

Chapter 3

Acting on an Intention

Claire Finkelstein*

I

Suppose a person intends to go to France on a vacation. He purchases a plane ticket for a certain flight to Paris. What relation does his intention to take the flight bear to his actual taking of the flight? What, more generally, is the relation between the intention to do something and the actual doing of that thing, in the ordinary workings of a rational agent?

I shall ask this question in the context of cases of an unusual sort. For, as J. L. Austin once memorably remarked, “the abnormal will throw light on the normal, will help us to penetrate the blinding veil of ease and obviousness that hides the mechanisms of the natural successful act.”¹ In the sort of case I shall examine, the following conditions obtain: (1) if the agent forms the intention to φ and subsequently φ -es, he will be better off than if he had never embarked on that course in the first place, at the same time that (2) φ -ing, considered in isolation from the rest of the plan, would leave the agent worse off. The question that is usually asked about such cases is whether it is rational to φ .

A familiar type of reciprocation problem will provide an example. You and I are visitors at a faraway research institute. Each of us would benefit from the other's comments on our book manuscripts. Neither particularly wants to read the other's book: It will be time-consuming and not as useful as devoting time to our own work. But each would benefit more from having the other's comments, counting the costs of reciprocation, than from foregoing the exchange altogether. You propose that we swap manuscripts and exchange comments. The only problem is that your manuscript is ready first, and you cannot wait because it is due to the publisher shortly. My manuscript will not be finished for several months. You suggest that I read and comment on yours first. You will read and comment on mine when it is ready. Let us call this example “Institute”.

The problem in Institute is that I will not undertake to comment on your manuscript unless I am fully convinced that you will later comment on mine. And I have ample reason to doubt that you will hold up your end of the bargain, given what I know your preferences to be. For we work in different fields and do not expect to encounter one

* I am indebted to David Gauthier, Ulrike Heuer, Leo Katz, Connie Rosati, Michael Thompson, David Velleman, Bruno Verbeek and Gideon Yaffe for comments on the manuscript at various stages in its development, as well as to the participants in the 2002 conference where the paper was first presented.

¹ J. L. Austin, “A Plea for Excuses”, *Proceeding of the Aristotelian Society* 57 (1957): 1–30.

another again, and a bad reputation would be unlikely to carry over from one field to the next. Given that reading my manuscript is all cost and no gain, then, it looks as though you have no reason to read it.

Except for the following consideration. Assume that "common knowledge of rationality" obtains between us, meaning that each of us is rational, each one knows the other is rational, and each one knows the other knows that he is rational. Against the background of this assumption, perhaps you *do* have a reason to read my manuscript after all. For you prefer promising to read my manuscript and then actually reading it to avoiding any such promise but being unable to secure my help. Thus when you consider the alternatives, you will see you have a reason to adopt the former course of action – promise to read and then read – over going your own way. Naturally you would prefer to secure my assistance by promising to read my manuscript and then backing out of the deal after I have read yours. But arguably the common knowledge assumption makes false promises infeasible.

The point can be made by *reductio*. Suppose it *were* rational for you falsely to promise to read my manuscript. Given common knowledge, I would know this and I would therefore have a basis for attributing an intention to break your promise to you, since I am also rational and know you to be rational. The result would be that you would not secure my assistance. But you want to secure my help, even more than you want to avoid reading my manuscript. It cannot, then, be rational for you to promise to read my manuscript at the same time that you have no intention to read it. And if this is the case, you actually have reason to read my manuscript, given your preferences.

It is, however, difficult to make sense of this suggestion. In particular, what reason should a rational agent give himself for undertaking a costly action that will produce no further benefits for him, such as would be the case with your reading my manuscript? One answer is that a rational agent should adopt a "package deal" approach to rational decision-making. That is, he should not choose actions singly, taking each decision as it comes. Instead he should settle on a sensible plan for the accomplishment of his aims and he should then choose his actions solely as a function of his chosen plan. On this approach to practical reasoning, while reading my manuscript may not be rational in and of itself, since you obtain no future benefit from doing so, it is rational as part of the intention–action package of which it is a part.

