be explained by reference to the roles that they characteristically play in our planning activity and the norms that govern those roles.<sup>2</sup> Here, it will be helpful to distinguish two such roles. When you intend E, you will typically plan toward E by doing what promotes and avoiding what conflicts with the realization of E, coordinating this with the pursuit of your other intended ends. You will often also plan from E, using the assumption that E will occur in thinking about what to do subsequently. It is possible to be committed to E in the first of these ways but not the second: you might be orchestrating your action, compatibly with your other intentions, toward E but without being confident enough of success to rely on the attainment of E in

1. For the view that deciding is forming an intention after deliberation, or by reasoning, see Joseph Raz, "Reasons for Action, Decisions and Norms," *Mind* 84 (1975): 481–99, 488–89; Michael E. Bratman, "Intention and Means-End Reasoning," *Philosophical Review* 90 (1981): 252–65, 252; and John Broome, "Practical Reasoning," in *Reason and Nature: Essays in the Theory of Rationality*, ed. Jose Bermudez and Alan Millar (Oxford: Clarendon, 2004), 28–55, 86. My usage is broader, in allowing that we can just decide to do things, without deliberation. Perhaps it is also narrower, if it is possible to form an intention unconsciously after deliberation. For discussion, see Storrs McCall, "Decision," *Canadian Journal of Philosophy* 17 (1987): 261–88.

2. Michael Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).

further planning. This opens up the possibility of a debate about whether true intentions play both roles, only the first, or neither.<sup>3</sup> I prefer to sidestep this verbal issue by making the following stipulation. ‘To-intentions’, I shall say, are states of the kind that characteristically plays the first role; ‘from-intentions’ are states of the kind that characteristically plays both. This distinction will be important later.

These remarks leave some things obscure. They claim that to- and from-intentions “characteristically” and “typically” play these roles in planning. Eventually, I shall be replacing these phrases with “to the extent that you are rational.” But I have some explaining to do before I reach that point. This is enough to get us started.

Next, a stipulation about ‘reasons’. It is not controversial whether decisions can be explanatory reasons why actions occur. The interesting question is whether decisions are reasons to do things—that is, whether they are reasons for action, reasons that count in favor of actions, or, as I shall follow others in saying, normative reasons for action. I need to make a further stipulation. ‘Reasons’, as I use the term here, will refer to those awareness-independent considerations whose normative bearing on action we deliberate about: what some writers call “objective” rather than “subjective” normative reasons.<sup>4</sup> This is the sense of ‘reason’ in which, if my doctor convincingly tells me that a certain drug will cure me when in fact it will kill me, there is no reason for me to take the drug. Confining the use of ‘reason’ to this sense does seem stipulative: ordinary usage does not clearly insist on it. But it is a good stipulation to make since it focuses our attention on what really matters. When my doctor or I deliberate about whether I should take the drug, what we are concerned about is what effect it really will have on my health.<sup>5</sup>

Several writers take reasons of this kind to be conceptually primitive,

3. As I read them, according to Bratman, the answer is the first; according to John Broome, the second; according to Hugh McCann, neither. For the disagreement between Bratman and Broome, see Michael E. Bratman, “Intention, Belief, and Instrumental Rationality,” in *Reasons for Action*, ed. David Sobel and Steven Walls (Cambridge: Cambridge University Press, forthcoming), sec. 5; and John Broome, “The Unity of Reasoning?” in *Spheres of Reason*, ed. Simon Robertson, John Skorupski, and Jens Timmerman (Oxford: Oxford University Press, forthcoming), sec. 7. For McCann, aiming purposefully toward an outcome suffices for intention, so that I can rationally intend two ends while recognizing that they are not co-realizable. He thus denies the agglomeration condition on intentions discussed in Sec. VIII below. See Hugh McCann, “Settled Objectives and Rational Constraints,” *American Philosophical Quarterly* 28 (1991): 25–36.

4. For example, Mark Schroeder, “The Scope of Instrumental Reason,” *Philosophical Perspectives* 18 (2004): 337–64, 348.

5. Compare Joseph Raz, *Practical Reason and Norms*, 2nd ed. (Oxford: Oxford University Press, 1999), 18.

and I am inclined to agree with that view.<sup>6</sup> But it is debatable—John Broome denies it—so I will not assume that here.<sup>7</sup>

‘Rationality’, as standardly used, is a term of evaluation. Evaluation of what? One of my main aims in what follows will be to make progress in answering that question. A first attractive-looking answer is this: “evaluation of my rationality is evaluation of my thoughts about reasons, and of whether my actions and attitudes are appropriately affected by those thoughts.” That is vaguer than we need.<sup>8</sup> But an account of this kind—call it an “apparent reasons” account of rationality—gives us a good place to start in looking for the right thing to say about examples like the one just given. If all the evidence I have suggests that the drug will cure me, then it is rational for me to think there is a reason to take it and (if I do think that) to take it. Taking the drug “makes sense.”

This suggestion, however, is only one of a variety of competing accounts of the relationship between reasons and rationality. Many such accounts can be classified as belonging to two broad families. “Responsiveness” accounts explain rationality as responsiveness to reasons or apparent reasons. “Proper functioning” accounts reverse this order of explanation. These think of rationality as proper mental functioning and then explain reasons as the considerations that have a motivational influence on agents to the extent that those agents are functioning properly and possess relevant information.<sup>9</sup> In fact, the most influential traditions of thinking about practical reason have taken this latter form. The label ‘proper functioning account’ naturally calls to mind the kind of neo-Aristotelian account advanced by Philippa Foot, but many neo-

6. See, e.g., *ibid.*, 11–12; T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), 17; Jonathan Dancy, *Ethics without Principles* (Oxford: Clarendon, 2004), 15, 29.

7. John Broome, “Reasons,” in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, ed. R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith (Oxford: Clarendon, 2004), 29–42. For a different attempt to identify a more primitive concept underlying that of a normative reason, see Sven Danielsson and Jonas Olson, “Brentano and the Buck-Passers,” *Mind* 116 (2007): 511–22. For Allan Gibbard’s expressivist treatment, see *Thinking How to Live* (Cambridge, MA: Harvard University Press, 2003), 188–91.

8. Noteworthy proposals for sharpening it up are made by Derek Parfit, “Reasons and Motivation,” *Proceedings of the Aristotelian Society*, suppl. ser., 71 (1997): 99–130, 99, and “Rationality and Reasons,” in *Exploring Practical Philosophy: From Action to Values*, ed. Dan Egonsson, Jonas Josefsson, Bjorn Petersson, and Toni Ronnow-Rasmussen (Aldershot: Ashgate, 2001), 17–39, 25, 29; Scanlon, *What We Owe to Each Other*, 22–32; Joseph Raz, “Explaining Normativity: On Rationality and the Justification of Reason,” in *Engaging Reason: On the Theory of Value and Action* (Oxford: Oxford University Press, 1999), 67–89, sec. 1; and Niko Kolodny, “Why Be Rational?” *Mind* 114 (2005): 509–63, sec. 5.

9. It is of course possible to think of rationality as proper functioning without making rationality explanatorily prior to reasons: see, e.g., Bratman, “Intention, Belief, and Instrumental Rationality,” sec. 1; and Joseph Raz, “The Myth of Instrumental Rationality,” *Journal of Ethics and Social Philosophy* 1 (2005): 2–28, 15.

Kantian and neo-Humean views have the same general structure, explaining what it is to have a reason to act by reference to prior accounts (albeit very different ones) of what it is to be functioning rationally.<sup>10</sup>

Apparent reasons accounts, since they are a kind of responsiveness account, are opposed by proper functioning accounts. They are also opposed by other kinds of responsiveness accounts. In particular, there are those that contain a more direct element, making rationality at least in part a matter of our responsiveness to reasons themselves rather than just our thoughts about them. In favor of this, the following sort of case can be offered (a “Huckleberry Finn” case, we might say).<sup>11</sup> There are facts that do amount to decisive reasons for you to perform an action (Jim’s humanity counts decisively in favor of helping him to hide from the slave hunters); you falsely believe there are decisive reasons not to perform that action, but then, in opposition to this judgment, you go ahead and do the right thing, in a way that is responsive to the facts that are indeed good reasons. If you do this, there is something wrong with the way you deliberate and act. But arguably, there is also something importantly right: you respond to the reasons in the right way. And perhaps we should mark what is right about your responsiveness to the reasons by saying that it is a kind of rationality. I return to this kind of case later.

The discussion that follows belongs, like responsiveness accounts,

10. For examples of the neo-Aristotelian view, see Philippa Foot, *Natural Goodness* (Oxford: Clarendon, 2001), esp. chap. 4; see also Gavin Lawrence, “The Rationality of Morality,” in *Virtue and Reasons*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Clarendon, 1995), 89–147. The neo-Humean view presented in Bernard Williams, “Internal and External Reasons,” in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 101–13, has this broad structure, as does the neo-Kantian view advocated by Christine M. Korsgaard, “Skepticism about Practical Reason,” *Journal of Philosophy* 83 (1986): 5–25. ‘Neo-’ in each case because the theories of Aristotle, Hume, and Kant themselves do not obviously deploy the concept of a “reason” at all. For a summary of some of the principal differences between views of these three broad kinds, see the introduction to Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Clarendon, 1997).

11. For several helpfully detailed examples of this kind, see Nomy Arpaly, “On Acting Rationally against One’s Best Judgment,” *Ethics* 110 (2000): 488–513. See also Harry Frankfurt, “Rationality and the Unthinkable,” in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 177–90; Alison McIntyre, “Is Akratic Action Always Irrational?” in *Identity, Character and Morality*, ed. Owen Flanagan and Amelie Rorty (Cambridge, MA: MIT Press, 1993), 379–400; and Robert Nozick, *The Nature of Rationality* (Princeton, NJ: Princeton University Press, 1993), chap. 4. The following formulation in Parfit also seems to imply that Huck is rational: “We are rational insofar as we respond to reasons, or apparent reasons. We have some *apparent* reason when we have some belief whose truth would give us that reason” (Parfit, “Rationality and Reasons,” 25). My own use of the phrase “apparent reasons” in the text differs from Parfit’s. As I use it, *F* is an apparent reason for you to *X* when it appears to you that *F* is a reason to *X*.

to the general project of explaining rationality by reference to reasons. My destination, however, is not a responsiveness account but instead what I call a “standard-fixing” account of the relationship between rationality and reasons. Preliminaries over, we can now embark on the main line of argument.

