Plans and Practical Reasoning

3.1 Plans

I have emphasized the fact that our intentions concerning our future actions are typically elements in larger plans, plans which facilitate coordination both socially and within our own lives, plans which help enable prior deliberation to shape later conduct. In the search for coordination and effective action we simply are not capable of constantly redetermining, without inordinate costs, what would be the best thing to do in the present, given an updated assessment of the likelihoods of our own and others' future actions. We are not frictionless deliberators. Rather, we settle in advance on prior, partial plans and tend to reconsider them only when faced with a problem. The ability to settle in advance on such plans enables us to achieve complex goals we would not otherwise be able to achieve. This ability to settle on coordinating plans is a kind of universal means: it is of significant use in the pursuit of goals of very different sorts. One of the legacies of the behaviorist tradition in the philosophy of mind is that contemporary theories of action—with some exceptions[1]—have tended to ignore the roles of such plans as inputs into further practical reasoning. But this has been a mistake.

Before proceeding I need to clarify what I have in mind when I talk about plans. The first distinction that needs to be made is between plans as abstract structures and plans as mental states. When I speak here of plans, I have in mind a certain kind of mental state, not merely an abstract structure of a sort that can be represented, say, by some game-theoretical notation. A more colloquial usage might reserve 'plan' for the notion of an abstract structure and 'having a plan' for the notion of a mental state. But this is frequently stylistically awkward. In any case, even after saying this there remains room for misunderstanding; for there are two significantly different cases of having a plan. On the one hand, I might have

only a kind of recipe; that is, I might know a procedure for achieving a certain end. In this sense I can have a plan for roasting lamb whether or not I actually intend to roast lamb. On the other hand, for me to have a plan to roast lamb requires that I plan to roast it. It is the second kind of case that I intend when I speak of plans. Plans, as I shall understand them, are mental states involving an appropriate sort of commitment to action: I have a plan to *A* only if it is true of me that I plan to *A*.

Plans, so understood, are intentions writ large. They share the properties of intentions recently noted: they resist reconsideration, and in that sense have inertia; they are conduct controllers, not merely potential conduct influencers; and they provide crucial inputs for further practical reasoning and planning. But because of their increased complexity (as compared with relatively simple intentions) plans reveal other properties that are crucial to an understanding of reasoning-centered commitment. In particular, the plans characteristic of limited agents like us typically have two important features.

First, our plans are typically *partial*. Suppose I decide this morning to go to a concert tonight. I do not settle all at once on a complete plan for the evening. Rather, I decide now to go to a concert, and leave till later deliberation about which concert to go to, how to get tickets, how to get to the concert in ways consistent with my other plans, and what to do during intermission. Later, as time goes by, I fill in my plan with specifications of means, preliminary steps, and more specific courses of action. Of course I am aware when I first decide to go to a concert that my plan is importantly incomplete. But I know that for now only a partial plan is needed; I can fill it in later as required.[2]

Second, our plans typically have a *hierarchical structure*. Plans concerning ends embed plans concerning means and preliminary steps; and more general intentions (for example, my intention to go to a concert tonight) embed more specific ones (for example, my intention to hear the Alma Trio). As a result, I may deliberate about parts of my plan while holding other parts fixed. I may hold fixed certain intended ends, while deliberating about means or preliminary steps; and I may hold fixed a more general intention, while deliberating about how more specifically to realize it.

The strategy of settling in advance on such partial, hierarchically structured plans, leaving more specific decisions till later, has a deep pragmatic rationale. On the one hand, we need to coordinate our activities both within our own lives and socially, between lives. And we need to do this in ways compatible with our limited capacities to deliberate and process information. Further, given these same limitations we need a way to allow prior deliberation to shape later conduct. This argues for being

planning creatures. On the other hand, the world changes in ways we are not in a position to anticipate; so highly detailed plans about the far future will often be of little use and not worth the bother. Partial, hierarchically structured plans for the future provide our compromise solution.

We not only have relatively specific plans; we also have quite general plans—for example, to pursue an academic career, to raise a family, to work for social justice. Such very general plans—projects, as we might say—structure our lives in a way analogous to the way in which more specific plans for a day structure deliberation and action that day. Of course, these very general plans are radically partial, and need to be filled in as time goes by. But that is a virtue of such plans.

Partiality and hierarchy combine with the inertia of plans to give many intentions and actions a *hybrid* character: at one and the same time, a new intention or action may be both deliberative in one respect and nondeliberative in another. An intention or action may be the immediate upshot of deliberation, and so deliberative. But that very deliberation may have taken as fixed a background of prior intentions and plans that are not up for reconsideration at the time of the deliberation. I may hold fixed my intention to earn a doctorate in philosophy while deliberating about what school to go to, what to write a thesis on, and so on.

It is by way of such plans—plans that are partial, hierarchical, resist reconsideration, and eventually control conduct—that the connection between our deliberation and our action is systematically extended over time. The partiality of such plans is essential to their usefulness to us. But on the other side of the coin of partiality are the patterns of reasoning I have been emphasizing: reasoning from a prior intention to further, more specific intentions, or to further intentions concerning means or preliminary steps. In such reasoning we fill in partial plans in ways required for them successfully to guide our conduct.

These connected phenomena of partial plans and reasoning aimed at filling in such plans are central to our understanding of intention. As I have said, a theory that approaches intention by way of these phenomena is a planning theory of intention. My aim is to sketch a plausible version of such a planning theory.