The "package deal" approach has been amply defended by others, and it is not my purpose to undertake yet another defence from the ground up.² Instead, I wish to try to explore in greater detail the precise nature of the relation among the elements of the package. My hope is that such package-deal cases, assuming at least provisionally that we do accept their logic, will tell us something about the nature of the relation between intention and action in the more standard cases of acting on an intention. That is, I shall provisionally assume it would be rational to reciprocate in Institute, and ask how we might best account for this being so. And I shall then attempt to draw some more general conclusions about the relation between intentions and actions in satisfaction of those intentions.

2 Gauthier, "Assure and Threaten"; McClennen, *Rationality and Dynamic Choice*; see also Bratman, *Intention, Plans, and Practical Reason*.

II

Consider two possible defences of the suggestion that it is rational for you to read my manuscript. On the first view, reading my manuscript need not be separately rationalized. Its rationality follows from the rationality of *intending* to read my manuscript. On this view, forming a prior intention to φ *causes* the agent whose intention it is to φ by impeding him, if he is rational, from performing any action other than that called for by the intention. As Scott Shapiro puts the point, intentions "causally constrain non-conformity".³ On this view, forming a prior intention is like setting up an external pre-commitment device to force oneself to act later, with the difference that the agent who forms an intention is *internally*, rather than *externally* constrained. There is no difficulty explaining why it might be rational to reciprocate in Institute on this view: It is rational to act in the way demanded by the prior intention just in case it is rational to form that intention in the first place. Thus if the causal view is correct, I can reliably expect that having formed the intention to read my manuscript, you will in fact read it, assuming that when the time comes to do so the situation is as you supposed when you formed the intention.

But the ordinary relation between intending to φ and φ -ing cannot be *merely* causal.⁴ A person who intends to φ , but who later does not regard himself as having any reason to φ , is judged irrational, not broken. If there is anything to such judgments, the relation between intentions and actions in satisfaction of those intentions must be more than merely causal. As John Broome puts the point, the process that leads from intending to acting is "guided or controlled by reason". It is "normatively sanctioned".⁵

Consider, then, a second view, at the opposite extreme from a purely causal view. On this view, a prior intention to φ provides a reason to φ , and thus serves to rationalize a person's φ -ing (when he φ -es in satisfaction of an intention to φ). On this view, having formed the intention to read my manuscript, you now have a *reason* to read it. And thus even if you previously lacked any reason to read it, you have now managed to acquire one. If I know you to be rational, I can thus rely on your reading my manuscript, given that your intention to read it gives you a reason to read it.

The intentions-as-reasons view is at first blush more attractive than the purely causal view. For it explains how intentions can be effective in motivating rational agents, but in a way that leaves those agents in charge of the actions they perform. And at the same time it makes the relation between intentions and actions in satisfaction of those intentions appropriately normative. But upon further reflection, this view turns out to be problematic as well. As Michael Bratman and John Broome have separately argued, the intentions-as-reasons view leads to unacceptable

3 Scott Shapiro, "Authority", in Jules L. Coleman and Scott Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford: Oxford University Press, 2002), p. 418. See also Shapiro's contribution to this volume.

4 I say "ordinary" relation because I mean to screen out cases where the intention produces the action in satisfaction of that intention via some wayward causal chain.

5 Broome, "Are Intentions Reasons?", p. 104.

bootstrapping.⁶ Consider once again the man who wishes to go to France. Suppose he forms the intention to board a plane to Sydney, Australia, all the while retaining his desire to go to France. If, knowing the destination of the flight, he resolves nevertheless to take it, it would be absurd to think that his intending to take the flight to Sydney provides a reason for him to take that flight, or even ought to weigh in favour of his taking it. For that would mean that in cases in which the intended action was only weakly irrational, the reasons supplied by the intention to perform it could outweigh the background reasons against performing it. And it would then be all things considered rational to perform the otherwise irrational action. Having formed an intention to φ thus cannot *by itself* give a person a reason, even a *prima facie* reason, to φ .

We have said that a prior intention does not straightforwardly cause the action in satisfaction of that intention, but that the intention is also not a reason for an agent to act as the intention requires. How, then, might we account for the transformation of intention into action in a case like Institute?