## II. BOOTSTRAPPING

Once we distinguish reasons themselves from our responses to them in the way just outlined, we confront a strong objection to the idea that decisions themselves could be reasons.<sup>12</sup> Return to the example of the drug. Before I have started thinking about it, there is no reason to take the drug since it will not cure me. If I form the belief that it will cure me, there is still no reason to take the drug since even if my belief is rational it is false. So if I form a second belief that there is a reason to take the drug, that is false too. Now suppose that having formed these beliefs, I decide to take the drug. On a view according to which decisions are reasons, the truth value of my second belief now changes. My decision, which was based on a false belief, itself turns that belief from false to true!<sup>13</sup>

Perhaps this will not seem so bad once we observe that the very strong reason not to take the drug—it will kill me—remains in place. My decision to take it may give me a reason to do so, but that reason is clearly outweighed.<sup>14</sup> However, there will be variants of the example for which that is not true. How strong a reason is my decision supposed to give me for taking the drug? Call that strength  $s$ . Now progressively weaken the bad effects of the drug, so that it produces not death but debilitation, pain, discomfort, annoyance, irritation . . . . Eventually, we will have a drug for which the reason not to take it has a strength less than  $s$ . And now we have an example in which, before my decision, there is a decisive reason for me not to take the drug, but if my false belief that there is a decisive reason for me to take it leads me to decide to take it, that belief thereby turns from false to true.<sup>15</sup>

What makes this hard to accept is that, in a case like this, the reasons for or against deciding to do something just are the reasons for or against

12. An idea advocated, e.g., by Raz, *Practical Reason and Norms*, 66–73.

13. There is also the converse case. Suppose I truly believe that the drug will not cure me and that I have no reason to take it, but I then irrationally decide to take it. Now, if my decision is a reason, I have turned the second belief from true to false.

14. This is advocated by Schroeder, “Scope of Instrumental Reason,” 344–45.

15. The reasoning of this paragraph would fail if, as the bad effects of the drug diminish, the strength of the alleged reason provided by the decision to take it also diminishes. But it is hard to see why that would be so.

doing it.<sup>16</sup> My decision to perform the action is a faulty decision—it is the wrong response to those reasons. So how can the fact that I do respond by forming that decision succeed in counting in favor of my action, when my decision is the wrong response to the very reasons which, before I made it, counted against the action as well as the decision?

This problem has come to be called the problem of “bootstrapping.”<sup>17</sup> The label is not entirely satisfactory since the complaint is not that decisions pull themselves by their own bootstraps from being decisions there was no reason to make into decisions there is a reason to make nor that actions do this to themselves. The problem is rather that it is hard to see how a decision which is one, incorrect, kind of response to a set of reasons could itself provide a further reason to add to that set, thereby transforming another kind of response—an action—into something that ceases to be incorrect. But since the label has become common, I shall continue to use it. The important point is that the problem it refers to seems serious.

I have concentrated on the cases in which my decision is the wrong response to my reasons since there the problem is at its most obvious. But if there is a problem with thinking that decisions are reasons for action in those cases, this will also be problematic in cases where my decision is the right one. The reasons that count in favor of my action are the reasons for deciding to do it. If decisions and actions are responses to the same reasons, it is hard to see how one kind of response, right or wrong, can add to the reasons for the other.

### III. DECISIONS AND REASONS

What exactly does the bootstrapping objection show? It might be thought to show that our decisions never make a difference to our reasons. But that is false, in three different kinds of cases. More carefully, it might be claimed that decisions themselves can never be reasons. But in the third kind of case, that is false too.

16. “In a case like this”—but not in unusual cases like Kavka’s toxin puzzle, discussed in Gregory S. Kavka, “The Toxin Puzzle,” *Analysis* 43 (1983): 33–36, sec. 4.

17. Michael Bratman sharply identifies the problem and gives it this label in Bratman, “Intention and Means-End Reasoning,” sec. 1—see also Bratman, *Intention, Plans, and Practical Reason*, 24–27. He calls it a problem of “bootstrap rationality,” though: in the terms I am using here, it is a problem about reasons, not rationality. For other acknowledgments of the problem, see John Broome, “Are Intentions Reasons? And How Should We Cope with Incommensurable Values?” in *Practical Rationality and Preference: Essays for David Gauthier*, ed. Christopher W. Morris and Arthur Ripstein (Cambridge: Cambridge University Press, 2001), 98–120, 98–100; Richard Holton, “Rational Resolve,” *Philosophical Review* 113 (2004): 507–35; Raz, “Myth of Instrumental Rationality,” 10; and John Brunero, “Are Intentions Reasons?” *Pacific Philosophical Quarterly* 88 (2007): 424–44, sec. 1.

Decisions are sometimes thought to make the following kind of difference to reasons: although your decision to do A is never itself a reason to do A, it makes a difference to whether you have a reason to take a means to doing A. Having a reason to do A (it is claimed) is not itself enough to give you a reason to take a means to doing A: for that, you must actually have decided to do A. After all, it is argued, you have no reason to take a means if you are not going to do A.<sup>18</sup> However, I find this argument unconvincing. If you are not going to do A, it is equally pointless to decide to do A, but that does not show that there is no reason to decide to do A: you cannot make the reasons go away just by ignoring them. It is normally true that you have no reason (to decide to do A and not to do A). But that does not mean that, when you are not going to do A, you have no reason to decide to do A. If not, then agreeing that you have no reason (to take a means to doing A and not to do A) should not lead us to conclude that, when you are not going to do A, you have no reason to take a means to doing A.

So this alleged difference is not one of the three I have in mind.<sup>19</sup> For the first of those three, consider the costs of reversing a mistaken decision. Suppose I am faced with a choice between A and B. I have a decisive reason to choose A, but I (rationally or irrationally) choose B. After I have done so, the cost of changing course to A may be too great for it to be worthwhile. If I am already standing in the supermarket queue when I realize that the tube of toothpaste in my basket is five cents more expensive than the better rival brand, it might not be worth going back and swapping. My decision to choose what I had most reason not to choose has meant that I now have most reason to do what I have decided.

This does seem to be a first way in which my decision can make a difference to my reasons. But it is not a very interesting way. Here, my decision has caused me to do something which has put me in a new situation whose features provide me with new reasons. My having made the decision is an explanatory reason why I have come to be in a situation in which there are further normative reasons to do the thing I have decided to do. The decision itself is not a normative reason for doing what I have decided to do.<sup>20</sup>

A second kind of case is more interesting. Suppose I am faced (as

18. See, e.g., Brunero, "Are Intentions Reasons?" sec. 4; Raz, "Myth of Instrumental Rationality," 6–7, 23.

19. I do not think this is a convincing argument that decisions do not make this kind of difference to reasons; I am suggesting that we need a better argument for thinking that they do.

20. The same point can apply in situations where there is no mistake—e.g., when one chooses one of a range of long-term activities, none of which is worse than any other. See Raz, "Myth of Instrumental Rationality," 21–22.



I often am) with a range of options, none of which is better than the rest. Either the best ones are equally good, or they are incommensurate—not equally good but not better or worse than each other either. In such situations there can be a decisive reason to choose one of the options, rather than none at all, but not to choose any particular option in preference to the others.

This is another kind of case in which our decisions seem to make a difference to our reasons. Suppose I have a choice between holidaying in Melbourne or Tasmania, and neither is better. Suppose I decide on Tasmania. The next day, you offer me some of your spare tourist maps. Is it not now true that I have a reason to take a map of Tasmania but not a map of Melbourne?

We should say yes, for the following reason. In this situation, two things are true initially:

- a) I have a sufficient reason to choose to go to Melbourne and a sufficient reason to choose to go to Tasmania.
- b) I have a decisive reason not to choose both.

After I have chosen to go to Tasmania, *b* remains true. Therefore, *a* becomes false. I no longer have a sufficient reason to choose to go to Melbourne, given that *b* is true and I have chosen to go to Tasmania.

From this, it does not follow directly that I no longer have a sufficient reason to go to Melbourne. For this, we must make a further supposition. Suppose my decision will play the role that decisions normally play—if I do not change it, it will issue in corresponding action. Next, notice that *a* and *b* are true because two other things are also true initially:

- a') I have a sufficient reason to go to Melbourne and a sufficient reason to go to Tasmania.
- b') I have a decisive reason not to do both.

The reasons reported in *a'* are the same as the reasons reported in *a*, and the reasons reported in *b'* are the same as the reasons reported in *b*. The merits of Melbourne and Tasmania are reasons both to go there and to decide to go; the inconvenience of going to both is a decisive reason not to do that and not to decide to do that. But once I decide to go to Tasmania, then provided my decision will issue in corresponding action, my situation has changed. If I am going to go to Tasmania and *b'* is true, then I do not have a sufficient reason to go to Melbourne. And if I do not have a sufficient reason to go to Melbourne, then I do not have a sufficient reason to take a map of Melbourne.

In this way, my decision has made a difference to what I have a reason to do. The difference it makes, though, is to subtract a sufficient reason for action, not to add one. I initially had a sufficient reason to

go to Melbourne; now I have not. If the first view discussed in this section is false, I also initially had a sufficient reason to get a map of Melbourne; now that is gone too.

This is different from the case of reversing a mistaken decision. There, my decision causes me to produce a further relevant change in my circumstances, and my new circumstances have features that supply me with new reasons. Here, my decision is itself the relevant change in my circumstances. So here, my decision is a normative reason, not just an explanatory one. But it is not a normative reason for action. It is a second-order reason. It is a reason why the attractions of Melbourne no longer give me a sufficient reason either to choose to go to Melbourne or to go there.<sup>21</sup>

My discussion of this example has relied on the supposition that, if I do not change it, my decision to go to Tasmania will issue in my going there. What if that is false? Then I am in an odd situation. Although *a* is now false, *a'* is still true. One of the problems with making ineffectual decisions is that they get us into odd situations of this kind, and that can make them irrational. I shall offer an argument for that assertion later.

But let me turn now to the third example. This concerns collective decisions—decisions that we together take to perform an action of which we collectively are the agent. Sometimes, a group ought to act collectively, but there are different opinions about how we ought to act. When the collective action will importantly affect many of us, and there is time for us to deliberate, it is often the case that we ought to do so collectively, in a way that is appropriately inclusive and regulated in an appropriately impartial way. This is right because it is fair. Moreover, fairness tells us not only to establish such a process and respect it. It also tells us to respect the outcome of the process—the decision that is reached. And our respecting the decision means doing what we have decided to do.

If those claims about fairness are right, then, in this sort of situation, our collective decision provides us with a reason to do what we have collectively decided to do.<sup>22</sup> It is not just that it would be irrational for us collectively to fail to do what we have decided to do. There is a normative reason that counts against it, and this reason can be present even when what we have decided to do was something that we ought

21. In his discussion of these cases, T. M. Scanlon, "Reasons: A Puzzling Duality?" in Wallace et al., *Reason and Value*, 237, also argues that decisions are second-order reasons. However, his conclusion is not that reasons are subtracted, and I disagree with one of the premises of his argument—as I explain in Sec. VII below.

22. I offer a defense of those claims about fairness in Garrett Cullity, "Public Goods and Fairness," *Australasian Journal of Philosophy* 86 (2008): 1–21.

not to have decided to do. Here, incorrect decisions can be reasons for doing what we have incorrectly decided to do.