3.2 Demands on Plans

Plans support coordination and systematically extend the influence of deliberation on later conduct. Let us reflect on demands that plans, other things equal, need to satisfy to serve these roles well.

First, there are *consistency constraints.* To coordinate my activities over time a plan should be, other things equal, *internally consistent.* Roughly, it should be possible for my entire plan to be successfully executed. Further, a good coordinating plan is a plan for the world I find myself in. So, assuming my beliefs are consistent, such a plan should be consistent with my beliefs, other things equal. Roughly, it should be possible for my entire plan to be successfully executed given that my beliefs are true. This is a demand that my plans be *strongly consistent, relative to my beliefs.*[3] Violation of either of these consistency constraints tends to undermine the contribution of my plan to coordination in the world I am in. If I plan both to leave my car at home for Susan and also to drive my car to Tanner Library, all the while knowing that I have only one car, I am unlikely to succeed in my effort at coordination.

Second, there is a demand for *means-end coherence.* Although plans are typically partial, they still must be appropriately filled in as time goes by. My plans need to be filled in with subplans concerning means, preliminary steps, and relatively specific course of action, subplans at least as extensive as I believe are now required to do what I plan. My plan to go to Tanner Library will need at some point to be filled in with a specification of a means to getting there (car? bus? bike?), and may also need to include a specification of appropriate preliminary steps (for example, checking the bus schedule). And my plan to go to a concert tonight needs at some point to be filled in with a specification of which concert. Failure to fill in my plans as needed in these ways will leave them means-end *incoherent.*[4]

Of course, means-end coherence does not require that my plans specify what I am to do down to the last physical detail. Rather, my plans will typically be at a level of abstraction appropriate to my habits and skills. My plan to take the bus to Tanner Library need not include a detailed specification of the foot with which I am to step onto the bus when it comes. Again, my plans need not specify what I am to do in every conceivable future circumstance; for many circumstances will seem too unlikely to be worth planning for. So even means-end-coherent plans will remain partial in significant ways. But means-end coherence does require that my plans be filled in with specifications that are as detailed as is, on my view, needed for their successful execution. That is not to say that they need to be filled in all at once; it is enough that they be sufficiently filled in before it is, by my lights, too late.[5]

So we have two important demands on an agent's plans: they are to be both internally consistent and consistent with the agent's beliefs; and they are to be means-end coherent. Both these demands are rooted in a pragmatic rationale: their satisfaction is normally required for plans to

serve well their role in coordinating and controlling conduct. Both these demands are also defeasible: there may be special circumstances in which it is rational of an agent to violate them. But they are nevertheless important demands, ones whose recognition is central to an understanding of the role of prior plans in ongoing practical reasoning and planning.

The recognition of these demands helps distinguish intentions and plans, on the one hand, from ordinary desires and valuations, on the other. First, we do not normally require our desires to be consistent in these ways. I might, without irrationality, both desire to play basketball today and desire to finish this chapter today, all the time knowing that I cannot do both. This is a sort of conflict that occurs countless times in the life of a rational agent. If, however, my *plans* include both actions then I *am* guilty of a criticizable form of inconsistency.

Second, simply desiring to go to Tanner this afternoon, or finding that prospect desirable, places me under no rational demand to settle on some means to getting there. But if I am planning to go there, I am subject to such a demand.

Intentions are the building blocks of larger plans. Of course, not all intentions are initially formed by way of a process of planning. We sometimes come to have intentions by way of very different sorts of psychological processes. Still, once I come to intend to A, my intention becomes part of my web of intentions and plans, a web subject to the plan-type demands of consistency and coherence. Perhaps I simply find myself with the intention to solve a chess puzzle I stumble on while glancing through the newspaper; I do not arrive at this intention through deliberation or planning.[6] Nevertheless, if I really do intend to solve the puzzle, I am subject to a demand to figure out how to solve it in a way that satisfies the demands of consistency on my total web of intentions.

3.3 The Framework of Prior Plans

So we have two major rational demands on intentions and plans. Associated with these two demands are two direct roles intentions and plans play as inputs in practical reasoning. First, given the demand for means-end coherence prior intentions not up for reconsideration frequently *pose problems* for further deliberation. For example, given my intention to go to Tanner later today I need soon to fill in my plan with a specification of some means for getting there. And to do this I might well deliberate between alternative, conflicting means. Second, given the need for strong consistency prior intentions not up for reconsideration *constrain* further intentions; in particular, they constrain the solutions to the problems

posed by the demand for means-end coherence. If I am already planning to leave my only car at home for Susan to use, and if I do not reconsider this prior plan, then I cannot settle on driving my car to Tanner without violating the demand for consistency. So my intention to leave the car at home constrains the solutions to my problem about means. The option of driving my car to Tanner is, given my prior intentions and beliefs, not an admissible option.

The demand for means-end coherence provides rational pressure for the addition of further intentions. In contrast, considerations of consistency do not by themselves provide pressure for the addition of further intentions. Rather, the need for strong consistency only provides constraints on what further intentions may be introduced.[7]

My prior intentions and plans, then, pose problems for deliberation, thereby establishing standards of *relevance* for options considered in deliberation. And they constrain solutions to these problems, providing a *filter of admissibility* for options. In these ways prior intentions and plans help make deliberation tractable for limited beings like us. They provide a clear, concrete purpose for deliberation, rather than merely a general injunction to do the best. They narrow the scope of the deliberation to a limited set of options. And they help answer a question that tends to remain unasked within traditional decision theory, namely: where do decision problems come from?[8]

They can also force us to consider more complicated options.[9] For example, suppose I intend to lose three pounds by tomorrow and consider drinking a milk shake for dessert. I know that if I have this milk shake I will only be able to lose the three pounds if I skip lunch tomorrow. So considerations of consistency force me to consider the more complex option of drinking a milk shake and skipping tomorrow's lunch.

