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.” In the sort of case I shall examine, the following conditions obtain: (1) if the agent forms the intention to ϕ and subsequently ϕ is, 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 Yaffee 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.

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—ever 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 reduction. 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 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. 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"; Mcclenon, Rationality and Dynamic Choice; see also Bratman, Intention, Plan, 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.” 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. 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 act is “guided or controlled by reason”. It is “normatively sanctioned.”

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 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.
A clear statement of this kind of account is offered by David Gauthier in *Morals by Agreement*. He argues that it is rational to develop a disposition to reason as a constrained agent, rather than as a straightforward maximizer, because constrained maximizers are likely to fare better than straightforward maximizers. Unlike straightforward maximizers, constrained maximizers behave cooperatively, for they have “internalized the idea of mutual benefit.” 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 behavior. 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 behavior 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 dispositional 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 dispositional theorist who begins with the premise of rational agency, for if agents are fully rational, they are not weak-willed. So dispositional 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. Dispositional 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 favor of the constrained or cooperative option. They must instead find a way of constraining their own maximizing behavior, 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 dispositional 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 dispositional view, he never clearly repudiated it. The problems with the dispositional 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.

the person who “always seriously reconceives 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 reconceives, 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 Gauchet, Bratman assumes that intentions or plans must be adopted on the basis of an agent’s current 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”.

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 Finkelein, “Rational Temptation”, in Christopher Morris and Arthur Ripstein (eds.), Practical Rationality and Preference: Essays for David Gauchet (Cambridge: Cambridge University Press, 2001).
15 Bratman does allow that the temporary preference reversal might sometimes make it rational to reconsider one’s intention.
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”, Ethos 107, no. 4 (1977): 612–43; and, under quite different guise, E. F. Mcclennen, “Constrained Maximization and Resolve 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”.
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”. 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 a 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:


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 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 fulfillment 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 fulfillment 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.” An intention, in other words, does not itself guide the agent to the action he must perform in fulfillment of that intention. The agent must do the guiding, making judgments about both how the intention can be fulfilled as well as whether to fulfill 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 dispositional dispositions, since it is an agent’s control over the actions he performs that impedes his performance of 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 dispositional 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 dis positionalist’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 preceded 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.

IV

I have argued that the performance of an action in satisfaction of a prior intention must be under the rational control of the agent if the action is to inherit the intention’s rationality. And I have argued that the dispositions view does not meet this condition. But someone may now wish to issue the following defence of that view.

The dispositions view arguably retains the control needed for rational action, despite the local relinquishment of control the acquisition of a disposition involves, given that it is up to the agent whether to form the relevant disposition in the first place. Moreover, it is up to the agent which intentions he wishes to execute with the help of those dispositions. So even if the disposition itself removes the action from the agent’s control at the moment of performance, it is up to the agent whether to perform such acts in a larger sense, since he would not perform them if he had not chosen both intention and disposition required for their execution. Thus the fact that an agent uses dispositions to help himself perform certain actions does not vitiate his control over them, any more than using a hammer to drive in a nail would, or standing on a ladder to reach a high object. And if this is correct, then an agent’s rational control over the acquisition and use of a disposition may be adequate to entail rational control over the actions he performs on the basis of those dispositions.

Now the first thing to notice about this move is that it does not allow the dispositionalist to distinguish cases involving dispositions from Schelling’s case.

After all, a person who uses an external pre-commitment device to ensure that he sticks to a promise also chooses the method by which he binds himself to the satisfaction of the promise. So if we think that the control an agent exercises over the formation of the disposition is sufficient to establish the action as under his control in the relevant sense, we would have to allow the same of the agent’s actions in Schelling’s case. And then we would have to allow, contrary to what we have thought, that an agent’s choice of a mechanism for realizing a prior intention is sufficient to establish an action that follows from that mechanism as one’s own, as long as the mechanism produces results the agent regards as beneficial. So let us take seriously this revised position. Should we accept choice of a disposition, plus benefit to the agent, as sufficient to establish an action that follows from that disposition as rational?