However, these cases are special. Asked to say why there is a reason in such cases to do what we have decided, we can say more than just that we have decided to do it. We can point to the further features of the situation that constitute reasons of fairness for acting on our decision. The reason we have to do what we have decided is still given by those further features of the situation—features that are antecedent to the decision itself. What this discussion shows, then, is that we need to be careful about the conclusion we draw from the bootstrapping objection, not that we have a refutation of it. The bootstrapping objection does not show that our decisions cannot make a difference to our reasons nor that they cannot be reasons. What it shows is that when there is a reason to do what you have decided, the fact that you have decided to do it cannot exhaust that reason. That you have decided to do something can only be a reason for doing it when there is some further reason for doing what you have decided to do.

This is enough to set us a puzzle. Outside these special circumstances, decisions are not reasons. However, decisions do still seem quite generally to have a bearing on our rationality. This is true when not even the agent thinks there is a sufficient reason to do what he has decided. Suppose I believe there is a decisive reason not to go to Sunday's concert because this will prevent me from preparing properly for the meeting I will be chairing on Monday; I decide to go anyway, but then I am too lazy to book a ticket before they sell out. Here, I am both akratic—I decide to do what I believe there is a decisive reason not to do—and incompetent. I am irrational twice over, in my akratic decision and my instrumental incompetence. But if not even I think there is a sufficient reason for my decision, how do we account for the second kind of irrationality—the irrationality I display in failing to carry it through?

The first task in responding to this puzzle is to find an accurate statement of the bearing of a person's decisions on her rationality. That occupies the next three sections. In the last two sections, I then attempt to explain it.

#### IV. BROOME'S SOLUTION

John Broome has made the influential suggestion that we should understand the relation between decisions and the means to their implementation as a relation not of giving a reason but of "normative requirement."<sup>23</sup> As he originally formulated it, the proposal was this:

23. See John Broome, "Normative Requirements," *Ratio* 12 (1999): 398–419; see also

1. You ought to see to it that (if you intend *E* and believe that your doing *M* is a necessary means to *E*, then you intend *M*).<sup>24</sup>

Broome treats “ought” as expressing what he takes to be the primitive normative concept. Here, that concept governs a material conditional; “see to it that” is inserted as grammatical padding.<sup>25</sup> The claim is that if you if you intend *E* and believe that *M* is the necessary means to *E*, but do not intend *M*, then you are not as you ought to be.

Reasons, for Broome, are potential contributions to explaining oughts. So they are not conceptually primitive: they can be analyzed using the more primitive concepts of ought and explanation.<sup>26</sup> On Broome’s view, it is possible that you ought to *F* without your having a *pro tanto* reason to *F*. But when you do have a *pro tanto* reason to *F*, then its being true that you ought to *F* requires that the reasons in favor of *F* outweigh the reasons against.<sup>27</sup> So ordinarily, when I is true it will be because:

2. You have a decisive reason not (to intend *E*, to believe that your doing *M* is a necessary means to *E*, and not to intend *M*).

Broome’s proposal neatly offers to explain the normative relevance of our decisions to our subsequent actions without turning decisions themselves into reasons. The ‘ought’ in 1 has wide scope, governing the whole conditional, not just its consequent. There is only one way to violate the requirement: retaining the end intention and the means-end belief but lacking the means intention. But there are three ways to satisfy it: not having the end intention, not having the means-end belief, or having both of them together with the means intention. When you ought not to intend *E*, you ought to satisfy the requirement in the first way.

Nonetheless, this proposal is problematic. One problem is that the antecedent of I is too rarely satisfied since it is not often that you believe that a certain means is strictly necessary for the attainment of your intended end.<sup>28</sup> (I need only believe that it is possible that I will receive

---

Broome, “Are Intentions Reasons?” secs. 5 and 6, and “Practical Reasoning,” secs. 4 and 5.

24. This is my own rendering of a claim that Broome formulates slightly differently in various places. See Broome, “Normative Requirements,” 410, and “Are Intentions Reasons?” 110; John Broome, “Normative Practical Reasoning,” *Proceedings of the Aristotelian Society*, suppl. ser., 75 (2001): 175–93, 179, and “Practical Reasoning,” 96.

25. Broome, “Normative Requirements,” 399. Broome treats it as a merely grammatical fact that ‘ought’ does not ordinarily take propositions in English. For an objection to this, see Schroeder, “Scope of Instrumental Reason,” 342–44. Broome has a much fuller discussion of these issues in his unpublished typescript *Reasoning*, chap. 2.

26. Broome, “Reasons,” 34–42.

27. *Ibid.*, secs. 3 and 4.

28. See Broome, “Practical Reasoning,” 97.

a surprise gift of a concert ticket in order to avoid violating 1.) Second, there are counterexamples. You might have a sufficient reason to believe that *M* will occur without your intending it. Or (what is admittedly less likely) you might find yourself in a situation like the following variant of Kavka's "toxin puzzle."<sup>29</sup> An eccentric billionaire offers you a million dollars if you form and sustain, until midnight, the intention to drink a (nonlethal) toxin by midnight. A benevolent pharmacologist then offers you a pill that will induce in you, until midnight, the odd state of intending to drink the toxin, believing that a necessary means to doing so is to pick up the bottle that contains it but not intending to pick up the bottle. You clearly have a decisive reason to take the pill, and this is because you have a decisive reason to be in the odd mental state it produces: a state that wins you a million dollars while preventing you from getting sick. (Assume that you have no other way of achieving that result.)<sup>30</sup> So here, 2 fails and therefore, given Broome's claims about "ought," so does 1.

Situations of this kind are not common. But the example suggests that what we should be looking for is a requirement that tells us about rationality, not about reasons. The odd state produced by the pill is surely a state of irrationality, but one that you have a decisive reason to be in. And this points toward a more general shortcoming in the idea that 1 and 2 can help us to explain what we are seeking to explain. We are trying to describe the impact of decisions on instrumental rationality. But given Broome's claims about "ought," 1 does not do that: what ordinarily makes it true is 2, a requirement concerning reasons, not rationality.<sup>31</sup>

More recently, though, Broome has made another proposal that overcomes these problems. I render it:

3. Rationality requires of you that (if you now intend *E*, believe that *E* will only happen if *M* does, and believe that *M* will only happen if you now intend *M*, then you now intend *M*).<sup>32</sup>

29. Kavka, "Toxin Puzzle." In Kavka's example there is a different requirement—I must, at midnight, intend to drink the toxin tomorrow afternoon—and no pharmacologist.

30. Some will want to resist this example by granting that you have a reason to get yourself into that state but not a reason to be in that state. (This is a distinction suggested, e.g., by Parfit, "Rationality and Reasons," 24.) But Broome himself would, I think, want to reject that distinction, and it looks problematic here. The only reasons you have for getting yourself into that state are the advantages of being in that state, and those advantages might in some circumstances not be reasons for getting yourself into that state (suppose you will be fed the pill in your sleep).

31. On this point, see John Broome, "Does Rationality Give Us Reasons?" *Philosophical Issues* 15 (2005): 321–37, 325, where he says that what he really meant to express by at least some of his original normative requirements were requirements of rationality. (He does also say that in "Normative Requirements," 419.)

32. See John Broome, "Have We Reason to Do as Rationality Requires?—a Comment on Raz," *Journal of Ethics and Social Philosophy* 1 (2005): 1–8 n. 5. A formulation intermediate between 1 and 3 is given in Broome, "Does Rationality Give Us Reasons?" 322.

This presents us with a requirement of rationality, not (or at least, not directly) a claim about reasons or ought.<sup>33</sup> It avoids the counterexamples to 1 and 2. And its antecedent is no longer restricted to cases in which you believe that *M* is strictly necessary for *E*. (Although I believe it is possible that I'll receive a surprise gift of a concert ticket, I also believe that in fact I will only go to the concert if I buy one myself.)

This is a big improvement. However, the usefulness of 3 is still very limited since its antecedent is rarely satisfied for future-directed intentions. If I intend to go to the concert next week, I may believe that I will only go if I book a ticket, but it would be unusual to believe that I will only book a ticket if I now intend to do so.

We need to find a way to account for the irrationality of the procrastinator who never gets around to intending any means to his end. This might seem to be covered as follows. As long as I retain my intention to go to the concert, eventually I will arrive at the point at which the concert is about to start. If I haven't bought a ticket by then, I will then satisfy the antecedent of 3, believing that I'll only get to the concert if I now buy a ticket. If I keep procrastinating and miss the concert, then 3 does correctly tell us that I was irrational in doing so. However, there are two problems with this suggestion. One is that leaving things to the very last moment is itself often irrational, and 3 does not explain why. The other is that many future-directed intentions are not time limited in the same way as this example, where there is a deadline for achieving *E*. Suppose I intend to become a professional opera singer, believe that this will only happen if I have singing lessons, but never get around to intending to have singing lessons. This could surely be irrational. But provided I never form the belief that I will only have singing lessons if I now intend to have them, 3 does not account for my irrationality.

So it seems that we need a further, diachronic requirement of instrumental rationality to account for many cases of instrumental irrationality in relation to the implementation of future-directed intentions. Another, more common kind of procrastination is also often irrational. Suppose I do form the intention to take the means to my end—I intend to have singing lessons—but I never do anything about it. Any variant on 3 will struggle with this kind of case since 3 only gets us as far as intending the means to our ends. Broome rightly emphasizes that rationality cannot require us to get all the way to doing what we intend. If some unforeseen obstacle prevents me from doing what I intend, that

33. For the debate over the normativity of requirements of rationality, see Kolodny, "Why Be Rational?" and the subsequent exchange: John Broome, "Wide or Narrow Scope?" *Mind* 116 (2007): 359–70; and Niko Kolodny, "State or Process Requirements?" *Mind* 116 (2007): 371–85; see also Broome, "Does Rationality Give Us Reasons?"

does not make me irrational.<sup>34</sup> But rationality does seem to require us to go beyond intending to take the means to our ends, at least as far as trying to take them.

A further limitation of 3 is illustrated by the following kind of case. Suppose I intend to write a conference paper.<sup>35</sup> I could start work on it early and produce a good, considered paper without undue stress. Or I could write it in a hurry at the last minute and miss out on those desirable things. So there are those two broadly described means to my end. I believe that I will only write the paper if I do either the first thing or the second. So 3 tells me that retaining my intention to write the paper, retaining my beliefs, and not intending to do either the first thing or the second is irrational. So far so good. But we need to go further. Given how much better the first means is than the second, it is irrational for me not to intend that. Instrumental irrationality is not just a matter of failing to intend what you believe to be the only means that will bring your intended end about; it also includes intending what you believe to be a bad means.

A final point is that 3 tells us nothing about cases in which I lack either of the beliefs it mentions. But I do not escape irrationality by lacking those beliefs. If it should be obvious to me that I am only going to achieve *E* if I intend *M*, but I fail to see this and consequently fail to intend *M*, then that can be irrational too, although it does not violate 3.

In these remarks, I have not presented a reason for thinking that 3 is false.<sup>36</sup> But it seems undesirably limited, in the ways I have just listed. It captures some cases of instrumental irrationality but misses many others. Let me now make a proposal that seeks to overcome these limitations.