All this is equally true of intentions formed in Buridan cases. Once I form the intention to take route 101, I am faced with a problem about means, and must filter my other options accordingly. My turning right at Page Mill is an admissible, relevant solution to this problem; my turning left (toward route 280) is not. The fact that my prior intention was in a way arbitrary does not prevent it from playing these roles.

Note that it is not merely that prior intentions in fact tend to lead us to consider certain options and not consider others. Such a causal role might be played by a variety of sensitivities. But in addition to this causal role prior intentions provide a kind of rationale for considering some options but not others, a rationale rooted in the demands for consistency and means-end coherence.[10]

Prior intentions and plans, then, provide a *background framework* against which the weighing of desire-belief reasons for and against various

options is to take place. This framework helps focus deliberation: it helps determine which options are relevant and admissible. Prior intentions provide this background framework directly: the framework does not depend on the agent's seeing her intentions as evidence for something else, or on the presence of some special desire—for example, to stick to her guns. Nevertheless, in playing these roles intentions do not provide reasons for action to be placed on the scale with desire-belief reasons in determining what to do. Their role is to help determine which options are to be considered in the process of weighing conflicting reasons for action, rather than to provide reasons to be weighed in favor of one considered alternative over another. The reasons to be weighed in deliberation remain desire-belief reasons. In this way we go beyond the modest extension of the desire-belief model, giving intentions a direct role as inputs in practical reasoning, just as common sense would suppose. But we do this without following the intention-based-reasons view in supposing that intentions provide reasons for action analogous to those provided by one's desires and beliefs.

So, do intentions provide reasons for action or not? On the one hand, intentions do provide—by way of demands for coherence and consistency—considerations that are directly relevant in deliberation to the rationality of the ensuing intention and action. On the other hand, intentions do not provide reasons that are to be weighed along with desire-belief reasons in favor of one considered alternative over another. The best thing to say is that intentions provide special kinds of reasons—*framework reasons*—whose role is to help determine the relevance and admissibility of options. These reasons do not compete with desire-belief reasons, but rather structure the process of weighing such reasons. Further, this role of intentions in providing a background framework for the weighing of desire-belief reasons is itself grounded in pragmatic considerations concerning the satisfaction of (rational) desire.[11]

A complication is that I may well intrinsically desire to act in a way I also intend to act. This is frequently true about those general intentions and plans—for example, my intention to work for social justice—that are naturally describable as projects. But it can also be true of relatively specific intentions—my intention to hear Pavarotti at the Opera House tonight, for example. In many contexts it may not be important to separate these two attitudes toward so acting. But for the purpose of understanding the structure of practical reasoning we need to make this distinction. My intention to hear Pavarotti partly constitutes the background framework of my further reasoning, whereas my desire to hear him can provide reasons for relevant and admissible options.

As I noted at the end of Section 2.4, a central problem for a theory

of intention is to provide a satisfactory model of the relation between two kinds of practical reasoning: the weighing of desire-belief reasons for and against various options, and reasoning from a prior intention to intentions concerning means, preliminary steps, or more specific courses of action. My solution to this problem is to see prior intentions as elements of plans which provide a background framework within which the weighing of desire-belief reasons is to occur—a framework that poses problems for such further reasoning and constrains solutions to those problems.

Practical reasoning, then, has two levels: prior intentions and plans pose problems and provide a filter on options that are potential solutions to those problems; desire-belief reasons enter as considerations to be weighed in deliberating between relevant and admissible options. This two-level structure is an essential part of the way in which intentions and plans play their coordination-facilitating role, and so part of the way in which intentions enable us to avoid being merely time-slice agents—agents who are constantly starting from scratch in their deliberations. So this two-level structure of practical reasoning has a pragmatic rationale, one grounded in its long-run contribution to our getting what we (rationally) want—given our limits and our complex needs for coordination. We need not leave a broadly instrumental conception of practical reason in order to allow intentions to have direct relevance to the rationality of action.

It is commonly noted that sometimes in means-end reasoning one sees a certain means as necessary to one's intended end, whereas at other times one must choose among several means, each of which would suffice but no one of which is necessary for one's end. Our model of practical reasoning provides a unified treatment of these two types of cases. In both cases one settles on an option so as to avoid a threatened incoherence in one's plans. In the case of reasoning to a necessary means one is presented with what is (at the appropriate level of abstraction) a unique solution to the problem of avoiding this incoherence. In the latter sort of case one is presented with several solutions to this problem, and one needs to appeal to one's desire-belief reasons to determine the best solution. But in both cases one's prior intentions play the same role, namely: the role of posing a problem of coherence and of constraining admissible solutions to that problem.

3.4 Intention and Belief

This discussion of plans and their role in practical reasoning raises a cluster of issues about belief, its relation to intention, and its role in the

background framework against which further planning takes place. Though some of these issues go beyond the scope of this book, I proceed to a brief discussion of some relatively pressing matters.