To answer this question, let us return to consider how an agent is expected to form an adaptive disposition. At least on Bratman’s view and Gauthier’s former view, the disposition is selected on the basis of the preferences of the agent at the time he forms the disposition. But why should we think of the choice an agent makes on the basis of those preferences as more rational than the choice he would make on the basis of the preferences he later comes to have? On the dispositions view, the relevant disposition only serves to entrain a particular set of preferences, and these are often not the preferences the agent has when the time for action arrives. If the choice of a disposition is supposed to rationalize acts that follow from that disposition, the choice must itself be supported by something deeper or more authentic than the preferences that now conflict with the dictates of the disposition. But if the disposition can only be chosen on the basis of the same sort of item as the acts that would conflict with the disposition, there are no grounds for thinking that the acts that follow from the disposition are better, more rational, or more endorsed by the agent than the alternatives.

Now it might be argued that this is not the right way to think of the choice of a disposition. Perhaps instead of choosing dispositions as a way of maximizing some occurrence desire they have, agents select dispositions as part of an overall choice of identity. That is, the choice of a disposition might be more like the acquisition of a virtue: I choose to acquire the disposition of courage because I want to be a brave person. I choose to acquire the disposition of liberality because I want to be a generous person. Dispositions, in short, might be a way of realizing a self-ideal, rather than a way of maximizing preferences. If we are prepared to uncouple dispositions from preferences in this way, it would be easier to see acts that follow from dispositions as more authentic, more genuine and more reflective of a person’s agency than acts that would conflict with a person’s dispositions. Although the foregoing is not typically what the dispositions theorist has in mind, it might be a more defensible approach.

But once we allow that an agent can choose to acquire a disposition in this more reflective and autonomous way, the disposition itself arguably becomes otiose. For if I can choose a disposition on the basis of an ideal, instead of a preference, I should be able to choose to stick to a previously formed intention in the face of desires I have to the contrary, when those intentions realize the same ideal. If, in other words, I am capable of making autonomous choices that are unconfined by my current desires or preferences, then presumably I have the ability to realize an intention I have settled on, regardless of the state of my desires when the time comes to act. Acting in the way demanded by the prior intention may be as well supported by my self-ideal as developing the disposition in question. And if this is so, I will not require a disposition in order to stick to a prior intention, even if the action required by that intention is unsupported by my preferences when the time for action arrives.

In Part III, we identified a dilemma for the proponent of the dispositions view. We have now identified a second dilemma for the dispositionalist if he tries to solve this difficulty by availing himself of a second-order strategy. If the second-order choice or endorsement of a disposition must itself be based on desires, the second-order judgment will not have the normative force to rationalize the acts that flow from the disposition. On the other hand, if I have the ability to choose my disposition fully and reflectively, that is, irrespective of the state of my desires, then one would assume that I have the ability to choose whether or not to perform the relevant actions, irrespective of the disposition. For all of the above reasons, then, the second-order strategy does not seem to be an improvement over the basic turn to dispositions, and the difficulties with the dispositions view remain.

V

We started out by claiming that the relation between intentions and actions in satisfaction of those intentions cannot be merely causal; it must instead, or also, be a normative relation. But we also rejected the most straightforward way of cashing out the idea of a normative relation between the two, namely that the intention itself supplies a reason to perform the action in furtherance of that intention. Can we identify any normative connection between intentions and actions in satisfaction of those intentions once we reject the idea that the former can be a reason to perform the latter? Can we go further and identify some sort of alternative rational relation between the two?

John Broome offers one type of “normative relation” between intentions and actions in satisfaction of those intentions, which is not itself the normativity of reasons for acting. Instead, he suggests we think of the action as entailed by the intention, much in the way that accepting the conclusion of a piece of theoretical reasoning is entailed by our belief in its factual premises. Just as we are required to believe q once we believe p, when p entails q,

24 For a fuller discussion of this point, see Finkelstein, "Rational Temptation".

25 It is worth noting that the problem with this second order move is reminiscent of one Warren Quinn identifies for noncognitivists in metaethics, according to which ethical judgments are nothing more than dispositions to behave in certain ways based on pro- or con-attitudes agents possess. Quinn argues that pro-attitudes that dispose us to act cannot, by themselves, travel any distance toward rationalizing the acts that agents who have those attitudes are disposed to perform. He then considers whether a “second-order endorsement” of the agent’s dispositional state would improve the dispositional account. He concludes that it would not, since the noncognitivist is not entitled to appeal to anything other than a disposition to endorse or reject first-order dispositions. And if first-order dispositions cannot serve to rationalize the acts that flow from them, a second-order disposition to choose first-order dispositions could not fare any better. Warren Quinn, "Putting Rationality in Its Place", in R. G. Frey and Christopher Morris (eds), Value, Welfare, and Morality (New York: Cambridge University Press, 1993).