34. See Broome, "Does Rationality Give Us Reasons?" 323.

35. I adapt an example from R. Jay Wallace, "Normativity, Commitment, and Instrumental Reason," *Philosophers' Imprint* 1 (2001): 24.

36. The only potential counterexample I can think of depends on the claim that a person can rationally intend *E* while believing *E* will not happen. Suppose I am jailed and credibly assured that I will be released if I betray my best friend. It could be rational to refuse to betray my friend and spend my time pursuing ways of getting out of jail that I know are very unlikely to succeed (filing the bars, perhaps). Here, we might think that I to-intend to get out of jail—that is the goal toward which my action is orchestrated—but believe that I will only in fact get out of jail if I betray my friend. However, the claim on which this depends is controversial: see, e.g., Bratman, *Intention, Plans, and Practical Reason*, 35–42. Notice that although, if the controversial claim is correct, this is a counterexample to 3, it is not a counterexample to 4. Betraying my friend is not a suitable way of getting out of jail.

## V. DIACHRONIC REQUIREMENTS OF INSTRUMENTAL RATIONALITY

The proposal comes in two parts. The first part is this:

- 4-I. Rationality requires of you that (if, during a period through which you have an orderly succession of temporal beliefs,
- i) you intend *E* throughout that period,
  - ii) you believe throughout that period that *E* will only suitably be achieved if, by *t*, you intend some particular means to *E*,<sup>37</sup> and
  - iii) at the end of that period you believe that it is *t*, then at the end of that period you intend what you believe to be a suitable means to *E*).

This has a similar structure to 3 but extends the requirement cross-temporally. Clause ii should be read as covering both *de re* and *de dicto* beliefs—that is, you may or may not have identified a particular means but, in the latter case, you believe you’ll need to do so by *t*. If, intending a Christmas vacation, you have believed since the start of the year that you need to pick a holiday destination by September, and you believe that September has arrived but you have not picked one, then something has gone wrong.

However, there is one way in which what has gone wrong might not be a defect in your instrumental rationality. You might wake up one morning in April, convinced that it is September but realizing that you have not yet picked your Christmas holiday destination: that is not enough to make you instrumentally irrational, hence the condition that you have an orderly succession of temporal beliefs. On a view according to which the rationality of intentions is intimately connected with their role in planning, that should not be surprising. You can suffer a disruption in your temporal beliefs without there being anything amiss with your planning. Notice, however, that the condition is not that your temporal beliefs be true. If you spend the morning having forgotten to reset your watch at the end of daylight saving, your failure to get to the bus stop by the time you thought the bus was due there could still display instrumental irrationality.

37. I offer no definition of ‘means’. These are usually thought of as temporally prior causes of the attainment of ends and therefore distinguished from ways of attaining an end that constitute its attainment—see, e.g., Williams, “Internal and External Reasons,” 104. They can also be distinguished from preconditions and other “facilitating conditions” for the attainment of an end—see Raz, “Myth of Instrumental Rationality,” 5. However, I think 4-I is correct when ‘means’ is interpreted very broadly, as covering a wide variety of ways of attaining an end. In that way, it can serve as a principle of more than just instrumental rationality.



What does it mean to say that *E* will only “suitably” be achieved by your taking some means? It cannot mean that this is the best way of achieving *E*.<sup>38</sup> General epistemic humility might make you wary of ever thinking you’ve identified the very best way of achieving your ends; conversely, you might believe the best way is the best of a set of alternatives, none of which is suitable. It cannot mean that you have a decisive reason to do this, given that you intend *E*. You might be akratic and believe (either rightly or wrongly) that *E* is an end you have a decisive reason to frustrate. And it cannot mean you would have a decisive reason to do this, if you had a decisive reason to intend *E*. The reasons you would have if the world were different in the ways required to make that true need not be reasons that you actually have, and if so your beliefs about those reasons will not bear on your rationality as things actually are.<sup>39</sup>

Instead, we should simply say this. *E* will suitably be achieved by your taking a means whenever taking that means will achieve *E* without frustrating any of your other intended ends. When ‘suitably’ is understood in this way, 4-I expresses the idea that rational intention formation is coordinated with a view to facilitating the co-realization of one’s intentions.

The variable *t* should be read as allowing any temporal reference. For “by *t*,” we can substitute “before I die,” “while the kids are still young,” or “immediately.”<sup>40</sup> Moreover, notice that anyone who successfully meets an instance of 3 will also meet a corresponding instance of 4-I by the end of the period it refers to (which might be a very short one).<sup>41</sup> So the requirement imposed by 3 is subsumed under 4-I.

With 4-I, we have a diachronic requirement that captures the ir-

38. It might be translated as “a sufficiently good means”—but only if that is interpreted in line with the stipulation I go on to give in the text.

39. Suppose I intend to go to a concert. Actually, the music will be bad, and I don’t have a sufficient reason to go. If I did have a sufficient reason to go, the music would be good. And if the music was good, a lot more tickets would have been sold by now, and the only available seats would be at the back of the auditorium. Suppose I believe all of that but akratically decide to go anyway. Surely that does not mean that in order to avoid instrumental irrationality I must buy a seat at the back.

40. This properly allows for considerable vagueness. If I think I need to pick a marriage partner by middle age, the indeterminacy surrounding whether I have reached middle age will be reflected as time passes in a changing degree of belief whether I have done so and a corresponding variation in the extent to which it is irrational for me not to have done so yet. That seems a plausible result.

41. Suppose I form a spur-of-the-moment decision to contact my confessor immediately, I believe that in order to do so I must pick up the phone, and straight away I form the intention to pick up the phone. Then I do satisfy a version of 4-I. The relevant period is the very short period between making the former decision and forming the latter intention; “immediately” is the substitution for “by *t*.”

rationality of a procrastinator who never manages to intend a means to his end. However, we also wanted to cover another form of procrastination, which gets as far as intending a means but no further. The second part of the proposal is needed to cover this.

- 4-T. Rationality requires of you that (if, during a period through which you have an orderly succession of temporal beliefs,
- i) you intend *E* throughout that period,
  - ii) you believe throughout that period that *E* will only suitably be achieved by your taking *M* at or before *t*, and
  - iii) at the end of that period you believe that it is *t* and that you have not taken *M*,
- then you believe that you are taking or trying to take *M*).

This says that once I have identified a means as the only suitable one, it is not enough simply to intend it.<sup>42</sup> I should get beyond that, to the point where it at least seems to me that I am taking or trying to take that means. (“Taking or trying to take” allows for the point that doing something effortlessly need not involve trying to do it.)<sup>43</sup>

The consequent of 4-T is phrased in terms of your belief. Normally, you will have this belief when it is true. But that is not always the case. You might have taken *M* in the past and forgotten about it. It even seems possible to be taking *M* now and not be aware of it.<sup>44</sup> However, in either of those cases, it will be irrational, given i–iii, not to respond to this by acting in a way that does give you that belief. Perhaps there could be odd cases in which there is some systematic impediment to this: whenever you try to take *M* you are prevented from believing it. It seems to me that 4-T is right to classify those as cases of irrationality: I shall say something to justify this judgment later.

Conversely, you might believe you have taken *M* but be wrong about that. Any irrationality here would have to come from a source other than 4-T. There might be something irrational about the belief. But given that you have it, there is no further instrumental irrationality in your failure actually to take *M* or try to take it.

Notice, however, that while 4-T has a doxastic consequent, 4-I does not. This is not because whenever we have an intention we believe it. Suppose that, with my planning for the Tasmania trip well under way, you send me your maps of Melbourne by mistake. I might be halfway

42. If there is more than one suitable means, the disjunction of them can be substituted for *M*.

43. See, e.g., Severin Schroeder, “The Concept of Trying,” *Philosophical Investigations* 24 (2001): 213–27.

44. See Sarah-Jayne Blakemore, Daniel M. Wolpert, and Christopher Frith, “Abnormalities in the Awareness of Action,” *Trends in Cognitive Sciences* 6 (2002): 237–42.

through my e-mail thanking you for them before I remember that it is Tasmania, not Melbourne, that I am intending to visit.<sup>45</sup> But, despite the fact that intentions and beliefs about them can come apart too, 4-I seems to be correctly formulated as it is. If I falsely believe that I intend a suitable means, that is not enough to make me rational. Suppose I irrationally intend what I know is an unsuitable means to my end; I then forget this and, confabulating, form the false belief that I intend a suitable means. The false belief does not make my instrumental irrationality disappear.

Thus, it does seem that 4-I, the requirement on intending, is correctly formulated without a doxastic consequent, whereas 4-T, the requirement on taking or trying, requires the doxastic formulation. Intention is a dispositional state, and the (wide-scope) requirement on me concerns putting myself in that state. Taking and trying to take a means are events, and the requirement on me concerns satisfying myself that those events have occurred or are occurring.

I shall refer to the conjunction of 4-I and 4-T as requirement 4. In 4, we have a diachronic requirement that covers the same cases as 3 but now captures the irrationality of the two forms of procrastination that 3 misses. However, I also drew attention to two other shortcomings of 3. It is not so clear that they have been met.

In encompassing your beliefs about suitable means, 4 does capture cases of instrumental irrationality that 3 misses. Specifically, it captures cases in which I can see that a means to one of my ends, although not the only way of achieving that end, is the only way in which it will be achieved compatibly with my other ends. Thus, as long as I intend not only to write a conference paper but also to avoid undue stress, 4 correctly implies that leaving it until the eve of the conference is irrational. That is a step in the right direction. However, it still seems to miss some obvious-looking kinds of instrumental irrationality—at least the following four. I might rule out some means to *E* on grounds other than their frustrating my other ends. For example, not stealing my lunch from the campus café does not seem to be an end of mine: it is a side constraint that I think I have a decisive reason to meet in pursuing my end of having lunch. Or consider the inefficient use of resources, in a way that will leave me able to achieve all the ends I currently intend but with fewer resources to serve the intentions I will go on to form in the future. Hiring a taxi each day to take me to work and wait with the meter running until 5:00 p.m. is an obviously bad means of arranging my transport to and from work: doing this could be irrational, without

45. Here, I follow Bratman, who gives similar examples in Bratman, “Intention, Belief, and Instrumental Rationality,” sec. 2, and Michael E. Bratman, “Intention, Belief, Practical, Theoretical,” in Robertson, Skorupski, and Timmerman, *Spheres of Reason*, sec. 4.

frustrating any of my actual ends. A third kind of instrumental irrationality that seems to have been missed occurs where I can see that the means I am taking, while not preventing outright the co-realization of my intentions, makes this unlikely. Suppose that in order to avoid execution, I must apply for a stay of execution: if I know that leaving my application until the day before the execution date will reduce its chances of success, that could surely be irrational. Moreover, to accentuate this problem (a fourth case), suppose I have decided to watch the entire TV coverage of the Olympics in the weeks leading up to my execution date—now I have saddled myself with a set of intentions that can only be co-realized by leaving the application to the last day, but surely that does not remove the irrationality.

