

On the Way to the Planning Theory

IN THIS chapter I note two aspects of the desire-belief model, describe several features of intention that will be central to later discussion, argue for the need to take intentions seriously as distinctive states of mind, and describe certain problems that will shape later discussion.

2.1 Two Aspects of the Desire-Belief Model

It will be useful to begin by saying more about the desire-belief model. On this view what makes an action intentional, or done with a certain intention, is the fact that it stands in an appropriate relation to the agent's relevant desires and beliefs. The relation in question is typically taken to be explanatory in some way. Suppose, for example, that Eve intentionally turns on the air conditioner. This action is intentional because it is explained, at least in part, by the fact that Eve wants to cool off and believes in the efficacy of air conditioners. Such an explanation of Eve's action has an important feature, one that distinguishes it from an explanation of Eve's sweating by an appeal to the fact that she is hot. What is special about the former explanation is that Eve's desire and belief explain her action in part because they provide Eve with a *reason* so to act. In seeking such an explanation of Eve's intentional conduct we are trying to see that conduct as at least to some extent evidencing Eve's rationality; we are trying to see Eve and her action as at least to some extent rational.

In developing a model of intention and action we are quickly led into questions about practical rationality: What makes actions rational? What makes it rational of agents to act as they do? These questions are normative in nature: they are queries about how an agent ought, rationally speaking, to act. As commonly understood, the desire-belief model goes on to provide a powerful and plausible approach to such questions of

practical rationality. The main idea is that the agent's desires and beliefs at a certain time provide her with reasons for acting in various ways at that time. What practical rationality requires is that her intentional action be at least as strongly supported by these desire-belief reasons as any of its supposed alternatives. On this view, then, the rationality of intentional action is primarily a function of the agent's desire-belief reasons for action: only these desire-belief reasons have *direct relevance* to the rationality of action; other considerations bear on the rationality of conduct only *indirectly*, by way of their impact on the agent's desires and beliefs.

To change examples: My desire to get a copy of *Actions and Events*, together with my belief that I can do this by going to Tanner Library, provides me with a reason for going there. And my desire to watch the Giants on television, together with my belief that I can do this only by staying home, provides me with a reason for not going to Tanner. On the desire-belief approach just sketched my intentionally going to Tanner is rational only if it is supported by the balance of such reasons.

So the desire-belief model, as I will be understanding it here, has both descriptive and normative aspects. It attempts to capture the basic structure of the conception of mind implicit in our commonsense understanding of intention and action. And it also attempts to articulate an associated, normative conception of practical rationality.

2.2 Intention and Commitment

Return now to commitment. Intention, I have said, involves a characteristic kind of commitment. A full characterization of the commitment involved in intention is one of the main jobs of a theory of intention. Here I will provide just an initial characterization.

The first step is to distinguish two dimensions of commitment. The first concerns the relation between intention and action, and I will call it the *volitional* dimension of commitment (or, for short, volitional commitment). To explain what I have in mind here I need to introduce some terminology. Both intentions and desires are, but ordinary beliefs are not, *pro-attitudes*. Pro-attitudes in this very general sense play a motivational role: in concert with our beliefs they can move us to act. Thus, both a desire to go to Tanner Library and an intention to go could, in concert with a belief that a certain bus will take me there, lead to my taking that bus.

But we may go on to distinguish between two kinds of pro-attitudes. To change examples, suppose I desire a milk shake for lunch, recognize that the occasion is here, and am guilty of no irrationality. Still, I might

not drink a milk shake; for my desire for a milk shake still needs to be weighed against conflicting desires—say, my desire to lose weight. My desire for a milk shake potentially influences what I do at lunchtime. But in the normal course of events I still might not even try to drink a milk shake.

In contrast, suppose that this morning I formed the intention to have a milk shake at lunch, lunchtime arrives, my intention remains, and nothing unexpected happens. In such a case I do not normally need yet again to tote up the pros and cons concerning milk-shake drinking. Rather, in the normal course of events I will simply proceed to execute (or, anyway, try to execute) my intention and order a milk shake. My intention will not merely influence my conduct, it will control it.