26 Broome, "Are Intentions Reasons?"
so we are required to \( p \) once we have formed the intention to \( q \). Moreover, intending to \( q \) normatively requires an agent to \( q \), even if the intention to \( q \) is irrational in the first place, just as believing that \( p \) normatively entails believing that \( q \), even if one ought not to have formed the belief that \( p \) in the first place. This could not be said if intentions supplied reasons to perform the required action. For if intending to \( p \) did supply a reason to \( q \), even if you had no reason to intend that \( p \), then you would have a reason to \( q \), just by virtue of having formed the intention to \( p \). And this would be bootstrapping. So instead, Broome thinks, we can explain how the process by which we come to execute our intentions is a rational one without saying that the intention gives the agent a reason to perform the action. Instead, the process is rational because the intention "normatively requires to be carried out", as long as the agent has not repudiated the intention. This solves the bootstrapping problem without abandoning the normativity of the relation between intentions and actions in satisfaction of those intentions.

A shortcoming of Broome's approach, however, is that it does not yet allow us to explain how it is that you can be thought of as having a reason to reciprocate in Institute. Indeed, because he does not think intentions themselves supply reasons, he does not think you do have a reason to reciprocate in such a case. For Broome assumes that if you have a reason to reciprocate, it could only be because the intention itself supplies you with that reason. But even if we agree with Broome's arguments against the intentions-as-reasons view, we need not reject your having a reason to reciprocate in Institute. For we need not suppose that your prior intention alone must supply you with that reason. In virtue of what, then, might you have a reason to reciprocate in Institute?

Instead of saying that the intention to read my manuscript supplies you with a reason to read it, we can say that the reason you had for adopting the package ("intention plus-action") supplies you with a reason to perform the action required by that package. In this way, your intention to read my manuscript can figure in the reason you have to read it, without itself supplying that reason. In the kinds of cases we are considering, then, the intention and action are additionally related by the fact that they are both parts of the plan whose adoption will leave you better off than any other plan it was open to you to adopt, over and above the general normative relation that holds between intentions and the actions that satisfy them. The reason for adopting the plan thus constitutes a reason both for forming the relevant intention and for executing the action as a way of fulfilling that intention. There is no bootstrapping, and there is no need to resort to non-deliberative forms of intention-execution.

It will seem mysterious how something that rationalizes your adoption of the package could supply a reason for an action that is not instrumentally necessary for obtaining the benefits from that package. Why should we think the fact that reading my manuscript is part of a plan it is rational to adopt provides you with a reason to read it, given that reading it is not by itself rational, and given too, as I have claimed, that the fact that you have adopted the plan is not itself of relevance? The answer lies in the thought that we can value something because of its role in constituting a larger item we value. That is, the relation between the action and the intention-action package that requires it is constitutive, rather than instrumental. The value of an individual note in a melody disappears when that note is played in isolation from the melody it helps to compose. But we certainly would not for that reason imagine that the note is instrumentally related to the piece of music it inhabits. Where constitutive goods are concerned, the fact that we cannot place independent value on the constituents does not mean that we do not value those items. And similarly, the fact that an act called for by an optimal plan is "suboptimal", considered in itself, does not mean that we place no value on the act. For we value the act as constitutive of the value we place on the overall plan. In such cases, there is a reason for performing suboptimal actions required by the plan, a reason that is inherited from the reason we have for valuing the larger item of which it is a part.