The other worry I raised about 3 concerned people who lack the beliefs mentioned in its antecedent. That also applies to 4: you might lack either of the beliefs mentioned in ii and iii. If you fail to form any views about the suitable achievement of your intended ends when this should have been obvious to you, your inaction still seems irrational, but 4 fails to say so. Indeed, the wide-scope formulation of 4 means that you will routinely satisfy it if, whenever you recognize that you satisfy the consequent, you simultaneously drop either of the beliefs mentioned in ii or iii. But that does not seem to be a way of being rational.<sup>46</sup>

The doxastic consequent of 4-T provokes a related worry. It was complained against 3 that it gets us only as far as intending our ends. It might seem worse that 4-T only gets us as far as believing that we are acting to implement them. This might not seem practical enough to serve as a requirement of instrumental rationality. Suppose you have a problem with procrastination, and I offer you a pill to cure you of your irrationality. If the pill does not stop your procrastination, but simply causes you to believe that you have taken *M* whenever you believe it is *t*, my description of the pill sounds misleading.<sup>47</sup>

In order to defend 4 as a significant improvement over 3 in identifying the core cases of instrumental irrationality, we need to respond

46. This is a common theme among critics of Broome-style wide-scope requirements: it is often expressed as a worry about the “symmetry” of the different possible ways of complying with them. See Schroeder, “Scope of Instrumental Reason,” 339; and Raz, “Myth of Instrumental Rationality,” sec. 2; see also, for discussion, T. M. Scanlon, “Structural Irrationality,” in *Common Minds: Themes from the Philosophy of Philip Pettit*, ed. Geoffrey Brennan, Robery Goodin, Frank Jackson, and Michael Smith (New York: Oxford University Press, 2007), 84–103.

47. A further worry: suppose you have tried some means to your end but failed. It still seems irrational for you to retain your intention without making any further effort to implement it. Does 4-T fail to capture the irrationality of this kind of case? I think not. For now you have entered a new stretch of time—the one commencing with the failure of your first effort—and 4-T can be reapplied.

to these outstanding worries. But it would be better to go further than this and explain why violations of 4 are irrational.

In support of 3, a “cognitivist” strategy of explanation has been intensively explored by several recent writers.<sup>48</sup> This seeks to explain the practical requirement of instrumental rationality by reference to the connection between intention and belief. As we have noted, it seems questionable whether that connection is tight enough to allow us to infer from your intending *E* that you believe *E*. But perhaps, as Broome suggests, from your belief that you intend *E* we can infer that you believe *E*.<sup>49</sup> Or perhaps, as Jay Wallace suggests, from your intending *E* we can infer that you believe that *E* is possible.<sup>50</sup> If so, this opens up the prospect of an argument of the following general form. If you believe *E* and believe that *E* requires *M* but do not believe *M*, then you have incoherent beliefs. Arguably, this incoherence in belief is at the heart of your instrumental irrationality. If you are committed to an end that you know requires steps that you do not see yourself as taking, you cannot see your own thought and conduct as making sense.

However, whatever the attractions of this strategy as an explanation of 3, it looks questionable whether it will help us with 4. Having broadened our requirements of instrumental rationality to capture cases where you believe that *E* cannot suitably happen without *M*, not just those where you believe *E* requires *M*, we seem to have cut off the prospect of an argument of this kind.

So a would-be defender of 4 has two tasks: dealing with the cases of instrumental irrationality that it seems to miss and explaining why it is a requirement of rationality. The rest of the article tackles these in turn.

## VI. FURTHER REQUIREMENTS OF RATIONALITY

We might consider trying to revise 4 to cover the kinds of irrationality it misses. But it is simpler, and therefore better, not to do that. Instead, following Broome’s lead, we should recognize that there are various

48. Broome is one leading proponent of this cognitivist strategy: see esp. Broome, “Practical Reasoning,” secs. 1–3, and “Unity of Reasoning?” secs. 7–11. Others include Gilbert Harman, “Practical Reasoning,” in *Reasoning, Meaning, and Mind* (New York: Oxford University Press, 1999), 46–74; Wallace, “Normativity, Commitment, and Instrumental Reason”; and Kieran Setiya, “Cognitivism about Instrumental Reason,” *Ethics* 117 (2007): 649–73. For criticism, see Bratman, “Intention, Belief, Practical, Theoretical,” and “Intention, Belief, and Instrumental Rationality”; and John Brunero, “Two Approaches to Instrumental Rationality and Belief Consistency,” *Journal of Ethics and Social Philosophy* 1 (2005): 1–20.

49. Broome, “Unity of Reasoning?” sec. 7.

50. Wallace, “Normativity, Commitment, and Instrumental Reason,” 20.

other requirements of rationality and that violating any of them is sufficient to make you irrational.<sup>51</sup>

One that Broome emphasizes is a requirement of *enkrasia*—the requirement which *akrasia* violates. A formulation of this requirement close to Broome’s own is:<sup>52</sup>

- 5-P. Rationality requires of you that (if you believe you have a decisive reason to *X*, you believe that you will *X* only if you intend to *X*, and you have had time to form the intention to *X*, then you intend to *X*).

The clause, “you believe that you will *X* only if you intend to *X*” is reminiscent of 3: it is introduced for the same reason. There are many things that you believe you have a decisive reason not to do but that you need not form an intention not to do. Adding the restriction, though, gives us a similar problem to 3. If we adopt the same synchronic reading as 3, the second belief is that you will *X* only if you now intend to *X*. But then, given how rare that belief is, 5-P seems to account for only a small subset of the cases where *akrasia* is irrational.

One response to this problem would be to formulate a diachronic *enkrasia* requirement, modeled on 4. But a different suggestion handles many of the relevant cases well. If we think of 5-P as a requirement of positive *enkrasia*, we can supplement it with the following requirement of negative *enkrasia*:

- 5-N. Rationality requires of you that (if you believe you have a decisive reason to *X*, then you do not intend not to *X*).

This allows us to account for the irrationality of intending means that you believe you have a decisive reason not to take. In doing so, it covers one of the kinds of irrationality missed by 4—the kind exemplified by stealing from the campus café. Given that not stealing is not one of my ends, stealing is not unsuitable, in the sense employed in 4, but it would be irrational as a violation of (negative) *enkrasia*.

51. Even if we could list them all, meeting all of them would not be sufficient to make you rational. Your meeting them would need to be guided in the right way. The crucial question of how to explain this guidance condition is not one I try to answer here.

52. Compare Broome, “Unity of Reasoning?” sec. 2. Broome’s own favored *enkrasia* requirement is formulated in terms of beliefs about ought rather than decisive reasons. The other difference is my addition of the clause, “and you have had time to form the intention to *X*.” If getting from a belief that you have a decisive reason to *X* to an intention to *X* is a transition between different states, it will take time—although perhaps not very long. Without the addition of this clause, 5-P would carry the false implication that at the first moment you hold the two beliefs, but before the process of forming the intention is complete, the state you are in is a state of irrationality. As Broome points out, the occurrences of ‘you’ within the scope of the beliefs have to be read as “reflexive” pronouns—i.e., you must know who you are.

Another concern was that 4 does not capture the irrationality of those who lack beliefs about the suitable means to their ends, when this should be obvious to them. Again, we should handle this not by modifying 4 but by noticing a further group of requirements of rationality—requirements that our beliefs are responsive to the evidence for their truth. One such requirement of rationality on beliefs appears to be this:

6. If you have no clear evidence that  $p$  is true, and no other good reason for believing  $p$ , you are rationally required not to believe  $p$ .<sup>53</sup>

What about the things I am rationally required to believe? Some care is required here. From the existence of clearly decisive evidence for a proposition, it does not follow that I am rationally required to believe it. As Gilbert Harman points out, it is rational for us not to clutter our minds with all the beliefs for which we have clearly decisive evidence.<sup>54</sup> But if a proposition is both obvious and important, then overlooking the evidence for it can be irrational. There seem to be several different kinds of importance for which this is true, and I do not propose to attempt a summary of them. However, I suggest that two of them can be captured as follows:

7. If you have clearly decisive evidence that you have a decisive reason to  $X$ , then rationality requires of you that you believe you have a decisive reason to  $X$ .
8. If you intend  $E$  and you have clearly decisive evidence that  $M$  is the most suitable means to  $E$ , then rationality requires of you that you believe that  $M$  is the most suitable means to  $E$ .

These are scalar rather than all-or-nothing requirements. The degree of irrationality depends on how clearly decisive the evidence is in either case, as well as on how decisive the reason is (7), or how suitable the means is (8). (If intentions can be satisfied more or less fully, then it is possible that more than one means of achieving an end is suitable but that one of those means allows the agent's intentions to be satisfied more fully and is in that way more suitable than the others.) They are best thought of as requirement forms—frameworks that can be filled in with different rival accounts of the kind of evidence that is relevant to the rationality of belief.

Given these requirements of evidence responsiveness on beliefs, we can account for further cases of irrationality that 4 misses. If, as a would-

53. If there are no good reasons for a belief that are not pieces of evidence for its truth, then the second clause is redundant. I include it to sidestep the controversy over that.

54. Gilbert Harman, *Change in View* (Cambridge, MA: MIT Press, 1986), 12.

be opera singer, I never get as far as believing that the only suitable way to achieve this is by having singing lessons, then my irrationality lies in overlooking clearly decisive evidence that this is true. (And if I do not have such evidence, it does become hard to see how I am irrational.) If, in spending a large part of my income on making taxis wait for me, I am depriving myself of resources that I will need in the future, this could be irrational in one (or more) of three ways. I might now intend to maximize the resources available to me in future: if so, the inefficiency makes the means unsuitable, so if I know that this is true, I am violating 4, and if I do not, but there is clearly decisive evidence that this is true, I am violating 8. I might believe that my action will waste resources that there is a decisive reason for me to conserve, in which case my irrationality is accounted for by 5-N. Or although I do not believe that, I might have clearly decisive evidence that that is true, in which case my irrationality is accounted for by 7. If I meet none of those conditions—I do not believe I am frustrating any intention of mine nor that I am acting contrary to any decisive reason, nor is there clearly decisive evidence that either of those things is true—then it does become hard to see how my inefficiency is irrational. Perhaps there is some further requirement of rationality that I am violating, but it is hard to see what it would be.

Likewise with the death row inmate who spends his days watching TV and leaves his application for a stay of execution until the last moment. The irrationality here lies either in ignoring the clearly decisive evidence that he has a decisive reason to make his application immediately instead of watching TV—in ignoring this he violates 7—or in forming this belief but akratically failing to form the appropriate intention and indeed intending the opposite, thereby violating both 5-P and 5-N. (Again, if there is no clearly decisive evidence of this, then it is hard to see how there is, after all, any irrationality.)

The other worry about 4 concerns the procrastinator who satisfies 4-T by taking a pill that manipulates her beliefs. This kind of irrationality lies in the violation of 6. If there is no clear evidence that she has taken *M* and she believes she has, then the belief she has induced in herself is irrational. (If there is clear evidence of this, then even if she has not done anything to further her intended ends, it is once more hard to see how this is irrational.)