In addition to the causal and the intentions-as-reasons view, there is a third possibility, one we might understand as lying mid-way between the first two. According to this third view, a rational agent executes a previously formed intention by first cultivating a disposition to stick to her intentions, or some other similar disposition that reliably carries her from intention to action in satisfaction of that intention. One might cultivate, for example, a disposition of steadfastness, in order to make oneself into the *kind of person* who generally sticks to her intentions. Such a disposition could mediate between an intention an agent has formed and an action in satisfaction of that intention: It would allow an agent to execute her intentions without subjecting them to further deliberation. Let us call this kind of disposition an “adaptive disposition”, and the view that makes use of it the “dispositions view”. Although my ultimate target is not the dispositions view per se, much of what I wish to say about the relation between intentions and actions in package-deal situations will be brought out by focusing on it, and in particular, by seeing what is problematic about that view. Let us then turn to an examination of the dispositions view and consider precisely how that view is supposed to bridge the gap between intention and action.

The dispositions view maintains that if it is rational to form adaptive dispositions, it is rational to act on those dispositions, even if the acts those dispositions require are suboptimal, considered in themselves. The point would apply to things other than intentions. We can explain the rationality of keeping a promise, say, in terms of a disposition to keep promises, and we can explain the rationality of being a promise-keeper in terms of self-interest. Like rule-utilitarianism, the dispositions view is a two-level theory, in which individual acts are justified only in terms of the disposition from which they issue; it is only the dispositions that are justified in terms of an agent's first-order reasons.⁷

6 Bratman, *Intention, Plans, and Practical Reason*. See also Broome, “Are Intentions Reasons?”, p. 99.

7 For an excellent discussion of dispositions accounts in the context of other two-level theories, see Michael Thompson, “Two Forms of Practical Generality”, in Christopher Morris and Arthur Ripstein (eds), *Practical Rationality and Preferences: Essays for David Gauthier* (Cambridge: Cambridge University Press, 2001).

A clear statement of this kind of account is offered by David Gauthier in *Morals by Agreement*.⁸ Gauthier argues that it is rational to develop a disposition to reason as a constrained, rather than a straightforward maximizer, because constrained maximizers are likely to fare better than straightforward maximizers.⁹ Unlike straightforward maximizers, constrained maximizers behave cooperatively, for they have “internalized the idea of mutual benefit.”¹⁰ They are able to enter into agreements and exchange promises, even in the absence of an enforcement mechanism, and in this way they can reap the gains from cooperative behaviour. Thus developing the disposition of a constrained maximizer is well-supported by an agent's self-interested reasons.

Why must an agent who wishes to behave cooperatively form a disposition to do so? If constraining one's maximizing behaviour is ultimately to one's benefit, why not just choose the action a constrained maximizer would choose, without bothering to develop a disposition to that effect? The dispositions theorist would probably explain the need to resort to dispositions in one of two ways. The first is that human beings are weak-willed, and that they cannot be trusted to carry out rational courses of action if resistance is high. Forming dispositions is a way of ensuring that rational agents adopt rational courses of action, against the background of their inherent weaknesses. But this explanation will not avail the dispositions theorist who begins with the premise of rational agency, for if agents are fully rational, they are not weak-willed. So dispositions theorists are confused if the reason they have turned to dispositions is to deal with imperfections of rationality.

The second reason has to do with the nature of rationality itself. Dispositions theorists assume that rational agents inevitably choose the option that gives them the highest local payoff. They assume, in other words, that it is not open to rational agents simply to reject the maximizing choice in favour of the constrained or cooperative option. They must instead find a way of constraining their own maximizing behaviour, and in the absence of any available pre-commitment device, a disposition will serve the purpose. The disposition is thus a way of obviating a limitation on instrumental rationality. Gauthierian constrained maximizers would thus have no difficulty cooperating in Institute, since the disposition they have acquired makes it easy for them to exchange commitments and stick to those commitments, in situations where both agents are better off under the terms of the commitment than they would be in its absence.