3.4.1 Flat-out Belief

In explaining the role of intentions and plans in the background framework I have been assuming that there is such a thing as flat-out belief (or, as it is sometimes called, acceptance), not just degrees of confidence or "subjective probabilities" ranging from 0 to 1. Recall my planning concerning how to get to Tanner. My planning proceeds against a background that includes intentions to go there and to leave my only car with Susan. But this background also includes a variety of relevant beliefs, for example: that Tanner Library is not in my house and that I have only one car. It is important to my treatment both of option admissibility and of means-end coherence that these beliefs can be all-or-nothing, flat-out beliefs. If I just assigned a high probability (less than 1) to the proposition that I have only one car and did not simply believe that I have only one car, then a plan to leave a car of mine at home with Susan while driving a car of mine to Tanner would not run into problems of inconsistency. It is my flat-out belief that I have only one car that combines with my prior plans to make inadmissible the option of driving a car of mine to Tanner. And it is because of my flat-out belief that I will not get to Tanner unless I decide between car, bus, and bike that my plan to go to Tanner is threatened with means-end incoherence.

The background framework against which practical reasoning and planning typically proceeds includes not only prior intentions and plans but also such flat-out beliefs. Together these attitudes structure the decision problem addressed in the reasoning. Of course, just as I can always stop and reconsider some prior intention, I can also stop and reconsider some background belief. I might stop and ask whether there is some serious chance that I have more than one car, or that Tanner Library is no longer in the Philosophy Department. And in each case it is possible for such reconsideration to lead to revision. Still, in a normal case in which there is no such reconsideration my planning will be framed in part by my flat-out beliefs that I have but one car and that Tanner is in the Philosophy Department and not in my house.

None of this assumes that there is a simple relation between flat-out belief and degrees of confidence. In particular, it does not assume that to believe flat out that I have only one car I must assign this proposition a subjective probability of 1. If you were to offer me a bet in which I pay one dollar if I own only one car but receive one million dollars if it turns

out that I own a second car, I might well accept this bet; for I judge that there is better than a one-in-a-million chance that, unknown to me, I own a second car. (Perhaps my aunt has just died and left me her car in her will.) Still, though I would take such a bet if offered, I believe flat out that I own just one car. What makes my attitude toward my having just one car one of flat-out belief, and not merely the assignment of some probability somewhat less than 1, is, at least in part, its distinctive role in the background of my further planning—in particular its role in providing a screen of admissibility for my options.[12]

3.4.2 The Asymmetry Thesis

Let us look more closely at flat-out belief and consider how the planning theory should see the relation between intention and belief so understood. Two main ideas have emerged so far. The first is that intentions and plans normally support coordination in part by providing support for expectations that they will be successfully executed. My intention to go to the meeting helps support interpersonal coordination by providing support for your expectation that I will be there, an expectation that will play its role in your decision to come to the meeting. And my intention also supports my own expectation that I will be there, an expectation that allows me to plan my afternoon accordingly. Intentions and plans can provide this support for associated expectations because they are conduct-controlling pro-attitudes, ones that have a characteristic inertia, and ones that play a crucial role as inputs into and constraints on further practical reasoning.

The second main idea is that there is a defeasible demand that one's intentions be consistent with one's beliefs. Violation of this demand is, other things equal, a form of criticizable irrationality.

Note that these two ideas, taken together, still do not entail that an intention to A actually requires a belief that one will A.[13] And, indeed, there is reason to reject such a purported connection between intention and belief as overly strong. Two sorts of examples are relevant here. First, there seem to be cases in which there is intention in the face of agnosticism about whether one will even try when the time comes. I might intend now to stop at the bookstore on the way home while knowing of my tendency toward absentmindedness—especially once I get on my bike and go into "automatic pilot." If I were to reflect on the matter I would be agnostic about my stopping there, for I know I may well forget. It is not that I believe I will not stop; I just do not believe I will.

Second, there seem to be cases in which there is intention in the face

of agnosticism about whether one will succeed when one tries. Perhaps I intend to carry out a rescue operation, one that requires a series of difficult steps. I am confident that at each stage I will try my best. But if I were to reflect on the matter, I would have my doubts about success. I do not have other plans or beliefs which are inconsistent with such success; I do not actually believe I will fail. But neither do I believe I will succeed.

Examples such as these do not *prove* that an intention to A does not require a belief one will A; it remains open to the defender of that view to insist that the intentions in such cases are conditional in some way, or otherwise qualified. But I do think such examples are worrisome enough that I would do well to develop my account of intentions and plans in a way that does not require the strong assumption that to intend to A I must believe I will A. And that is what I will try to do. I will suppose that a normal role of an intention to A is to support an expectation that one will A; and I will also suppose that there is, other things equal, an important kind of irrationality involved in intending to act in ways inconsistent with one's beliefs. But I will not suppose that each and every intention to A involves a belief that one will A.

This clarification in hand, notice an important difference between these two ways in which, on my view, intention is related to belief. An intention to A normally provides the agent with support for a belief that he will A. But there need be no irrationality in intending to A and yet still not believing one will. In contrast, there will normally be irrationality in intending to A and believing one will not A; for there is a defeasible demand that one's intentions be consistent with one's beliefs. Let us label cases of intending to A without believing one will cases of intention-belief *incompleteness*. We can express the difference I have in mind here by saying that intention-belief inconsistency is closer to criticizable irrationality than is intention-belief incompleteness. This is the *asymmetry thesis*.