Now one might object that in the case of constitutive goods, we normally do not disvalue the items when considered separately in the way that we do the action that is part of the intention-action package in Institute. One does not disvalue each separate note of a melody in the way that one would disvalue the act required by the relevant promise in Institute. But there are other cases in which the analogy can be more tightly constructed. Consider a common experience when attending opera, the theatre or a concert. As the second act grinds on, and your energy starts to fade, you would be glad to learn that the opera will last five minutes less than you had originally anticipated. It is not that you are not enjoying yourself. You are, but the enjoyment is combined with a certain impatience to have the event over and done with. If, once again, you learn that another five-minute segment had been eliminated from the opera, you might again be pleased. There seems no particular point at which you would not to be pleased to have five minutes from the opera removed. Yet you would not at all be pleased to have the entire event eliminated in this fashion, segment by segment. When the opera is over, you are glad you attended, and would not particularly wish to remove any five-minute segment from your experience.27 The suboptimal action required by the optimal plan may be like the five-minute segment of the somewhat tedious opera: considered in abstraction from the larger item of which it is a part, it is not particularly desirable. But taken as a component of that larger item, where the larger item itself confers great value, its value is no longer difficult to discern. Would we, after all, regard a person who did not choose to eliminate a five-minute segment from the opera as irrational under these circumstances? Surely not.

Perhaps a more significant objection runs as follows.28 In general, the properties of a composite whole cannot be imputed to the subparts that constitute it. And thus the fact that the opera as a whole has value does not mean that this value can be imputed separately to each five-minute segment. The clearest example is the relation between supervenient physical properties and the micro-level entities that constitute them, such as that between water and the individual molecules of hydrogen and oxygen of which it is composed. The fact that individual molecules of hydrogen and oxygen make water "wet" does not lead us to impute the quality of wetness to each individual molecule. The macro-level properties we attribute to physical objects are always a product, or

27 There is an obvious resonance here with slippery-slope intransitivities such as Warren Quinn presents in "The Puzzle of the Self-Torturer", Morality and Action (Cambridge: Cambridge University Press, 1993). The difference is that in Quinn's case, the value true places on the whole is negative, and the segments that make up the whole are positively valued. Each turn of the crank that increases the electric current is valued in and of itself positively, despite the fact that each turn of the crank is a constituent of the larger total disvalue you place on finding yourself at the highest setting and in excruciating pain.

28 I am indebted to Bruno Verbeek for bringing the following point to my attention.
function of, the micro-level structure of those objects, without it being possible to
impute the macro-level properties to the items that make up the micro-level. And if
the rationality of plans functions like the wetness of water, then we cannot impute
the rationality of the overall plan to the individual acts that help to constitute the plan.

To be sure, it may, however, be that supervenient physical properties function
differently from relations of value. After all, it is somewhat less intuitive to think of
value as “supervening” on elements entirely lacking in value, and thus more natural
to think of the value of the res as residing in the constitutive acts of the plan. But the above
observation is also suggested by G. E. Moore’s famous discussion of “organic units” in
Principia Ethica. Moore argues that certain entities of value are such that “the value of . . .
[the] whole bears no regular proportion to the sum of the values of its parts”. 29 Moore
offers the example of a beautiful object, as having little value if no one is conscious
of it, but having great value if there is someone there to appreciate it. Consider
an example Christine Korsgaard offers to elucidate Moore: a beautiful painting and
someone’s appreciation of the painting. 30 Neither considered in itself has particular
value: a beautiful painting forever locked in a closet where no one will ever see it
surely does not have much, or any, value. And a person’s ability to appreciate beautiful
paintings, in the absence of any actual paintings to observe, is also devoid of value. But
the combination of the appreciation and an actual painting has great value. Moore also
offers the example of vindictive punishment: the painful treatment of a man who has
an evil state of mind has positive value, even though the value of each of the separate
components is negative. 31 Thus Moore might have said that the five-minute segments
of the opera form an organic unity and that the value of the whole both exceeds the
sum of its parts, and cannot, in turn, be imputed to those parts.

While it is true that the wetness of water or the solidity of a table cannot be imputed
to the molecules that constitute these items, there is a sense, pace Moore, in which the
value of the opera can be imputed to the five-minute segments that constitute it.
In the latter case, the value of the whole can be imputed to its constituent elements,
conditional upon those elements remaining part of the whole. As Korsgaard argues, it seems
perfectly appropriate to say that the value of a beautiful painting is conditioned on
some human observer being present to appreciate it, without saying either that the
painting is unconditionally valuable, or that its value is part of an overall organic
unity that is not separately attributable to the painting itself. What we say is that the condition
of the goodness of the painting is that it is viewed. 32 Thus the painting is intrinsically
valuable without being unconditionally valuable. And this allows us to locate the value
in the painting itself, without thinking of the painting as valuable under all conditions.
In Korsgaardian terms, then, we can say that each five-minute segment of the opera
is valuable, conditioned on its occupying a place in an overall valuable opera. The
same cannot be said, however, for wetness, solidity, blueness and other macro-level

29 George E. Moore, Principia Ethica (Cambridge: Cambridge University Press, 1903),
chapter 1, § 18.
30 See Christine M. Korsgaard, “Two Distinctions in Goodness”, Creating the Kingdom of
31 Moore, Principia Ethica, § 128.
32 Korsgaard, “Two Distinctions in Goodness”, p. 264.

properties. We cannot say that individual water molecules are themselves “wet”, even in
the conditional way we can with relations of value.