The details of the formulations I have given to 5–8 could be debated. In particular, there is a debate over whether 5-P, formulated here in Broome's wide-scope way, should be given a stronger narrow-scope formulation—that debate will apply also to 5-N.<sup>55</sup> (Requirements 6–8

55. See the exchange: Broome, "Wide or Narrow Scope?"; and Kolodny, "State or



clearly need to take the narrow-scope form I have given them: you cannot satisfy the rational requirement in these cases by removing the evidence.) However, what is important here is to see the plausibility of a more general point. The cases of irrationality that 4 may seem to have overlooked are captured by other more general requirements of rationality. What 4 gives us is a requirement that governs the output of our thinking about the means to our ends: these other requirements of rationality govern the inputs to that thinking.<sup>56</sup>

I am not going to keep listing requirements of rationality: I mention these in order to show how we can capture the kinds of decision-relevant irrationality which 4 on its own does not account for. However, let me mention one other candidate requirement. At the end of Section I, I mentioned the view that rationality involves a responsiveness not just to apparent reasons but directly to reasons themselves—a view supported by appealing to Huckleberry Finn cases. The examples are not decisive—it might be responded that we should admire Huck for his compassion, not his rationality—but a suggestion that deserves consideration is this. At best, it is a responsiveness to evidence of his reasons that we should admire in Huck. (Had Jim seemed exactly like a fugitive slave when he actually wasn't, we should evaluate Huck's response in the same way.) The best way to put this suggestion, then, is as the proposal that we add to our list of requirements of rationality:

9. If there is clearly decisive evidence that you have a decisive reason to *X* now, then rationality requires of you that you believe you are now either *X*ing or trying to *X*.

Compare Huck to someone who has the same beliefs as him and turns Jim in. Huck and the slave returner both violate 7: they have the wrong beliefs about reasons. In addition, Huck violates 5-N and possibly also 5-P: he is akratic. But the respect in which he can be claimed to be more rational than the slave returner is that he meets 9 while the slave returner violates it.

## VII. A STANDARD-FIXING ACCOUNT OF RATIONALITY

I have made a case for the intuitive plausibility of 4–8. But I now want to raise a deeper question. Why are these all part of what rationality requires? What is rationality that explains why it should contain these

---

Process Requirements?" Schroeder, "Scope of Instrumental Reason," goes as far as defending a narrow-scope treatment of requirements of instrumental rationality.

56. Another way to avoid being classified by 4 as irrational is simply to stop intending *E*. Doing that when nothing in your situation has changed could be irrational. So it looks as though there is also an intention-constancy requirement of rationality on the inputs to your thinking about the means to your ends.

components? Is 9 really a requirement of rationality too? What would settle that?

A fruitful approach to these questions anchors the answer to them in an account of why rationality matters. When you evaluate someone's thought or conduct as irrational, what is the force of that evaluation? In pressing this question, we should pay special attention to contexts where judgments about reasons and rationality diverge. Suppose that, when I believed that I had a decisive reason not to go to Sunday's concert, I was right. Then, when my incompetence means I miss the concert and end up preparing for the meeting instead, I have done what I had a decisive reason to do. What is the force of the judgment that the way I arrived at that outcome was irrational?

Apparent reasons accounts offer the following answer. When you judge that my thought or action is irrational, you are judging that it is ruled out by the reasons I think I have. The force of rationality judgments, on this view, is the apparent force of the reasons that person who is the object of the assessment accepts himself as having. Niko Kolodny aptly dubs this a "Transparency Account" of the force of rational requirements.<sup>57</sup>

However, this view struggles to explain the instrumental irrationality that is present in some cases of *akrasia*. To handle such cases, it must claim, as T. M. Scanlon does, that it is not possible to adopt an end without thinking one has a reason to take the means to it: "To adopt a goal is to set oneself to pursue it, and to take the fact that certain actions would be means to that end as counting in favor of those actions."<sup>58</sup> If this is right, then even when I think I have no reason to achieve an end of mine, I will still think I have a reason to take the means to it. If so, my instrumental irrationality can still be described, even here, as contravening my own judgments about reasons. But this is unconvincing. Suppose I, the agent, believe that whenever there is no reason for me to have an end, there is no reason for me to take the means to it. That thought at least seems possible. I might still figure out means to my *akratic* ends, asking myself which means are compatible with the co-realization of my intentions (the things I have, for better or worse, set myself to do) without thinking of them as supported by reasons. And if I can see that the only way to reach my intended end, compatibly with my other intentions, is to take a certain means, my retaining the

57. Kolodny, "Why Be Rational?" 513, 557–60.

58. Scanlon, "Reasons: A Puzzling Duality?" 238. Despite other disagreements with Scanlon, Brunero, "Are Intentions Reasons?" agrees with this claim at 436. Compare also Raz, "Myth of Instrumental Rationality," 15: "I will here confine the term 'instrumental irrationality' to the malfunctioning of our capacity to react properly to perceived reasons that manifests itself in failure to pursue available means to our ends."

intentions to achieve those ends without intending to take that means would still seem instrumentally irrational.<sup>59</sup>

A second suggestion is that when beliefs and intentions are irrational, it is because they frustrate the constitutive aims of those states. In this vein, it is argued that it is constitutive of a state's being a belief that it represents its content as true.<sup>60</sup> But beliefs are not fitted to doing this unless they are governed by requirements that include sensitivity to the evidence of the truth of their contents. It might then be argued, in the spirit of David Velleman's recent work, that we can derive requirements of rationality on intentions from those on beliefs.<sup>61</sup> Or it might be argued, with Bratman, that intentions have their own constitutive aim and that this directly generates a parallel set of requirements of rationality. What makes intentions the kind of states they are is the distinctive role they play in our thought and practice: a role of facilitating the intra- and interpersonal coordination of action and thereby enabling the effective exercise of planning agency, and intentions are not fitted to playing this role unless they are governed by requirements such as 4.<sup>62</sup>

However, this suggestion faces some large problems too. One is that there is an important difference between the claims it is plausible to make about beliefs and intentions. It is plausible to claim that a particular state of mine cannot be a belief unless it represents its content as true. But it is not plausible to claim that a particular state of mine cannot be an intention unless it "aims at its realization in coordination with one's overall system of intentions."<sup>63</sup> It is plausible that it cannot be an intention unless it belongs to a kind of mental state the point of having which is to enable such coordination, but that is a different claim.

There are replies to such objections. However, the point of men-

59. It might seem better to resort to the weaker claim that "adopting a goal gives rise to a difference in what an agent must, in so far as he has that goal and is not irrational, see as reasons" (Scanlon, "Reasons: A Puzzling Duality?" 235). This claims that failing to see oneself as having reasons to take what one believes to be the means to one's intended ends is not impossible but irrational. But now this claim is too weak to explain the instrumental irrationality of the akratic incompetent. It implies that I am irrational in not seeing my means as reasons. But it does not explain the difference between two akratic agents, both of whom deny that they have reasons to take the suitable means, but one of whom is instrumentally incompetent while the other is not.

60. See J. David Velleman, "On the Aim of Belief," in *The Possibility of Practical Reason* (Oxford: Clarendon, 2000), 244–81; Nishi Shah, "How Truth Governs Belief," *Philosophical Review* 112 (2003): 447–82; Nishi Shah and David Velleman, "Doxastic Deliberation," *Philosophical Review* 114 (2005): 497–534.

61. See J. David Velleman, "What Good Is a Will?" in *Action in Context*, ed. Anton Leist and Holger Baumann (Berlin: de Gruyter, 2007), 193–215.

62. Bratman, "Intention, Belief, and Instrumental Rationality," sec. 3.

63. *Ibid.*

tioning the transparency and constitutive aim accounts is not to show that they can be swiftly refuted but to introduce some important insights to which any adequate alternative will need to do justice. These two views are trying to give expression to the idea that irrationality is a kind of self-frustration, and there seems to be something importantly right about that. Transparency accounts see evaluation of my rationality as evaluation of me from my own point of view.<sup>64</sup> This is suggestive, although our earlier discussion suggests we need to think of that point of view more broadly—as encompassing not only my beliefs about reasons but also my intentions, my beliefs about the compatibility of my actions with those intentions, and the evidence available to me. What remains to be explained is what the elements of a person’s point of view, thought of in this broader way, have in common, in virtue of which they all bear on a single topic—rationality. There also seems to be something importantly right about the suggestion of constitutive aim accounts that we can explain the requirements of rationality that govern states such as belief and intention by reference to the distinctive point of those states.

I think we can retain and combine these insights in a third, and to my mind more plausible, view. This third view explains the force of evaluations of irrationality by reference to what it sees as the practical function of such evaluations—the function of supporting dispositions that, in general, it is important for us to have. If, through instrumental incompetence, I end up spending Sunday evening preparing for the meeting, I have ended up doing what I had a decisive reason to do. But I have arrived at that happy outcome through failing to exercise an important disposition—a disposition to take what I believe to be the suitable means to my intended ends.<sup>65</sup> The general importance of this disposition gives us, collectively, a decisive reason to adopt evaluative standards that reinforce it. Standards of rationality and irrationality are evaluative standards we collectively have a good reason to adopt because they reinforce important dispositions.

This provides an attractive model for accounting for evaluative standards of many different kinds, by appealing to the different dispositions we have good reasons to support. There are many contexts in which we make negative evaluations of a person’s performance in only accidentally hitting the proper target of an activity, and an account of this kind deals with them neatly. We see this in many evaluations of technical skill. My clumsy miskick might score a goal: although it has achieved the result toward which my training is directed, there is still a good

64. Kolodny, “Why Be Rational?” 557; Scanlon, “Structural Irrationality,” 7.

65. I take “failing to exercise a disposition” to span failing to have it at all and having it but not exercising it.

reason for us to evaluate it as clumsy since it is a failure to exercise dispositions there are reasons for me to acquire as a football player. If acquiring those dispositions matters, there are good reasons for the footballing community to adopt an evaluative standard that supports them.

Among the various evaluative standards that can be accounted for along these general lines, we distinguish one such evaluative standard from the others by specifying the disposition(s) it supports. To pick out the standards for the evaluation of a person's rationality, we should say this. The dispositions reinforced by assessments of rationality are those which we exercise in thinking (or not thinking) about reasons. There are several of these—I described some of them in Sections V and VI above—and they give us more specific evaluative standards, particular ways of being rational or irrational. Associated with them is a distinctive vocabulary of appraisal of each other as “weak willed,” “gullible,” “resolute,” “procrastinating,” “efficient,” “impulsive,” and so on. Rationality, then, is a family of evaluative standards: what ties them together is their connection to the dispositions it is important for us to exercise in our thinking (or not thinking) about reasons.