One good reason for accepting the asymmetry thesis is that intention-belief inconsistency more directly undermines coherent planning than does intention-belief incompleteness. If I intend to go to the bookstore later in the day but am only doubtful that I will, I can make my plans for tomorrow appropriately more complex. I can plan to stop at the market tomorrow if I make it to the bookstore today, and plan to stop at the bookstore tomorrow if I don't stop there today. But if I actually believe I will not make it to the bookstore today, it seems I should be able to plan on the basis of this belief. After all, that is a main role belief plays in ordinary planning. So I will be in a position to plan to stop at

the bookstore tomorrow. But then I will be planning to go to the bookstore twice even though I know I need only go there once.

Consider another example. Suppose there is a log blocking my driveway; and suppose I intend to move the log this morning but believe that since it is too heavy I will not move it. So I need to fill in the rest of my plan for the day accordingly. In filling in my plan for this afternoon I plan on the basis of my beliefs about this afternoon, including my belief that (despite my efforts) the log will still be there. So I add an intention to have the tree company move the log this afternoon. So my plan for the day includes my moving the log this morning and my having the tree company move it this afternoon. But it seems folly to plan to cause the log to be moved twice.

In both cases a problem is created by having an intention concerning the nearer future and a belief that this intention will fail. It is this belief that plays a direct role in the background for planning concerning the further future. But the intentions in which this further planning results are at odds with the initial intention concerning the nearer future, undermining the coordinating role of such planning. And this provides a significant pragmatic rationale for a strong prohibition on intention-belief inconsistency.

In contrast, even though an intention to A normally supports the belief that one will A, the mere absence of this belief in success will not generate such odd consequences. Suppose that I intend to move the log this morning but neither believe I will move it nor believe I won't. When I plan for this afternoon, I am not in a position to plan on the basis of the belief that I will have moved it this morning; but I am also not in a position to plan on the basis of a belief that I will have failed to move it. So I will likely form two conditional intentions: to go to work if I have moved it, and to have the tree company move it if I haven't. And that seems fine. In contrast, a belief that I will fail to move it leads to a much odder plan for the day: to move it, and then to have the tree company move it. And this provides support for the claim that intention-belief inconsistency is closer to irrationality than is intention-belief incompleteness.

There is, however, a further complexity. One might reply to the argument so far by claiming that it depends on an overly simple conception of the role of belief in practical reasoning and planning. The argument has assumed that if I believe that I will fail to move the log this morning, this belief can be a basis for further planning about this afternoon: in planning for this afternoon I can plan on the assumption that I will fail to move the log this morning. But it might be objected that this is not generally true. Rather, in the special case in which you both intend to A

and believe you won't, your belief should not play the role in planning for the further future that beliefs normally play. If you intend to *A* but believe you will fail, your further intentions concerning what to do when you fail to *A* should only be *conditional* intentions so to act *if* you do fail. I should only intend to call in the tree company *if* I fail to move the log. Even though I believe I will fail to move the log, I should not go ahead and intend simply to call in the tree company. So the consequences of intention-belief inconsistency need be no worse than those of mere agnosticism about the success of an intention. In both cases we merely need to construct more complex plans concerning the further future.[14]

This raises a hard question about belief. If my attitude toward my not moving the log does not support planning on the assumption that I will not move it, is my attitude really one of *belief*? I am inclined to answer in the negative. To believe something is not merely to assign a high probability to its occurrence. I might assign a high probability to my failing to move the log without believing I will fail. On the planning theory an important difference between these two attitudes lies in their different roles in my further planning. If I merely think failure quite likely, I am not yet in a position to plan on the assumption of failure. Normally I will go ahead and construct conditional intentions both for success and for failure. In contrast, what seems distinctive about believing I will fail is that it puts me in a position to plan on the assumption of failure. (Of course, I may still want to have a contingency plan for what to do if my belief proves to be false.) But if this is correct, then we should retain the straightforward model of the role of belief in planning, and so reject the cited reply to our argument in favor of the asymmetry thesis.

Another concern that might be expressed here is that intention-belief inconsistency may not be so close to criticizable irrationality as I have been claiming. I can sometimes try to move the log, believe I will fail to move it, and still not be guilty of criticizable irrationality. After all, it is sometimes worthwhile to make an attempt despite one's pessimism. But for me to try to move the log is for me to act in order to move it. To act in order to move it is to act with the intention of moving it. And to act with the intention of moving it requires having the intention to move it. It follows that I may sometimes intend to move the log, believe I will fail, and yet still not be criticizably irrational. And this challenges my claim that such intention-belief inconsistency is normally irrational.[15]

My response is to reject the inference from my trying to move the log to my intending to move it. This may seem counterintuitive; in any case, it is a view that requires some defense. For now, however, I only want to flag this issue: I will return to it in Chapter 9. Until then I will take

as given the asymmetry thesis and the prima facie demand for strong consistency of intention and belief.

3.4.3 Option Admissibility

I now want to add some further remarks about the way in which prior intentions and plans provide a filter of admissibility on options.[16] The basis for this role of prior intentions in further reasoning is the need for consistency in one's web of intentions and beliefs: other things equal, it should be possible for me to do all that I intend in a world in which my beliefs are true. We must, however, be careful not to assume an overly simple relation between this consistency constraint and the nature of this option filter. Not every option that is incompatible with what the agent already intends and believes is inadmissible.

Consider an example. I intend to turn on my computer and I believe that turning it on will heat up my room. I do not intend to heat up my room, however: heating it up is just something I expect to do by turning on the computer. I also believe that turning on the air conditioner would keep the room cool even when the computer is on; but as of now I have no intention to turn on the air conditioner.