This example illustrates an important difference in the normal functioning of desires and intentions concerning what to do in present circumstances. Intentions are, whereas ordinary desires are not, *conduct-controlling* pro-attitudes. Ordinary desires, in contrast, are merely *potential influencers* of action. The volitional dimension of the commitment involved in future-directed intention derives from the fact that intentions are conduct controllers. If my future-directed intention manages to survive until the time of action, and I see that that time has arrived and nothing interferes, it will control my action then. As a *conduct-controlling* pro-attitude my intention involves a special commitment to action that ordinary desires do not.

Turn now to the second dimension of the commitment characteristic of future-directed intention. Here what are crucial are the roles future-directed intentions play in the period between their initial formation and their eventual execution. For reasons that will become apparent I will say that these roles constitute the *reasoning-centered* dimension of commitment, or simply, for short, *reasoning-centered* commitment.

There are two main facts here. The first is signaled by talk of being *settled* on a certain course of action. If I now intend to go to Tanner later today, I normally will not continue to deliberate about whether to go. I will normally see (or, anyway, be disposed to see) the question of whether to go as settled and continue so to intend until the time of action. My intention resists reconsideration: it has a characteristic *stability* or *inertia*.

Of course, given new information, or a change in what I want, I may well reopen the question and reconsider. As Austin emphasized, "Every intention . . . which regards the future, is ambulatory or revocable."¹ But revocability does not entail actual reconsideration. Lacking new considerations I will normally simply retain my intention up to the time of

action. Retention of my prior intention and nonreconsideration is, so to speak, the "default option."

The second fact concerns the role my intention normally plays in my further reasoning between now and the time for going to Tanner. I will frequently reason from such a prior intention to *further* intentions. I will frequently reason from intended end to intended means or preliminary steps: as when I reason from my intention to go to Tanner to intentions concerning how to get there. And I will frequently reason from more general to more specific intentions: as when I reason from an intention to take a bus to Tanner, and my reflections on the bus schedule, to an intention to take a particular bus. Further, my prior intention to go to Tanner this afternoon will constrain the other intentions I form for the day, since I will seek to make my intentions consistent with one another and with my beliefs.

My intention to go to Tanner this afternoon involves, then, certain characteristic dispositions concerning my reasoning between now and this afternoon: a disposition to retain this intention without reconsideration, and a disposition to reason from this retained intention to yet further intentions, and to constrain other intentions in light of this intention. These dispositions partly constitute the reasoning-centered dimension of the commitment characteristic of future-directed intentions.

So there are two dimensions to the commitment characteristic of future-directed intention: volitional and reasoning-centered. Of course, these two dimensions of commitment are closely related. When I decide this morning to go to the basketball game tonight, my intention will play its volitional role only if it first plays its role in reasoning: I must retain this intention, go on to figure out how to get tickets and get to the game, and also be careful not to schedule incompatible activities. And it is because of its ultimate volitional role that the inertia of intention, and its role in further means-end reasoning, is so important to us. Thus the dispositions behind both kinds of commitment are more or less directly tied to action. But this should not lead us to ignore the fact that some of these dispositions are, in the first instance, dispositions concerning reasoning.

A related and important point is that the combination of both dimensions of commitment involves a kind of synergy: taken together these two dimensions of commitment help explain how intentions play their characteristic role in supporting coordination, both intrapersonal and social. Both the inertia of intention and the fact that it is a *conduct-controlling* pro-attitude provide support for the expectation that when the time for action comes, an agent will at least try to do what she intends to do.

Further, the dispositions to figure out how to do what one intends, and to settle on needed preliminary steps, provide support for the expectation that an agent will both be in a position to do what she intends and succeed in doing it. And it is such expectations as these that are central to coordination—both interpersonal and intrapersonal. Because my intention to go to Tanner supports your expectation that I will be there, you can go ahead and plan to meet me there. And because it also supports my expectation that I will be there, I can go ahead and plan the rest of my day on the assumption that I will be there. For example, I can plan on reading a book tonight that I can only get at Tanner. So both dimensions of commitment help explain why an intention to *A* normally supports a belief that the agent will *A*. And this belief helps facilitate coordination.²

2.3 The Initial Challenge to the Desire-Belief Model

These characterizations of volitional and reasoning-centered commitment are, so far, only partial. We have focused on the underlying regularities and dispositions associated with each dimension of commitment. We have not yet explored the implications of these phenomena for our normative conception of practical rationality. But even at this stage there is a challenge to the descriptive aspect of the desire-belief model; for it seems doubtful that these complex phenomena can be adequately described while staying strictly within the desire-belief framework.