We need to include “or not thinking” in order to cover various ways in which thinking about one's reasons can be irrational. Sometimes, thinking gets in the way of acting: it is irrational to stand in the goalmouth pondering your reasons as the opposition striker bears down on you. And sometimes it can be rational to form and stick to a resolution to do something when, if you deliberated about your reasons at the time of acting, you would succumb to temptation and judge them differently.<sup>66</sup> It may seem that all dispositions are exercised either in thinking or not thinking about reasons, so the addition of “or not thinking” prevents the account from picking out a particular kind of disposition at all. But that is not true. If I am an effective goalkeeper (rather than a goalmouth philosopher), then the dispositions that constitute my technical skill in anticipating the striker's movements may be exercised while

66. Notice the way in which my view differs from the “two-tier” account of such cases recommended by Holton, “Rational Resolve.” On that account, the rationality of forming and holding a resolution confers rationality on its exercise. My different suggestion is that the dispositions we have decisive reasons to support fix the standards for rationality. To illustrate the difference, suppose you are convincingly taught from a young age that the natural selectionist theories you will be taught in school are the work of the devil and that you should shut your ears to the seductive arguments that will be used in their favor. It might then be rational for you to resolve not to listen to them. As I read it, a two-tier account carries the implication that it is therefore rational to go ahead and ignore the evidence for natural selection. On my view, this is irrational since your disposition, even if it is rationally acquired, is one which violates the standards for evidence sensitivity that we have decisive reasons to uphold.

not thinking about my reasons, but they are not exercised in not thinking about them. If I do have a disposition not to ponder my reasons in situations like this, then that disposition is exercised in not thinking about my reasons, and in that respect I can be assessed as rational, but exercising my technical skill is neither rational nor irrational.

The view I am proposing, then, makes the following claim: standard *S* is part of the set of standards of rationality only if:

- a) the disposition to conform with *S* is important because of its exercise in thinking (or not thinking) about one's reasons, and
- b) the importance of this disposition gives us collectively a decisive reason to support it by employing *S* as an evaluative standard.

Any account of rationality that makes this claim, I call a "standard-fixing account." It places two necessary conditions on a standard being part of rationality. One kind of standard-fixing account—the broadest—claims that these two conditions are also jointly sufficient. Other versions will seek to add further conditions to these two. However, rather than exploring the differences between rival versions of this view, let me instead draw attention to two attractive features possessed by any standard-fixing account.

A first significant attraction is that, when a disposition is important because of its exercise in thinking about one's reasons, we can still have good reasons to support that disposition by evaluating a person's conformity with it when she is not thinking about her reasons. This is the key to accounting for instrumental irrationality in cases of *akrasia*. Dispositions toward the co-realization of one's intentions—dispositions of the kind spelled out in 4—are exercised by *enkratic* agents in thinking about their reasons. *Akratic* agents can also exercise those same dispositions, without doing so in thinking about their reasons (*contra Scanlon*). The importance of supporting these dispositions comes from the role they play in thinking about our reasons. But we support them by making evaluations of instrumental rationality and irrationality wherever we find it, in agents who are *akratic*, *enkratic*, or neither. The force which pointing out the instrumental irrationality of an *akratic* agent usually has is that it draws attention to a failure with respect to a standard which this agent still has a decisive reason to be disposed to meet because of its importance in thinking about reasons, even if she does not have a sufficient reason to meet that standard on this occasion. It should matter to me that I have only ended up preparing for the meeting through irrationality because it should matter to me that I repair the defective disposition that has caused me to do so.

Notice that this does not mean that the objects of evaluation as rational or irrational must be dispositions. (It does not preclude this either. There is, after all, good reason to reinforce our dispositions by

subjecting them to evaluation, and we do so.) We evaluate our performance directly as rational or irrational. But the standards for this evaluation are fixed by reference to the dispositions there is a good reason for us to support.

The second attractive feature is this. Normally, the force of an evaluation of my thought or conduct as irrational is something it is appropriate for me to feel since, normally, the standard I am failing to meet is one that I have decisive reasons to be disposed to meet. However, that is not always true. I might be in a situation like the earlier variant of the toxin puzzle. Or I might be in a “Schelling case,” in which acquiring irrational dispositions is the best way to render myself invulnerable to others’ threats.<sup>67</sup> It is still true that there is a good reason for us—and that may include me<sup>68</sup>—to evaluate my thought and conduct as defective, even if it is a defect there is a decisive reason for me to have in these circumstances. Evaluating me by reference to this standard is still part of a practice that gets its point from the way in which a disposition not to have such defects is generally important. After all, Schelling cases are parasitic on ordinary rationality. They are cases in which the suitable means to your end is to dispose yourself in opposition to the standards of rationality that generally apply and thereby dissuade others from harming you. Indeed, it is only because it is a defect that you have a good reason to have the defective disposition.

It is a corollary of this that if we all lived in a toxin puzzle or Schelling case world, and collectively ceased to have good reasons to have dispositions toward instrumental rationality, there would cease to be any point in evaluating anyone as instrumentally irrational. But that sounds right. (It still sounds right to me to say that that would be a world in which frustrating one’s intentions is irrational. The reference of ‘rationality’ is fixed by the standards to which there are actually good reasons to hold each other accountable. But that is merely a report of linguistic intuition rather than a normative point.)<sup>69</sup>

Standard-fixing accounts of rationality differ from both transparency and constitutive aim accounts. In their treatment of the instrumental irrationality of akratic agents, they differ from transparency accounts. And standard-fixing accounts make no claims about the con-

67. See Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960).

68. Perhaps I will perform better in a Schelling case if I am oblivious to my own irrationality. But perhaps not.

69. Another corollary is that our practice of evaluation in terms of rationality and irrationality only makes sense on the assumption that we are similar enough to give us reasons to have the same dispositions in thinking about reasons. But that sounds correct too.

stitutive aims of beliefs and intentions.<sup>70</sup> However, they do draw on the thought that such states have a point, in the sense that we have decisive reasons to have such states and consequently to govern them in particular ways: to govern our beliefs with dispositions to evidence responsiveness and to govern our intentions with dispositions to co-realizability. Having a disposition to evidence responsiveness does help us to form true beliefs, which we have both instrumental and noninstrumental reasons to form. And being able to commit ourselves to ends and to coordinate those commitments with each other is important in enabling us to achieve the things we have good reasons to achieve. Just as importantly, our intentions help us to deal with situations in which we have sufficient reasons to do each of several things but cannot do all of them. Committing ourselves to some of them and following through the commitment allows us to achieve more of what we have good reasons to achieve. As agents on whose actions many different reasons bear, our plans and the intentions that structure them form some of our most important responses to those reasons. Moreover, just as it is plausible to think that we have noninstrumental reasons to have true beliefs, it is also plausible to think that we have noninstrumental as well as instrumental reasons to be effective agents.

This equips a standard-fixing account to give a plausible explanation of why rationality matters—when it does. The dispositions we support through evaluations of rationality are crucial to our practical and theoretical functioning—most of the time. So normally there is a kind of self-frustration in my irrationality: I am failing to comply with standards that I have decisive reasons to be disposed to meet. Rationality matters because these dispositions matter. This is why it should matter to us to correct our irrationality and to participate in an evaluative practice that reinforces those dispositions. But that is consistent with the possibility of unusual cases in which irrationality is actually helpful.

Notice also how standard-fixing accounts depict the relationship between rationality and reasons. Reasons enter such accounts at two points: (a) the dispositions that fix the relevant standards are dispositions concerning our reasons, and (b) they are dispositions we collectively have decisive reasons to support. This means that standard-fixing accounts are not a kind of proper functioning account since they explain rationality by reference to reasons, rather than vice versa. But nor are

70. Nor are they inconsistent with such claims. If beliefs and intentions have constitutive aims, then it is plausible that our reasons to have those states will give us reasons to be disposed to meet the standards that fit them to being the states they are.



they a kind of responsiveness account since they deny that rationality is exhausted by responsiveness to reasons or apparent reasons.<sup>71</sup>

As I see it, standard-fixing accounts of rationality have several further attractions. One is that they offer us a unified treatment of both practical rationality and the rationality of belief. Notice that irrationality in akratic action has its counterpart in irrational belief formation, and it is handled in the same way. Suppose I form several beliefs in defiance of the evidence. Suppose I even recognize this: my beliefs are not responsive to apparent reasons. I might then exhibit a further irrationality in retaining those beliefs while denying what they obviously entail. This further irrationality can be neatly explained in the same way as before. It is a failure to exercise a disposition which we have a decisive reason to have because of its importance in governing the beliefs which are responsive to our reasons. Defects of this kind are still defects even when they are exhibited in beliefs which are not responsive to our reasons.

Standard-fixing accounts allow for a range of further plausible judgments about rationality and irrationality. They allow for a distinction between reasons to evaluate my thought and conduct as defective in relation to standards of rationality and reasons to criticize or blame me for those defects. (They imply that insanity is a rational defect, without implying that the insane should be criticized for their irrationality.) And they also allow for a plausible context-specific calibration of the dispositions which fix the standards for assessments of rationality. For example, what disposition to question the instructions of our superiors should we support? The answer varies from one context to another: it varies greatly, for example, between the context of a philosophy class

71. My remarks have some close affinities with those made by other writers. For example, Bratman writes: "In assessing the rationality of an agent for some intention or intentional action our concern is to determine the extent to which the agent has come up to relevant standards of rational agency. A failure on the agent's part to come up to such standards makes that agent guilty of a form of criticizable irrationality. In reaching such assessments our concern is with the actual processes that lead to the intention and action and with the underlying habits, dispositions and patterns of thinking and reasoning which are manifested in those processes. Our concern is with the extent to which these processes—and the underlying habits, dispositions, and patterns they manifest—come up to appropriate standards of rationality" (Bratman, *Intention, Plans, and Practical Reason*, 51). Similarly Raz, after quoting this passage with approval, says that irrationality "can consist in *faulty functioning*, that is in ways of thinking and of forming beliefs or intentions, and so on, which do not conform to standards of rationality" (Raz, "Myth of Instrumental Rationality," 15). I agree with these remarks. However, they raise two questions that I am trying to answer in the text: (1) Which processes, and which kinds of functioning, are candidates for assessment as rational or irrational? and (2) What fixes the standard of assessment? The answer I offer presents a picture of the relationship between reasons and rationality. It is not clear to me that either Bratman or Raz is doing that, so I hesitate to describe them as offering a standard-fixing theory of that relationship.

and a battlefield, where a disposition to obey instructions unquestioningly may save one's own life and those of one's comrades.<sup>72</sup>

However, this does not imply that whatever disposition you have decisive reasons to have in thinking about your reasons in a given context amounts to what is rational in that context. In Schelling cases, you may have decisive reasons to acquire dispositions to behave wildly: the wild behavior is still irrational. Someone who is adept at detecting Schelling cases and getting himself into a state of threat invulnerability—a skilled gangster or statesman, perhaps—might exhibit rationality (in the form of “shrewdness”) in getting into that state. But his behavior in that state, violating standards such as 4–8, is still behavior we have good reasons to identify as rationally defective because of the importance that dispositions to comply with such standards ordinarily have.