Consider the option of turning on the air conditioner. If I were simply to add a new intention in favor of this option to my prior intentions and beliefs, I would introduce inconsistency. In a world in which I execute both my prior intention to turn on the computer and an intention to turn on the air conditioner, one of my prior beliefs will turn out to be false. This is because my prior beliefs include the belief that turning on the computer will heat up the room and also the belief that turning on the air conditioner will prevent the room from heating up. But this should not make the option of turning on the air conditioner inadmissible. It seems perfectly reasonable for me to give this option serious consideration and to do this without in any way reconsidering or bracketing my prior intention to turn on the computer.

In contrast, suppose I intend to use my computer to heat up my room. Again, a new intention to turn on the air conditioner would introduce intention-belief inconsistency. This is because I could not successfully execute both my prior intention and this new intention in a world in which my prior beliefs about the effects of each intended act are true. But in this case this threatened inconsistency does make the option of turning on the air conditioner inadmissible.

Turning on the air conditioner is incompatible with my prior intention and beliefs in both cases. Yet in only the second case is it inadmissible. Why is this? Consider how my web of intentions and beliefs would change

in each case under pressures of strong consistency if I were to decide to turn on the air conditioner. In the first case this decision would force a change in my *beliefs*. I will no longer believe that if I use the computer I will heat up my room. And this change in belief would suffice to prevent the threatened inconsistency. No change in intention would be needed.

In the second case, however, no such reasonable change in belief is available. To avoid the threatened inconsistency, what is called for is a change in my prior intention to use the computer to heat the room. Of course, a change in my belief that the air conditioner would cool off the room is possible. But nothing about a decision to turn on the air conditioner would justify such a change of belief. If I did go on to change this belief, and the rationale for this change was only that it would preserve consistency in my web of intentions and beliefs, I would be guilty of wishful thinking.

This leads to the following understanding of the relation between the demand for strong consistency and the admissibility of new options. Consider a new option, O. Hold fixed the agent's prior intentions, but add to the agent's web of intentions and beliefs a new intention to O. Also add changes in belief that would be justified given that new intention, but without any other revision in the agent's prior intentions. The option O is admissible if these changes in the web of intentions and beliefs would introduce no new inconsistency in that web. What matters for admissibility of a new option are one's intentions *prior to* a decision concerning that option and the beliefs one would reasonably have *after* a decision in favor of that option.[17]

3.5 Internal versus External Points of View

Let us briefly review the discussion so far. The modest extension of the desire-belief model failed to do justice to the direct roles of intentions as inputs to practical reasoning. But the intention-based-reasons view both failed to recognize the special role of intentions and plans in practical reasoning and ran into a bootstrapping problem. We have improved upon the modest extension of the desire-belief model by noting how prior intentions and plans guide and focus deliberation: they pose problems and constrain solutions to those problems. In doing this we have been able to recognize the special nature of the contribution of prior intentions and plans to further practical reasoning: in contrast with the intention-based-reasons view we do not see intentions and plans as contributing just one reason for action among many.

But how do we avoid bootstrapping difficulties? Our account may

seem to be subject to the same worries about unacceptable bootstrapping as those that plagued the intention-based-reasons view. Return to Mondale. As long as he continues to intend to attack Star Wars, his relevant and admissible options will include asking certain questions (for example, the three he actually considers asking), but will not include asking certain other questions (such as a question about Middle East policy). Though Mondale knows he *could* ask a question that bypassed Star Wars in favor of Middle East policies, such an option will be inadmissible, for its performance is known by Mondale to be incompatible with his intended attack on Star Wars. Suppose now that Mondale correctly judges that his asking his third question is better supported by his desire-belief reasons than are its relevant and admissible alternatives. It seems that Mondale should then suppose that this is what it is rational to do, all considered—that this is what he ought on balance to do. But this seems to allow Mondale's prior intention to attack Star Wars—an intention that was irrational when formed—to bootstrap into rationality his derivative intention to ask his third question. So we again seem to be in danger of sanctioning unacceptable bootstrapping.

My response to this challenge is rooted in three main ideas. The first I have already emphasized. An agent's prior intentions play a direct role in his further practical reasoning—a role I have been at pains to elucidate. But the rationale for having attitudes that play such a role is at bottom a pragmatic one, grounded in a concern with the satisfaction of (rational) desire. This leads naturally to the second idea. Mondale's judgment that he ought to ask his third question is made from the *internal perspective of his deliberation,* a perspective within which Mondale's prior plans play the roles I have been emphasizing. But we can also take an *external* perspective in assessing Mondale's asking his third question, a perspective within which we bracket the influence of Mondale's prior plans and from which we are able to note the superiority of Mondale's instead pursuing issues about Reagan's Middle East policies. From this external perspective, roughly, we assess intentional actions solely on the basis of a concern with the expected satisfaction of (rational) desire. Since this concern is foundational, relative to the roles played by prior intentions and plans, criticisms of Mondale's actions from this perspective will have force. In making such external assessments we will need to appeal to standards of practical rationality that are not intended for direct use in the agent's deliberation about what to do. But, and this is the third idea, such external standards will play an important role in a theory of practical rationality that takes intentions and plans seriously. I proceed to expand on these ideas.