At issue here is thesis (4), the reduction of future-directed intention to appropriate desires and beliefs. A natural first step in the direction of such a reduction might go by way of the notion of a predominant desire. I have a predominant desire to *A* if I desire to *A* strictly more than I desire to perform any option thought by me to be incompatible with my *A*-ing. But it seems clear that my intention, say, to go to Tanner, cannot be identified just with such a predominant desire; for such a predominant desire does not ensure either volitional or reasoning-centered commitment.

Consider reasoning-centered commitment. A crucial point here is that even if I now have a predominant desire to go to Tanner—and so prefer to go there—I still might not see the issue as *settled*: I might be disposed to continue to give serious consideration to the possibility of taking the afternoon off and going to a concert. But if I were to intend to go to Tanner, I would be disposed not to continue to deliberate in this way: this is what is involved in the resistance to reconsideration characteristic of intention. For similar reasons my predominant desire does not ensure

that I have an appropriate disposition to reason about means to going to Tanner. After all, if I am still deliberating about whether or not to go to Tanner, I may well have as yet no disposition at all to reason about means to going.

Turning to volitional commitment, there is an analogous problem with the reduction of intention to predominant desire. An intention to *A* is a conduct-controlling pro-attitude. But my predominant desire to go to Tanner at noon does not guarantee that when I see that it is noon my desire will control my conduct. I might still be disposed to deliberate about what to do; for I might still not see the issue as settled.³

A final point is that if intention is seen as reducible to such a predominant desire, we will misdescribe the role of a prior intention in constraining other intentions, given the search for consistency. In requiring that I desire to go more than I desire to perform actions thought incompatible with going, the account makes it impossible to intend to perform actions thought incompatible. But while it is, other things equal, criticizably irrational to have such conflicting intentions,⁴ it does seem possible to be guilty of such irrationality. And we want our theory to allow for this possibility.

Can these objections be overcome by the addition of further conditions? A natural proposal is that to intend to go to Tanner is not only to have a predominant desire to go, but also to believe that one will go because of this desire to go.⁵ One worry here—one I will be discussing further in Chapter 3—is that it is too much to require that an intention to *A* involve a belief that one will. A second worry is that the addition of this other condition by itself does nothing to block the objection, already mooted, that we have made conflicting intentions impossible, rather than merely irrational. But here I want briefly to consider a different aspect of the question of whether such an account captures the commitment characteristic of intention.

Does the addition of the further belief condition guarantee that I see the issue of whether to go to Tanner as settled in the relevant sense—a sense that precludes a disposition for continued deliberation about whether to go? Suppose that I presently have a predominant desire to go to Tanner and (knowing my own work habits) expect that as a result of this desire I will go to Tanner. Could I nevertheless continue to be disposed to deliberate about whether to take the afternoon off? Suppose that I suspect that my predominant desire to go to Tanner is a result of my workaholic tendencies and want to reflect further on the matter. When I step back and try to make a prediction about what I will do, I continue to expect that I will end up going to Tanner. Still, I am suspicious of my motivation and want to think about it some more. Am I settled on going to Tanner,

in the sense of being settled that is involved in reasoning-centered commitment? I am inclined to say no.

Another case. Suppose I have a fleeting craving for a chocolate bar, one which induces a fleetingly predominant desire to eat one for dessert. And suppose that just as fleetingly I notice this desire and judge (in a spirit of resignation, perhaps) that it will lead me so to act. But then I stop and reflect, recall my dieting plans, and resolve to skip dessert. On the present desire-belief account I had a fleeting intention to have a chocolate dessert. But I am inclined to say I had no such intention, for I was never appropriately settled in favor of such a dessert.