Most importantly, a standard-fixing account of rationality coheres with the claims I made earlier about the specific requirements of rationality that bear on our decisions. We have decisive reasons to support dispositions toward *enkrasia*, not *akrasia*, hence requirements 5-P and 5-N. We have decisive reasons to support dispositions toward the kinds of evidence responsiveness set out in 6–8. And we have decisive reasons to support dispositions to take intention-compatible steps toward the achievement of our intended ends, as embodied in requirement 4. (Notice that standard-fixing accounts do seem to support the claim entailed by 4-T that an affliction under which, whenever you try to take *M* you are prevented from believing it, would be a kind of irrationality. The disposition to form such beliefs is important because of its exercise in thinking about one's [instrumental] reasons, and that gives us a good reason to regard a failure of this kind as a defect.)

In claiming that there is a relationship of coherence between a standard-fixing account and these requirements, I am claiming that the support runs in both directions. The plausibility of a standard-fixing account as an outline description of what rationality is supports my claims that 4–8 are indeed requirements of rationality; the intuitive plausibility of those claims, seen through their implications for judgments about particular cases, supports a standard-fixing account.

Finally, note that this account of rationality makes room for the claim that 9 is a requirement of rationality. It is one of the strengths of standard-fixing accounts, though, that they also offer a plausible explanation of why that should seem a controversial claim. On a standard-fixing account, the question we need to ask is whether, in Huckleberry Finn cases, there is a robust psychological disposition that we could

72. I would not want to suggest that a completely unquestioning obedience to military superiors is ever (actually) what there is a sufficient reason to have. See Jessica Wolfendale, *Torture and the Military Profession* (New York: Palgrave Macmillan, 2007).

encourage by means of the standard embodied in 9—a disposition to respond appropriately to the reasons you have evidence to believe you have, even when you lack those beliefs. We might worry that there is nothing I could do to improve my tendency to comply with 9, beyond being sensitive to the evidence of what I really do have a decisive reason to do—that is, beyond following 7. Given our fallibility as judges of our own reasons, we do have good reasons to be disposed to be self-mistrustful and to listen to our emotions—but we had better not encourage in ourselves a general distrust of our own beliefs about reasons. It would be good to have a disposition to distrust such beliefs which is sensitive to whether one is in a Huckleberry Finn case, in which there is strong evidence against them. But it is not clear that there is a robust disposition of this kind that such a standard could be used to encourage.

#### VIII. EXPLAINING REQUIREMENTS OF INSTRUMENTAL RATIONALITY

At the end of Section V, we were left wanting to explain why 4 is a requirement of rationality. One way of responding to this would simply be to stop with what has just been said, observing that the disposition to conform with 4 is a disposition which is important because of its exercise in thinking about reasons (although it is not always exercised in doing that) and whose importance gives us a decisive reason to support it. So a standard-fixing account does uphold the claim that 4 is a requirement of rationality.

However, on top of this, we are now in a position to notice that there is, after all, a cognitivist explanation of at least some core applications of 4. I shall close by explaining this.

At the outset, I made a broad distinction between to-intentions and from-intentions, introducing them as commitments that “typically” play two different roles—roles in planning toward and planning from the goals to which we are committed. I said I would eventually replace this with “to the extent that we are rational.” I think that a standard-fixing account of rationality allows me to do that. We have decisive reasons to commit ourselves to goals and to structure our planning around those commitments. Because of the importance of planning to effective agency, we need to make such commitments, and we need to have certain dispositions toward them if they are to play the roles we need them to play in facilitating action. To-intentions are commitments whose effectiveness depends on our being disposed to do what promotes and avoid what conflicts with their co-realization. This means that it is true (indeed, it is definitive)<sup>73</sup> of to-intentions that they are subject to the following “agglomeration” requirement:

73. Whether a state counts as a to-intention is settled by whether it is a commitment

10. Rationality requires of you that (if you intend *E* and you intend *F*, then you intend [*E* and *F*]).<sup>74</sup>

From-intentions are commitments of which this is true and which play the role we need them to play within our planning through our being disposed to assume that they will be achieved—doing this in order to help us think about what to do subsequently. So from-intentions are those that are also subject to a requirement of belief implication:

11. Rationality requires of you that (if you intend *E* then you believe *E*).

I do not claim that states that lack the features of agglomeration or belief implication cannot be to- or from-intentions. Rather, the claim is that we form commitments which play the roles which we need them to play in our deliberation only through our being disposed to conform with these requirements; on a standard-fixing account, they are therefore requirements of rationality, and ‘to-’ and ‘from-intentions’ are my labels for the states to which these requirements apply.

If this is right, then there is a cognitivist explanation of some applications of 4: applications of 4-T to from-intentions.<sup>75</sup> Here is 4-T again:

- 4-T. Rationality requires of you that (if, during a period through which you have an orderly succession of temporal beliefs,  
 i) you intend *E* throughout that period,  
 ii) you believe throughout that period that *E* will only suitably be achieved by your taking *M* at or before *t*, and  
 iii) at the end of that period you believe that it is *t* and that you have not taken *M*,  
 then you believe that you are taking or trying to take *M*).

Suppose I have a from-intention that satisfies the antecedent of 4-T. I from-intend *E*, I believe that *E* will only happen suitably if I take *M* by *t*, I believe that *t* has arrived, and I believe that I have not taken *M* yet. For *E* to happen suitably is for it to happen without frustrating any of my other intended ends. So consider all of my intended ends: *E* and all the others. Since I intend each of them, then if I also fail to intend all of them I am not as rationality requires me to be: the agglomeration

---

to which this requirement properly applies. Thus, merely orchestrating your action toward the achievement of a goal is not sufficient for a to-intention since it may be rational for you to aim at *A* and at the same time to aim at *B*, even though achieving both is undesirable. See Bratman’s video game example in Bratman, *Intention, Plans, and Practical Reason*, chap. 8, for illustration and discussion.

74. Compare *ibid.*, 134–38.

75. I present what follows as an explanation of why these applications of 4 are requirements of rationality. Unlike Broome, I offer no cognitivist explanation of the correctness of practical reasoning.

requirement (10) tells us this. And if I do intend all of them, then if I also fail to believe they will all be achieved I am not as rationality requires me to be: this follows from the belief-implication requirement (11). In believing that *E* will only suitably happen if *M* happens at or before *t*, I believe that the only way all of my intended ends will happen is if I take *M* at or before *t*. And I believe it is *t*. So, if I am to be as rationality requires me to be, the contents of my beliefs must entail that I am taking *M* now. One thing I could believe which would be consistent with that is that I am trying to take *M* and will now succeed. Or I could just believe that I am taking *M*, without believing that I am trying. But unless I believe either of those things, I am failing to believe what is entailed by my other beliefs. In general, I am not rationally required to believe what is entailed by my other beliefs. But I am rationally required to believe what is important and is obviously entailed by my other beliefs. This seems right and is upheld by a standard-fixing account of rationality. If so, then that does imply that if I do not believe that I am taking or (successfully) trying to take *M* now, I am not as rationality requires me to be. So 4-T is a rational requirement.

This gives us a cognitivist defense of the application of 4-T to from-intentions. Since from-intentions are states whose rationality depends on supplying us with beliefs, it is not a surprise to find such an explanation. If I am in violation of 4-T, the incoherence of the beliefs associated with my intentions makes me irrational. At the end of Section V, I remarked that the prospects of a cognitivist defense of a principle of instrumental rationality may seem to be removed once we move from the narrow range of cases in which you believe that *E* requires *M* to the broader range in which you believe that *E* cannot suitably happen without *M*. That turns out not to be true. When *E* cannot suitably happen without *M*, that does not mean that *E* requires *M*, but it means that achieving all of one's ends requires *M*.

However, the conditions required to make this argument work are very strong. The argument should make us wonder how many ends any person has for which it is rational to form from-intentions, as defined here. After all, the larger the number of ends you adopt, the less rational it is to believe they will all be achieved.

A parallel argument could be attempted for 4-I. But the most it would show is that I violate a requirement of rationality in from-intending *E*, believing that *E* will only be suitably achieved if I intend some means but not believing I intend such a means.<sup>76</sup> If I do succeed in forming a false belief of this kind without the intention, an argument of this kind cannot show that the failure to form the intention is irra-

76. Note that in doing so, this goes beyond 4-I, which does not imply (although of course it does not deny) that this combination of states is irrational.

tional. But the failure to form such intentions is irrational. Standard-fixing accounts correctly imply this since the disposition to form such intentions is one that we have reason to support. However, the cognitivist argument does not extend to showing this.

More importantly, a cognitivist argument cannot account for the application of 4 to commitments that are to-intentions but fall short of from-intentions. We often have good reason to form such commitments, coordinating our action toward the co-realization of ends without being likely enough to succeed to give us reason to base our deliberation about what to do later on the belief that we will succeed. Commitments of this kind are not subject to the belief-implication requirement (11). Yet they clearly are subject to requirements of instrumental rationality.<sup>77</sup>

Again, this seems to be correctly handled by the account I have offered. I do not have to believe that I will be successful in carrying out my to-intentions in order to form and hold them rationally. However, I am conducting myself irrationally if I can see that there are things I could do that are required to make my to-intentions co-achievable, and I retain the intentions without doing those things. In such a case, my beliefs are not incoherent, but I am practically irrational. The issue is not whether I can rationally believe that my to-intentions will be achieved; it is whether I am making the right kind of difference between their being achieved or not. Not being disposed in my thought and conduct toward making that kind of difference is something a standard-fixing account tells us is irrational, independently of its impact on the coherence of my beliefs.

I think this provides further support for a standard-fixing account of rationality. Such an account of rationality supports the premises from which it is possible to provide a cognitivist explanation of instrumental irrationality in the cases where such an explanation is plausible but not where it is not.

To sum up, I have argued that a plausible account can be given of the relationship between reasons and rationality and that thinking about the bearing of decisions on both helps us to find it. In making a case for the attractiveness of this kind of standard-fixing account, I have not tried to defend it against objections which it needs to address. Answering those objections would mean developing a particular version of this view by adding further detail to the framework described here. One issue that needs attention is to develop a full account of exactly what “exercising a disposition in thinking about reasons” includes and excludes. Another is to consider whether the account needs to be narrowed by adding further conditions to the two I supplied. As it stands, there are

77. Bratman makes this point in “Intention, Belief, and Instrumental Rationality,” sec. 5.

various forms of insensitivity to obvious reasons which, without further restriction, this account will classify as irrational. To many people, that will seem too inclusive. It may be important to support the disposition not to tell risqué jokes at funerals by means of an evaluative standard that censures this behavior as crass, but that is not enough to make it irrational. If that is right, the most plausible kind of standard-fixing account is one that adds to what I have said, restricting itself to forms of self-frustration more narrowly conceived. However, a version of the standard-fixing view that refines it in this way will not remove the attractions set out above. I hope at least to have made a case for thinking that this kind of theory of rationality deserves to be taken seriously.