Another version of the dispositions view posits the need for a disposition pertaining specifically to whether a person reconsiders his intentions. Instead of acquiring the disposition to keep promises directly, or a series of other dispositions having to do with virtues of cooperation, this version says that rational agents should cultivate a disposition to execute their rationally formed intentions. Michael Bratman, for example, calls for sensible strategies or “habits of non-reconsideration”. These habits would be located between two extremes. On the one hand, they must avoid the mistake of

8 Gauthier no longer subscribes to the dispositional version of his account, but others have followed his earlier lead in this respect. And, although Gauthier has moved away from the dispositions view, he never clearly repudiated it. The problems with the dispositions view will show us why we should favor the more recent, non-dispositional formulation of Gauthier's account.

9 This is subject to the condition that they are dealing with other constrained maximizers.

10 Gauthier, *Morals by Agreement*, p. 157.

the person who “always seriously reconsiders his prior plans in the face of any new information, no matter how trivial”.¹¹ On the other hand they must avoid the opposite mistake of the overly rigid planner, “one who almost never reconsiders, even in the face of important new information”.¹² The planning theory of agency for which Bratman argues thus calls for agents to be reasonably steadfast in executing prior plans, but to be open to reconsideration, so that they can adjust their plans in the face of significant new information.

Now Bratman does not himself think it rational to reciprocate in cases like Institute, and so the habits of non-reconsideration he advocates are not intended to apply to the sort of case on which we have been focusing.¹³ But he does think that habits of non-reconsideration play a role in other cases. Consider “Rational Temptation”.¹⁴ Suppose that Barbara is deciding whether to have a second glass of wine with dinner. Before dinner she prefers that she forego the second glass, and after dinner she wishes that she had foregone it. But she may have a strong temporary preference during dinner for drinking the second glass of wine. What should a rational agent faced with such a preference reversal do? Like Gauthier, Bratman assumes that intentions or plans must be adopted on the basis of an agent’s occurrent preferences. Thus it is not open to Barbara to form the intention *during* dinner to stick to only one glass of wine. A helpful strategy, he therefore suggests, would be for her to form the intention *before* dinner to have only one glass of wine, and then hold fast to that intention during the change in preference, refusing to allow herself to reconsider the intention until the change of preference has passed.¹⁵

A possible difficulty with this strategy, Bratman allows, is that although Barbara must be capable of reconsidering previously formed intentions, the intention to drink only one glass of wine will probably be abandoned if Barbara does reconsider. For once Barbara’s preferences shift, she will have every incentive to reconsider her intention, and once she reconsiders, she will see no reason why she should not change her mind and have that second glass. Indeed, if Barbara is rational, it seems she *must* decide to have that second glass of wine, for otherwise she is failing to maximize her preferences. The solution, Bratman thinks, is for Barbara to rely on her habits of non-reconsideration. For if these habits are sensible, she will not deliberate about whether to abandon her intention. Assuming that the situation at dinner is substantially as she expected it would

be when she formed the intention to stick to one glass of wine, Barbara will be able to resist the second glass of wine.¹⁶

I shall not here elaborate all the different versions of the dispositions view.¹⁷ For what interests me is not the variations among them, but what they all have in common – the idea that rational-intention execution requires shutting off one’s deliberation after the intention has been formed. In my view, we should reject the non-deliberative aspect of such accounts. My thought is that in exposing the difficulties with shutting off deliberation in this way, we will learn something about the nature of the relation between intentions and actions more generally.¹⁸

My argument against the dispositions view is that it makes use of a non-deliberative mechanism for executing previously formed intentions, and that any such mechanism will fail to vindicate the rationality of the actions it requires. If an agent must adopt a mechanism for shutting off his deliberation in order to move from intention to action, we are no longer entitled to think of the intention and the action as standing in the kind of rational relation we require in order to vindicate the rationality of acting on the intention. That is, like the causal view, the dispositions view uses the relevant sort of disposition like a pre-commitment device: The choice of the disposition is supposed to eliminate the possibility of choosing in a way that is inconsistent with the disposition. But precommitment also obviates one of the central characteristics of rational intention-guided behaviour – that it is entirely up to the agent and very much the product of his will. This feature of intention-guided behaviour should allow us to articulate a condition that any account of the relation between intentions and the actions in furtherance of those intentions should meet, namely that the two must be related deliberatively. They must satisfy what I shall call the “deliberative requirement”.