Such objections are not conclusive; nor do they do justice to all attempts to defend reductions along the lines I have sketched. But I do think such objections are plausible. And they help raise the basic question of why we should seek such a reduction in the first place. Presumably such a reductive approach to intention should be judged by its theoretical fecundity. But I believe that in the end such attempts at reduction do not promise to tell us a great deal about how intentions function in the lives of rational agents like us. I think we gain more insight into the kinds of agents we are by putting aside such attempts at reduction and taking seriously the idea that intentions are distinctive states of mind, on a par with desires and beliefs. Intentions are conduct-controlling pro-attitudes, ones which we are disposed to retain without reconsideration, and which play a significant role as inputs into reasoning to yet further intentions. I propose to consider this network of dispositions and functional roles on its own terms, without trying somehow to reduce it to ordinary desires and beliefs. In the end the main argument for this richer conception of mind and action lies in its theoretical usefulness: the proof of the pudding is in the eating. The main burden of much of this book is to demonstrate the theoretical fecundity of this more complex conception of mind and intelligent agency.

2.4 The Modest Extension of the Desire-Belief Model

Having departed this much from the desire-belief model, we are faced with a further issue, an issue about practical rationality. The issue arises because, as we have seen, the desire-belief model has both descriptive and normative aspects. In granting intentions status as distinctive attitudes, we reject as inadequate the descriptive aspect of the desire-belief model. But we do not yet reject its normative aspect: the powerful and extremely plausible conception of practical rationality as, roughly, a function of desire-belief reasons for action. Should our rejection of the de-

scriptive aspect of the desire-belief model bring with it a rejection of its conception of practical rationality as well?

At this point a desire-belief theorist may be attracted to a conservative strategy. This is to grant intentions status as distinctive states of mind and yet hold onto the desire-belief conception of practical rationality. On this strategy one insists that, unlike desires and beliefs, an agent's intentions do *not* have direct relevance to the rationality of her actions. Intentions are distinctive states of mind, states that play distinctive causal roles in the connection between deliberation and action. These causal roles include those described in our discussion of the volitional and reasoning-centered dimensions of commitment. But in playing these roles intentions do not provide considerations that are directly relevant to the rationality of the actions in which they issue.

Of course, it can still be granted that intentions (like anything else) can be *indirectly* relevant to what it is rational to do. We may distinguish three ways in which this might happen. First, intentions may have indirect *practical* relevance. This would happen if the agent's desires concerned the realization of earlier intentions. For example, she might just desire to stick to her guns or to improve her reputation for steadfastness. Second, intentions may have an indirect *epistemic* relevance. The agent might see her prior intention to *A* as evidence that she will *A*, and so take her *A*-ing for granted (or, anyway, assign it a high probability) in her further reasoning. Third, intentions may have an indirect *second-order* relevance. The agent might see her prior intention to *A* as evidence that *A*-ing is in fact favored by the balance of her present desire-belief reasons; or as evidence that the costs of reconsidering that intention would not be worth it, from the standpoint of her present desires and beliefs. In all such cases, however, the view in question insists that the relevance of a prior intention to the rationality of action is at most *indirect*: it goes by way of the agent's desires and beliefs.

I call such a view a *modest extension* of the desire-belief model; for it retains that model's conception of practical rationality, even while rejecting the underlying psychology. While modifying the descriptive aspect of the model, it retains its normative aspect. I am going to argue, however, that this is an unstable compromise. Once we take seriously the roles of intentions as inputs to reasoning to yet further intentions, we are, I think, forced to complicate our normative account of practical rationality as well. But before explaining why, let me note a limitation of the present discussion.

There are various more traditional ways in which philosophers have challenged the idea that it is the agent's desires and beliefs that provide considerations that are directly relevant to the rationality of her actions.

For example, one might argue that only desires that would remain after full rational reflection on available information provide such direct considerations.⁶ Or one might urge, in an Aristotelian spirit, that only facts about what things are part of or a means to an objectively good human life can provide such direct considerations. These challenges, and others in a similar spirit, are deep and important. But I shall put them to one side here; for they do not address the particular problems about practical rationality that I think are raised by taking intentions seriously. Even a person of Aristotelian practical wisdom, with only fully rational desires, will—if she is like us—form future-directed intentions of a sort that will raise the problem I want to focus on, a problem that challenges the modest extension of the desire-belief model.