We may thus have found a way of accounting for the value of an individual part
of an overall valuable plan, when the part is subterminal, considered in and of itself.
In a case like Institute, this means that we have a way of arguing for the rationality of
reciprocating directly, consistent with the deliberative requirement. It is rational for you
to reciprocate in Institute, as long as the reason you have to reciprocate depends on the
place that action occupies in the overall plan of which it is a part. Alternatively, we can
say that the value of reciprocating is conditioned on its occupying a certain position in the
overall plan of agreeing to read, and then reading, my manuscript, much in the way that the
value of a beautiful painting is conditioned on the existence of a human viewer to
appreciate it. That reciprocating is valuable in this way is what makes it rational, for the
logic of rationality should surely track the logic of value more generally. The rationality
of reciprocation in such a case is thus directly supported by an agent’s reasons, rather
than indirectly, as it is when we make use of a disposition. The rational support in
question is contextual and holistic. 33

VI

At the outset of our inquiry, we expressed a hope that cases like Institute would
have something more general to teach us about the relation between intentions and actions.
Let us consider whether that hope has paid off. What we have identified is a kind of
normative relation that holds between an intention and an action that satisfies that
intention: both are ineliminable parts of a larger structure which is itself well supported
by the agent’s reasons for acting. It follows that each part of this structure is also
supported by an agent’s reasons, but only on the condition that the other parts also
remain. We found that this indirect or conditional normative relation supplements the
more direct normative relation that holds between intentions and actions identified by
Broome. The question now is whether it supplements the general normativity of the
relation between intentions and actions in every case, or whether it only does so in cases
like Institute.

Consider once again the man who intends to travel to France. What must he do
to satisfy that intention? Notice that he can satisfy that intention in a variety of ways:
He can take any number of different flights, he can purchase a ticket in advance or
purchase one only at the airport, he can take an ocean liner, and so on. The general

33 Gauthier has most recently suggested that it is rational to constrain one’s maximizing in
accordance with a prior plan when one can expect to do better under the plan than one would
have done if one had never adopted the plan at all. When this condition is met “in each possible
situation in which it would require some particular action or decision”, Gauthier says the plan is
“fully confirmed”. David Gauthier, “Intentions and Deliberation”, in Peter Danielson (ed.),
notion of confirmation produces the same results as our deliberative requirement. For a plan that
is fully confirmed can be deliberatively open, meaning that the agent need not restrict his rational
facilities in order to implement it. See also Gauthier, “Resource Choice and Rational Deliberation:
sort of normative relation that holds between intentions and actions is ill-defined.34 But endorsing an action as rational requires more than that it is an instance of a type that satisfies the intention in question. The action must make sense against the background of the agent’s plans as a whole and the larger structure of his life. Thus it may be rational to form the intention to go to Paris, but not rational to plan to go there by ocean liner if one is planning only a two-night stay. Which action in satisfaction of a prior intention is rational thus typically depends on the larger structure of which both intention and action are a part.

What we might say, then, is that the background normative relation that holds between intentions and the actions that satisfy them assists in the rationality of intention-execution only insofar as the intention-action pair is embedded in a larger structure that is supported by an agent’s reasons for acting. In the absence of this structure, the relation between intention and action is so weakly normative that it cannot carry the force of reason at all. That is, intention and action are like individual notes in an overall melody: taken separately, neither carries much normative weight. Taken together as a package, however, they may be strongly supported by an agent’s reasons. The intention and the action thus each receive separate endorsement from the agent’s reasons, conditional on their occupying a place in an overall rational plan.

34 Of course it is not strictly a relation between intentions and actions that we want; it is rather act-types on which we must focus. We can speak casually, however, of “actions”, as long as we bear in mind what is actually meant.