16 Given that Bratman is willing to avail himself of a non-deliberative mechanism for moving from intention to action in Rational Temptation, it is curious that he nevertheless does not allow that it could be rational to reciprocate in cases like Institute. His reason for thinking Institute different from Rational Temptation is that, as he puts it, “What is up to the agent is what to do from now on”, rather than what to intend later. Bratman, “Toxin, Temptation, and the Stability of Intention”, p. 72. But I do not see why reciprocation cases like Institute require this. In Institute, I require a sincere commitment from you to *read my manuscript* – not just an intention on your part to read it. Institute is in this respect different from Kavka’s toxin puzzle (“Toxin”), in which the rewards one receives are explicitly based on the intention one forms, rather than on an act one performs in furtherance of that intention.

17 Other examples are Joe Mintoff, “Rational Cooperation, Intention, and Reconsideration”, *Ethics* 107, no. 4 (1997): 612–43, and, under quite different guise, E. F. McClennen, “Constrained Maximization and Resolute Choice”. McClennen appears to have at least partially recanted the idea since. See McClennen, “Pragmatic Rationality and Rules”, as well as his contribution to this volume.

18 Insofar as the object of my criticism of the dispositions view is its non-deliberative aspect, my complaint need not extend to other accounts that seek to justify individual actions according to a two-tiered structure. A social practice view of promising, for example, that justifies an individual act of promise-keeping in terms of a more general practice strikes me as untouched by the criticism here. This is despite Michael Thompson’s compelling arguments to the effect that all such two-tiered accounts should be seen as instances of the same type of logic. See Thompson, “Two Forms of Practical Generality”.

11 Bratman, “Planning and Temptation”, p. 52.

12 Bratman, “Planning and Temptation”, p. 52.

13 Bratman, “Toxin, Temptation, and the Stability of Intention”, p. 73. He also does not think it rational to drink the toxin in Kavka’s “toxin puzzle”, as he regards both sorts of cases as instances in which the agent receives “autonomous benefits” from forming the prior intention. See Kavka, “The Toxin Puzzle”.

14 The label comes from the fact that the temptation stems from an underlying preference reversal, rather than from weakness of will. For a more detailed discussion of such cases, see Claire Finkelstein, “Rational Temptation”, in Christopher Morris and Arthur Ripstein (eds.), *Practical Rationality and Preference: Essays for David Gauthier* (Cambridge: Cambridge University Press, 2001).

15 Bratman does allow that the temporary preference reversal might sometimes make it rational to reconsider one’s intention.

Now perhaps having a deliberative requirement as a condition on acting on an intention will make it impossible to explain how a person could rationally form the intention to reciprocate in a case like Institute. For if it is not rational for you to read my manuscript for its own sake, and it is also not possible for you rationally to execute a prior intention to read my manuscript by making use of some kind of non-deliberative mechanism, then perhaps it is simply not rational for you to read my manuscript. But I do not think the case for rationally committing to read, and then reading my manuscript, need be given up so quickly. Thus far, the deliberative requirement only serves to raise the stakes for anyone who wishes to argue for the rationality of cooperative action in this kind of case. I shall suggest below that one way of arguing for the rationality of cooperating in Institute *is* compatible with the deliberative requirement. First, however, we must explore the nature of the deliberative requirement more thoroughly than we have done thus far.

III

Compare Institute with another case of non-deliberative intention execution, which I shall call "Schelling's case".¹⁹ A robber is trying to coerce you into opening your safe and turning over your gold, and he is trying to do this by threatening to kill your children. You realize it is likely that once you actually do open the safe, he will kill you and your children anyway, since you are witnesses to his crime. You dare not open the safe, yet if you do not, you firmly believe he will start killing your children at any moment. The only way out you can see is to render yourself irrational by drinking an "irrationality" potion you happen to have. The potion will make you immune to coercion, because it will make you reason incoherently with respect to people and things you care about. The robber will have no choice but to abandon his plan, since threatening to kill your children will not have the effect for which he hoped.