So let me say what I think this problem is. Intentions, we are now supposing, are states of mind on a par with desires and beliefs. We have identified three kinds of dispositions associated with such states. They are conduct-controlling—and not merely potentially conduct-influencing—pro-attitudes; they resist reconsideration, and so have a characteristic inertia; and they play characteristic roles as inputs into further practical reasoning to yet further intentions. In each case we could ask about the impact on our normative conception of practical rationality; and I will have a great deal to say later about such matters. But for now let me focus only on this third type of disposition: the disposition to enter as an input into further practical reasoning.

Consider my intention in January to go to Boston in April. Having formed this intention I might go on to reason from it to a more specific intention (for example, to go during the first week of April) or to a further intention concerning means (for example, to take a United flight) or to a further intention concerning preliminary steps (for example, to ask Jean to cover my teaching responsibilities). For simplicity let us focus on the case in which I reason to an intention concerning means. In such reasoning I begin with my intention to go to Boston and deliberate about how. I treat my intention to go to Boston as *directly* relevant to the rationality of the further intention concerning means that I reach in such reasoning. I see it, for example, as directly relevant to the rationality of my decision to take a certain United flight. And when I actually take the United flight, I see the fact that so acting is a means to what I intend (going to Boston) as directly relevant to its rationality. I see my prior intention to go to Boston as directly relevant to the rationality of both my later intention to take the United flight and my eventual action of taking that flight. But this understanding of the role of my prior intention is in conflict with the view that only desire-belief reasons could have such direct relevance.

The problem is particularly clear in Buridan cases, cases in which I

arbitrarily form an intention in favor of one of several incompatible but (so far as I can see) equally desirable options. My desire-belief reasons in favor of taking route 101 to San Francisco may seem on reflection equal in weight to those in favor of route 280. Still I must decide. As it happens, I decide—albeit arbitrarily—in favor of 101.⁷ Now I must figure out how to get there: I reason from my intention to take 101 to an intention to turn right at Page Mill Road. In this means-end reasoning I treat my prior intention to take route 101 as directly relevant to the rationality of my derivative intention to turn right at Page Mill. That is why I see my intention to turn right as rational and an intention to turn left (toward route 280) as irrational, whereas before I formed the intention to take route 101 this was not so. Having decided to take 101 I think that I ought to turn right at Page Mill; but before I decided to take 101 I did not think this. But my thought that my prior intention to take 101 is directly relevant to the rationality of my turning right is in conflict with the desire-belief conception of practical rationality.

There is, then, a serious tension between an agent's commonsense understanding of such means-end reasoning and this desire-belief conception. To the extent that we take this commonsense understanding seriously we should find the desire-belief approach problematic.

Here is another way of looking at the problem. Associated with the desire-belief conception of practical rationality is a model of practical reasoning as the weighing of desire-belief reasons for and against various options. And such weighing is doubtless one common form of practical reasoning. But once we take intentions seriously as distinctive attitudes, we must also come to terms with reasoning concerning means to intended ends.⁸ In such reasoning we treat our prior intention as having a sort of direct relevance to the rationality of further intention and action that is, on the desire-belief approach, reserved to desire-belief reasons. We need a satisfactory theoretical account of how these two kinds of reasoning are related. Barring a successful demonstration that we are just confused in seeing prior intentions in this way, the search for such an overall account forces us to go beyond the modest extension of the desire-belief model.

2.5 The Intention-Based-Reasons View

How then are we to understand the way in which prior intentions provide considerations that are directly relevant to the rationality of derivative intentions and actions? One approach is just to build on the reasons-for-action structure provided by the desire-belief model, but to supplement

its account of reasons for action with *intention-based reasons*. The idea is that not only do our desires and beliefs give us reasons for action; so also do our intentions (together with our beliefs). Just as there are desire-belief reasons for action, there are also intention-based reasons for action. Intentions are not only distinctive states of mind; they also provide reasons for action, reasons over and above ordinary desire-belief reasons. It is by providing these reasons for action that intentions provide considerations that are directly relevant to the rationality of the conclusions of means-end reasoning.⁹

For example, my intention to go to Boston provides a reason to take the United flight, a reason over and above my relevant desire-belief reasons. It is this reason that I recognize when I treat my prior intention as directly relevant to the rationality of an intention to take that flight. Again, in our Buridan case my intention to take route 101 gives me a reason in favor of turning right at Page Mill, a reason over and above my desire-belief reasons. And it is this reason I recognize when I see my prior intention as having direct relevance to the rationality of my turning right there.