Mondale's perspective on his decision to ask his third question is *plan-*

constrained: it is limited to options that get through the filter of admissibility provided by his plans. This is a pervasive and justifiable feature of practical reasoning for limited agents like us. Nevertheless, *we* can step outside of Mondale's perspective, bracket Mondale's prior intention to attack Star Wars, and ask what action would be rational in a *non-plan-constrained* way. From such an external, non-plan-constrained perspective we can examine Mondale's option of asking his third question. And we can determine that this option is, after all, inferior to his pursuing issues about Reagan's Middle East policies—inferior, that is, even relative to Mondale's own desires and beliefs about what he would achieve through each course of action. Of course, this Middle East option is inadmissible for Mondale, given his intention to attack Star Wars. But it is nevertheless an option Mondale believes he *could* perform, and its inadmissibility for Mondale need not stop *us* from considering it and noting its superiority to Mondale's asking his third question.

In reaching such an assessment we are approaching the question of what Mondale ought to do from an *external* perspective. From this perspective we try to determine what course of action is best supported by Mondale's own desire-belief reasons for action, once we bracket the influence of Mondale's prior intentions and plans. Of course, from the internal perspective of Mondale's deliberation, these prior intentions and plans guide deliberation and constrain options. But they do not provide reasons for action in the basic way in which relevant desires and beliefs do. When we step out of the internal perspective of deliberation, then, it is natural to bracket such intentions and plans and try to determine what ways of acting are best supported by Mondale's relevant desires and beliefs, unconstrained by prior intentions and plans. We limit ourselves to options Mondale believes he *could* perform: for we want an assessment of what it would be rational for him to do relative to *his* desires and beliefs. But we need not limit ourselves to those options that are *admissible*, given Mondale's prior intentions. Granted, in attempting to reach such a determination we may need to engage in reasoning more extensive than that in which it would be wise of Mondale to engage, given limits of time and other resources. But never mind. We are not asking whether it would be wise for Mondale so to reason, but rather what it would be rational for him to do, relative only to his relevant desires and beliefs and putting to one side his prior intentions and plans.

To be sure, a defender of the intention-based-reasons view could also observe that once we bracket Mondale's prior intentions, his asking a question about Middle East policies is rational relative to his relevant desire-belief reasons. But this observation would not have the same significance for the intention-based-reasons view as it does for my account.

This is because on the intention-based-reasons view we would see such bracketing as the removal of reasons for action that are, so to speak, on a par with the desire-belief reasons that remain. So we would merely be observing that asking a question about Middle East policies would be rational relative to some proper subset of the agent's reasons for action. And such an observation can be made concerning very many of the actions open to us. There is frequently something to be said for even the silliest alternatives. In contrast, on my account the expected satisfactions of (rational) desires provide reasons for action that are basic in a way in which the framework reasons provided by intentions are not. And that is the source of the significance of a failure to be rational relative just to relevant desire-belief reasons.

So we need to distinguish two points of view from which the rationality of intentional action may be assessed. There is, first, the *internal point of view of the deliberating agent*. It is from this point of view that the agent's prior intentions and plans play their role in providing standards of relevance and admissibility for options. This internal perspective is a *plan-constrained* perspective on rationality. There is, second, the *external point of view* within which the influence of the agent's prior intentions is bracketed, and we seek to determine which options, among those the agent believes he could perform, are best supported by the agent's relevant desire-belief reasons. This is a *non-plan-constrained* perspective on rationality. As the case of Mondale illustrates, there is a clear potential for divergence between the assessments from these internal and external perspectives. An option that is rational relative to the internal perspective of deliberation may fail to be rational relative to the cited external perspective. Mondale's asking his third question was rational relative to the internal perspective of his deliberation at the time of the debate; but from the external, non-plan-constrained perspective we can see the superiority of asking instead a question about Middle East policy.

The possibility of this divergence also sheds more light on Buridan cases. Suppose I arbitrarily decide to take route 101 rather than route 280, even though at the time of my decision these routes seem to me equally attractive. Once I make this decision, my taking route 101 will be rational from my internal perspective, whereas my taking route 280 will not be, for it will be inadmissible. But from the external perspective each option may well remain equally desirable—until I begin driving toward route 101 and away from route 280.[18]

What makes such divergence possible is the divergence between two different constraints on options. From the internal perspective what is crucial is an option's *admissibility* given prior plans. As we have seen, such admissibility is determined by considerations of consistency applied

to the agent's web of intentions and beliefs. From the external perspective, in contrast, what is crucial is whether the agent believes he *can* so act—whether he believes he has it in his power so to act. And these two conditions are significantly different. I might believe I have it in my power to drive my only car to Tanner Library even though the option of so acting is inadmissible given my prior intention to leave my car for Susan. And Mondale believes he has it in his power to ask a question about Middle East policy though this option is inadmissible given his intention to attack Star Wars instead.

So our assessment of Mondale's decision in the debate to ask his third question (as well as of his eventual action of intentionally asking that question) may vary depending on the perspective from which we make it. If we take the internal, plan-constrained perspective of the deliberating agent we may find it rational, all considered; and yet we may still determine that it is, all considered, not rational from the external perspective of non-plan-constrained rationality. Now, the rationale for being planning creatures is, I have urged, ultimately grounded in the long-run contribution of the associated patterns of reasoning and action to our getting what we want. So it is natural to take quite seriously assessments grounded in the external perspective, with its sole emphasis on the agent's desire-belief reasons. This means that Mondale's asking his third question may fail to be rational in an important way, even though from his own present perspective it is rational, all considered. So our theory can provide for a form of negative assessment of such unacceptable bootstrapping.