Assume that the robber is a very good judge of psychological states. He knows immediately after you drink the potion that you are immune to coercion. It is not necessary, then, for you to *do* anything to convince him of this fact. But let us also assume that after you drink the potion you begin smashing your antique china. Smashing the china is an unfortunate side effect of drinking the potion. Now apply the reasoning from Institute to Schelling's case, and consider how it sounds: It is rational for you to smash your antique china in this case, because doing so stems from a psychological state it is rational for you to be in.

Indeed, Gauthier once argued just this: He claimed that you are best off being irrational in the face of the robber's threats, and so best off acting as you do (given that you cannot make yourself irrational without acting in that way). And if you are best off acting that way, it must be rational to act in that way. As he wrote:

19 The example is drawn from Schelling, *The Strategy of Conflict*.

In Schelling's answer the disposition that will make one's life go as well as possible is to act in a quite random and uncontrolled manner. On my account ... it is therefore the rational disposition in such situations, and the actions to which it gives rise are rational actions.²⁰

On the dispositions view, then, Schelling's case and Institute are the same. In both cases, the rationality of the action depends on the rationality of the disposition, and the rationality of the disposition depends solely on what would be of benefit to the agent. In Institute, being a promise-keeper would be to your benefit. In Schelling's case, the benefit lies in being irrational. In each case, because it is beneficial to the agent to acquire that disposition, its acquisition, and hence the actions that flow from it, must themselves be rational.

But I would not treat Institute and Schelling's case alike: Your reading my manuscript is potentially rational, but I do not see how smashing your antique china could be. Why not? It would be tempting to explain the difference between the cases by saying that in Institute, the relevant actions are beneficial to one's interests, whereas in Schelling's case they are detrimental. But this would not quite be correct. Reading my manuscript in Institute is detrimental to your interests, considered in itself. More importantly, the detrimental nature of the actions is an incidental feature of Schelling's case. Consider a further variation. Instead of smashing your antique china, the potion makes you behave in ways that are beneficial. You immediately start paying old bills, filing papers, cleaning up the house and so on. The drug makes you lose all procrastination from which you ordinarily suffer. It also happens to make you immune to coercion with regard to the lives of your children. Now it seems to me that the fact that the actions you perform under the influence of the potion are beneficial is neither here nor there from the standpoint of their rationality. Indeed, I think we should accept this as a general thesis: The fact that an action you perform is beneficial is insufficient by itself to establish it as rational. Gauthier, at any rate, ought to accept this, since he subscribes to the corresponding thesis on the detriment side: The fact that an action one performs at a given time (such as reading my manuscript) is detrimental, considered in itself is insufficient to establish that action as irrational.²¹

If the fact that the actions you perform in Schelling's case are detrimental is insufficient to establish them as irrational, why do I maintain that they *are* irrational? Why not think of them like your reading my manuscript in Institute? The salient feature of Schelling's case is the fact, as Gauthier himself notes, that your actions are "not under [your] *reasoned* control".²² The actions you perform under the influence of the drug are

20 David Gauthier, "Rationality and the Rational Aim", in Jonathan Dancy (ed.), *Reading Parfit* (Oxford: Blackwell, 1997), p. 38.

21 The situation might be different at the level of *courses of action*, for the fact that a course of action is beneficial in Gauthier's scheme should be sufficient to establish it as rational to adopt – unless we think of courses of action as contributing to something more general still, such as policies. Further, policies may themselves be in the service of something more general, such as *modes of reasoning*. At the most general level, benefit and rationality should finally line up.

22 Gauthier goes on to say that rationality is a "technical term" in his account, and that we have reason to reject our pre-theoretical account of control as essential to rationality. Gauthier, "Rationality and the Rational Aim", p. 38. I am unclear what he means by "technical term" in this context, and so I see no reason to reject widely shared pre-theoretical intuitions on this point.

not ones you have willed or chosen. While they are reason-induced, in the sense that they are caused by a potion which you had a good reason for drinking, they are not, we might say, under the *authority* of your reason while they are being performed. And this seems another way of saying that the actions you perform in Schelling's case do not satisfy the deliberative requirement. To understand more clearly what that requirement consists in, however, we must understand better the distinction I have invoked between an action's being *reason-induced* and being *under the control or authority of reason*.