Seeing intentions as reason-giving in this way would help explain how intentions could be directly relevant to the rationality of derivative intentions and actions. But as stated this view seems to me doubly flawed. The role it provides for prior intentions as inputs into further practical reasoning is in one way too weak, and yet in another way too strong.

First, it is too weak because it treats my intention to go to Boston as just one reason among many—including many desire-belief reasons—that weigh for and against taking the United flight. But when I reason from intended end to means I do not see my intended end in this way. Rather, I take this end as fixed (in a sense that needs to be made clear) for the purposes of my deliberation, and proceed to try to figure out how to achieve it. Of course, difficulties in discovering acceptable means may force me to give up or, anyway, reconsider my intended end. But unless and until I do give up or reconsider my prior intention, its role in my means-end reasoning will be to set an end for that reasoning and not just to provide one reason among many. As Alan Donagan once put it, to see my intention as providing just one reason among many is to fail to recognize the peremptoriness of reasoning-centered commitment.¹⁰

Second, the intention-based-reasons view is too strong; for it leads us to sanction unacceptable forms of bootstrapping. To see why, let me introduce an example I will be using quite a bit in the discussions to follow: the example of Walter Mondale before the second Presidential Debate with Ronald Reagan. Let us suppose that prior to the debate Mondale considers whether to attack Reagan's "Star Wars" plan. Mon-

dale realizes there are serious political risks involved in such an attack. Indeed, let us suppose (what is probably contrary to fact) that Mondale should have seen that in light of his information it would be better to focus on Middle East policy instead; focusing on this different issue is better supported by Mondale's desire-belief reasons than is attacking Star Wars. But Mondale is so annoyed about Star Wars, and so susceptible to wishful thinking, that he gives in to temptation and irrationally decides to attack that plan in the debate. Mondale, however, leaves open which of three different questions to use to pursue this issue, waiting to see just how Reagan will act.

Now comes the debate. Mondale still intends to attack Star Wars and, after listening to Reagan, decides in favor of his third question. What are we to say about the rationality of this decision and the ensuing act of asking that third question? Asking the third question is, Mondale knows, a means to attacking the Star Wars program; and so it is supported by Mondale's various desire-belief reasons in favor of attacking that plan. But we may suppose that Mondale still has strong desire-belief reasons in favor of pursuing issues about Middle East policies, and so in favor of asking a different question. And these desire-belief reasons are stronger than those in support of attacking Star Wars. May we then conclude that just as the original decision to attack Star Wars was irrational, given Mondale's reasons for action, so is his later decision concerning means?

Surprisingly, the defender of the intention-based-reasons view may not reach this conclusion. When Mondale comes to decide about means, his situation has changed in a crucial respect from his situation prior to the debate when he reached his initial decision: Mondale now has the intention to attack Star Wars. On the intention-based-reasons view this intention—together with Mondale's means-end beliefs—gives Mondale yet a further reason for asking his third question, a reason over and above his various desire-belief reasons for so acting. So even if the balance of Mondale's desire-belief reasons presently favors asking a different question, once we "add in" his intention to attack Star Wars, the balance of *all* of Mondale's reasons for action may well favor his asking his third question. This means that Mondale's irrational earlier decision to attack Star Wars can bootstrap his later decision to ask his third question into a decision that is rational, all considered. But this seems to let his decision about means (and the ensuing intentional action) too easily off the hook of criticizable irrationality.

One way to appreciate the oddity of such bootstrapping is to note that if Mondale had, prior to the debate, decided *all at once* to attack Star Wars by asking his third question, this entire decision would have been, on balance, irrational. Yet, on the present view, by proceeding in

two steps Mondale can make his later decision about means rational, all considered. And this seems odd.