We may put the point in terms of a distinction between two kinds of ought judgments. In his deliberation Mondale aims at reaching a judgment of what he ought, on balance, to do—where this ought judgment is to be made from the internal, plan-constrained perspective of his deliberation. Call this an *internal-ought* judgment. From the external point of view of non-plan-constrained rationality we seek an, on balance, *external-ought* judgment. Although both ought judgments are relativized to Mondale's attitudes, Mondale's prior intentions and plans play a direct role with respect to the former but not the latter. We should not suppose that either ought judgment is more objective than the other. Both concern a relation between a type of action and certain attitudes of the agent's. Nor should we suppose that one generally takes precedence over the other. Rather, each has its distinctive role to play in our complex practices of deliberation and rational assessment. Internal-ought judgments will be central to deliberation; while the availability of external-ought judgments is one part of our solution to problems about bootstrapping.

*   *   *

This anyway is the basic idea. But I need to clarify what is involved when we take the external, non-plan-constrained perspective. Three features characterize this perspective. First, our interest is in determining what course of action is recommended by the agent's relevant desire-belief reasons. Second, in making this determination we bracket the screen provided by the agent's relevant prior intentions. And, third, in making this determination we are to ignore the costs to us of the reasoning and calculation that are required.

When we bracket Mondale's prior intention to challenge Reagan on Star Wars, we put aside this intention for the purpose of determining what options are to be considered. Options incompatible with this intention need no longer be blocked from consideration. This does not mean that we put aside any *desire* of Mondale's (for example, his desire to limit the arms race) that he sees such a challenge as promoting. Such desires remain in force and continue to provide reasons for various courses of action. But we do newly allow options incompatible with this intention to be considered; we partially lift the screen of admissibility.

Now, prior intentions not only contribute to the screen of admissibility. They also indirectly affect the agent's desire-belief reasons. When we take the external perspective and bracket these prior intentions, should we bracket their influence on these desire-belief reasons as well?[19]

Distinguish three different ways in which Mondale's prior intention to attack Star Wars might affect his desire-belief reasons. First, this intention might provide grounds for beliefs about how he himself will act, for example, a belief that he will attack Star Wars. Second, the presence of the intention might itself affect the costs of his acting differently by adding certain costs associated with the very process of changing his mind; and Mondale may well know this. Third, Mondale may have a desire to stick to his guns, and as a result his prior intention gives him a desire-belief reason for attacking Star Wars.

When we bracket Mondale's prior intention, we will want to bracket its influence on his beliefs about what he will do. We do not want to be able to argue, from the external perspective, that since Mondale is going to attack Star Wars it would be foolish for him to try also to challenge Reagan's Middle East policies! In contrast, it seems to me that we will not want to bracket indirect impacts of the second and third sorts on Mondale's desire-belief reasons. If Mondale supposes there would be costs associated with the very process of changing his mind concerning whether to attack Star Wars, we will want to include such costs in our external comparison of attacking Star Wars and challenging Middle East

policies. And if Mondale in fact has some desire to stick to his guns, we will want to factor that in as well.

This brings us to another complication. We may bracket prior intentions and plans to varying degrees. There are two dimensions along which the extent of such bracketing can vary. The first is rooted in the hierarchical structure of plans. A lower-level element in a plan may be merely a means or a preliminary step to some intended end in that plan; or it may be merely a specification of a more general intention that lies behind it.[20] When this is so, we can bracket the lower-level element without bracketing the intended end or more general course of action that lies behind it in the agent's plan. For example, we might bracket Mondale's intention to attack Star Wars without bracketing his general intention to focus on defense policies. In this way our bracketing may vary in its *depth*. Second, an intention to act in certain ways will frequently be part of a larger plan that concerns the yet further future. For example, Mondale's intention to attack Star Wars may be part of a plan that includes this attack and also an attack on fiscal policy in the next news conference. In bracketing the intention to attack Star Wars we may or may not go on to bracket the entire plan in which it is embedded, including the intention to attack fiscal policy in the next news conference. So our bracketing may vary in its *length*.

When we take the external perspective of non-plan-constrained rationality, how deep and how long should our bracketing be? It would seem that the bracketing should be complete, that we should put aside all prior intentions and plans. But we need not do this all at once. When we take the external perspective, we may begin by bracketing only the prior intentions specifically at issue, for example: Mondale's intention to attack Star Wars. Making appropriate adjustments in the agent's beliefs, we proceed to determine what he should do, relative to his desires and beliefs. We then modestly increase the depth and length of our bracketing, without moving immediately to total bracketing. For example, we include in our bracketing Mondale's general intention to focus on defense policy, and his specific intention to focus on fiscal policy at the next debate. And we proceed again to determine what he should do. And so on. At each stage we make only modest increments in the extent of bracketing. At a certain point we may well have reasonable confidence that no further increases in the depth or length of bracketing would change our verdict. Once we reach such a point, we can stop and say that our verdict at this stage determines what would be, all considered, rational to do from our external perspective.

In Mondale's case it seems plausible to supppose that as we increase the depth and length of our bracketing, the superiority of posing the

question that challenges Reagan's Middle East policies, rather than a question that challenges Star Wars, will be a stable result. Assuming that this is so, we can say that posing the question concerning Middle East policies is superior from the external perspective of non-plan-constrained rationality. And this may be so even if Mondale's posing his third question concerning Star Wars is, on balance, rational from the internal, plan-constrained perspective of his deliberation.