Being induced by the agent's reason just means that the action originated in an exercise of reason on the part of the agent. But this cannot be sufficient for rational justification. For a burp or a sneeze could be rationally induced as well, viz. I gulp down air to produce a burp, or smell some pepper to induce a sneeze. Of course burping and sneezing are not normally *actions*, as they are not intentional under *any* description. But the point of Schelling's case is precisely that there are some odd cases where actions are like burps or sneezes – they can be produced by reason without themselves being rational actions. Thus actions that are under the control of the agent's reason must be more, or differently, connected with reason than simply being caused by it. What more ought we to require?

If an action is under the control of the agent's reason, there are two prominent characteristics beyond the merely causal factor we have already noted. First, it must be up to the agent whether to perform the action. That is, the agent can herself decide whether to act on the intention she has formed or whether to reconsider and revise her intention. This being "up to the agent" is crucial for distinguishing a deliberative from a causal account. For on the causal theory, the prior intention must so thoroughly constrain the action in satisfaction of that intention that it is simply not open to the agent whether to reconsider and revise the intention.

Second, the agent must guide and direct the actions she performs in fulfilment of her prior intention. As David Velleman writes, "[the agent's] role is to intervene between ... intention and bodily movements, ... guided by the one to produce the other. And intervening between these items is not something that the items themselves can do."²³ An intention, in other words, does not itself guide the agent to the action he must perform in fulfilment of that intention. The agent must do the guiding, making judgments about both how the intention can be fulfilled as well as about whether to fulfil the intention, in light of the balance of reasons. It is thus the agent's on-going commitment to the course of action he has selected that will convert intention to action, a commitment that requires the continual deliberative engagement of the agent.

Dispositional accounts fail to satisfy the two aspects of the deliberative requirement: first, whether to execute the intention is not up to the agent, on such accounts, and second, the agent does not guide or direct the action once the intention is formed. For on this account, once you have acquired the promise-keeping disposition and promised to read my manuscript, the problem is out of your hands: the disposition will take over and you will find yourself reading my manuscript when the time for action comes. Indeed, we might say that relinquishment of control is the point of acquiring adaptive dispositions, since it is an agent's control over the actions he performs that impedes

his performing the cooperative action. Thus if I am correct to impose satisfying the deliberative requirement as a condition on rational intention execution, the actions that follow from adaptive dispositions are not themselves rational actions.

Now it might be objected that forming a disposition does *not* involve relinquishing control, in the way that imbibing a foreign substance would. For a person can always seek to change a disposition once it has been acquired, and thereby avoid performing the actions to which he has become disposed. The person under the influence of the potion, by contrast, can do nothing to avoid its effects, short of discovering an external antidote that might eliminate its hold on him. But while we might admit to this difference between the cases, it does not detract from their similarity in the relevant respect. For, by the dispositionalist's own account, if the disposition does not deprive the agent of control, it will not be of assistance in executing prior intentions. And this seems to suggest that the dispositionalist is in a bit of a bind. For either forming an adaptive disposition involves relinquishing rational control or it does not. If it does, it will be possible for an agent in the kind of example we are considering to execute her prior intention. But in that case, the rational connection between intention and action will be severed, and we cannot say that the action is rational just in virtue of the rationality of the intention that precedes it. If it does *not*, the intention is vulnerable to reconsideration, and, according to the proponents of the dispositions view, the agent will fail to act on that intention. If this is correct, then while it may be possible to reciprocate in Institute, it is not possible to do so rationally. It is possible rationally to adopt a mechanism by which one causes oneself to do so, but such mechanisms cannot vindicate the rationality of the action required by the mechanism itself.

23 J. David Velleman, "What Happens When Someone Acts?", *The Possibility of Practical Reason* (Oxford: Clarendon Press, 2000), p. 125.