We must, however, be careful here; for even the modest extension of the desire-belief model will support a kind of bootstrapping. This is because reconsidering and changing a prior intention typically has costs of its own. In certain circumstances these costs are obviously significant; and in the heat of the debate Mondale well knows that he is in such circumstances. In such debates time is quite limited, and audiences tend to interpret hesitation as a sign of weakness. Once in the debate Mondale already intends to attack Star Wars. For him to take a different tack he would have to reconsider and change that intention. And this itself might have—and be known by Mondale to have—nontrivial costs. So Mondale might have corresponding desire-belief reasons against reconsidering that intention. And these reasons might make it rational, relative just to Mondale's desire-belief reasons, for him to ask the third question.

In contrast, if prior to the debate Mondale had decided all at once to attack Star Wars by asking his third question, such desire-belief reasons against reconsideration would not have been available to support the decision to ask the third question. So even on a desire-belief conception of practical rationality we might get a kind of bootstrapping. By reaching his decision in two steps, Mondale might bring it about that his second decision—concerning means—is rational, all considered. And this might be so even if, had he decided prior to the debate and all at once to attack Star Wars by asking his third question, his total decision would not have been rational, all considered.

It is important to see, however, that the intention-based-reasons account justifies yet a *further* kind of bootstrapping. On this account Mondale's intention to attack Star Wars might bootstrap into rationality his intention to ask his third question, even if the latter intention fails to be rational relative to his desire-belief reasons, including those that reflect the expected costs of reconsideration and change of mind. On the intention-based-reasons view Mondale's decision to ask his third question and his intentional action of asking that question may be bootstrapped into rationality by an intention it was irrational to form. And this may happen even though this decision and action would not be desire-belief rational, even taking into account the costs of reconsideration. And this seems too easily to let this decision and action off the hook of criticizable irrationality.¹¹

The desire-belief conception of practical rationality avoids sanctioning such unacceptable bootstrapping. But it fails to do justice to means-end reasoning and other kinds of practical reasoning in which one's prior intentions provide crucial inputs. The intention-based-reasons view at-

tempts to take seriously the way an agent sees his prior intentions as directly relevant to the conclusions of his further practical reasoning. But this view itself faces two difficulties: it mistakenly sees a prior intention as providing just another reason among many, and it leads us into problems about bootstrapping. We need a view of intention that escapes both these objections to the intention-based-reasons view, and yet still goes beyond the modest extension of the desire-belief model in its account of the normative role of intentions in reasoning and action.

2.6 A Look Ahead

Let us take stock. We have isolated three central facts about intentions: they are conduct-controlling pro-attitudes, they have inertia, and they serve as inputs into further practical reasoning. These facts were a source for the challenge to the descriptive aspect of the desire-belief model. This challenge led us to see intentions as distinctive states of mind, on a par with ordinary desires and beliefs. We then turned to the normative aspect of the desire-belief theory. There I have argued that the role of intentions as inputs into further practical reasoning exerts pressure for a change in our normative conception of practical rationality. My conclusion was that the desire-belief conception of practical rationality could not stand unchanged. Yet it turned out not to be obvious how to change it. The most direct form of change—the introduction of intention-based reasons—led to significant problems of its own: especially problems about unacceptable bootstrapping.

We may cast these last problems in terms of the notion of commitment. Future-directed intention involves a characteristic commitment to later conduct. It is natural to interpret this to mean at least that the agent who intends now to *A* later thereby has a reason both for means to *A*-ing and for *A*-ing itself, a reason he would not have if he did not so intend. And it is natural to suppose that it is such a reason that one recognizes when one reasons from a prior intention to an intention concerning means. But we have seen that this approach to commitment is problematic.

In the next five chapters I will try to develop a framework within which we can solve this cluster of problems about intention, practical reasoning, practical rationality, and commitment. Much will depend on the basic idea, noted in Chapter 1, that intentions are typically elements in the larger, partial plans characteristic of limited agents like us. So I will begin by returning to this idea and providing some important detail.