Introduction

Understanding the relationship between the mind and the body requires understanding both the influence of body on mind, as in sensory perception, and the influence of mind on body, as in action. Motions of the body, it seems, are actions only if springing in some way from something in the mind. It seems likely that describing the aspects of the mind from which actions spring, not to mention describing what this “springing” amounts to, will require saying something about trying to act. After all, in many cases trying to act seems to be the first thing that a person does in translating her plans and her aims into action. In order to act, it seems, we ordinarily first have to try. To try to act, however, is not merely to prompt action; it’s also to act. Some behavior, even if only a twitch, is required for an agent to be trying. Thus trying seems to have a foot both in the mental and the physical and so it seems, and has seemed to others, to be the crucial element in the mind-body connection exhibited in action.

The mental state of intention, and its close allies volition and choice, is also a crucial element in the story of the mind’s progress into the world in the form of action. Is trying, however, merely a conglomerate of other things, of, in particular, intention (or its allies) and behavior? Or is the mental component of trying to act something distinct altogether? To put this question in its ontological form, does the furniture of a universe that includes trying to act include more than just intention and behavior? This paper examines this question and argues that one tries to act only if one intends to act. This result strongly suggests that the basic and fundamental element on the mental side of action is intention and that trying to act is nothing
more nor less than intending to act and taking some steps. The result does not indicate this uncontrovertibly, of course. It’s possible that while trying to act always involves intending to act (as will be argued here) it also involves some other mental component (as well as some behavior). But, as we’ll see, the argument for the claim that one tries only if one intends suggests that this possibility is, at best, remote.

This is not a purely abstract issue of import only to an abstruse subfield in the philosophy of mind. Consider the 1997 case of Jones v. State. In the middle of a June night in Fort Wayne, Indiana, Curtis Jones, a member of a gang, went to the home of a rival gang member where a party was taking place. Jones fired four bullets into a crowd of people inside the home. Three of the bullets struck Troy Williams who bled to death. The remaining bullet struck Latrail Gamble who was injured but not killed. Jones was charged with the murder of Williams and the attempted murder of Gamble.

In Indiana, and in much of the rest of the United States, an intention to kill is not required for murder. A person who causes another’s death is guilty of murder in that jurisdiction if he knew his action was likely to kill another person, even if he did not intend to do so. The jury found that Jones did indeed know that it was likely that firing bullets into the crowd would kill someone and so convicted Jones of Williams’s murder. However, in Indiana, attempted murder, by contrast, does require an intention to kill. The jury acquitted Jones of the attempted murder of Gamble. After all, they presumably reasoned, although Jones knew that it was likely that he would kill someone, he didn’t intend to; he intended, instead, to fire his gun, to frighten those associated with the rival gang and maybe even to injure them, but killing, the jury thought, was not his intention, even if he thought it might happen. Or, to put it another way, the jury seems to have thought that Jones could have done all that he intended even if nobody died. But if Jones didn’t intend to kill, then, under Indiana law, he didn’t attempt to. The Supreme Court of Indiana affirmed this pair of verdicts.
As it is in Indiana, so it is in much of the United States: intention to complete a crime is required for the attempt of it, even when the crime in question can be completed without such intent. Without additional propositions concerning the relationship between attempts and completed crimes, there is no contradiction here: why shouldn’t the mental states required for attempt be different from those required for the crime attempted? However, there’s an appealing view of the relationship between attempts and completed acts under which there is a contradiction in Indiana law. Ordinarily, we think that attempts are to completed acts as eggs are to chickens: they hatch into them. But so understood, completed acts are simply attempts plus more; to get a completed action one takes an attempt and adds, for instance, a particular consequence that may or may not have come about. Under such a view, an attempt cannot require anything of the agent that the completed act does not also require. Under Indiana law, Jones succeeded in doing something—namely, kill Williams—that he never attempted; after all, had Williams survived his injury, Jones would have been no more guilty of attempting to kill him than he was of attempting to kill Gamble. But if the view that completed acts are attempts-plus is correct, then Indiana law rests on a contradiction, for under such a view this is flatly impossible.

In fact, in some jurisdictions lawmakers have been sufficiently impressed by reasoning such as this as to broaden laws proscribing attempted crimes so that, if intention to complete the crime is not required for the completed crime, nor is it required for the attempted. In Colorado, for instance, “A person commits criminal attempt if, acting with the kind of culpability otherwise required for commission of an offense, he engages in conduct constituting a substantial step toward the commission of the offense.” So, had Jones done what he did in Colorado, instead of Indiana, he would have been guilty of the attempted murder of Gamble: he knew he was likely to kill someone in firing the shot that hit Gamble and only that “kind of culpability” is required for attempt in Colorado. Legislators in Colorado, then, seem to think that one can try without
intending, but not those in Indiana. Because of a disagreement that can only be thought of as a disagreement in philosophies of action, then, we find vast differences in the treatment of defendants who do not differ at all either in their actions or in the mental states from which those actions spring. However, given the appealing view that criminal punishment is to be issued only to the deserving, and given the just as appealing view that there has been an injustice committed when equally deserving defendants are treated differently, it follows that either defendants in Indiana or those in Colorado are treated unjustly. Which statutory scheme is unjust, it appears, turns on the question of whether or not one can try to act only if one intends to.

To clarify, it is important to note that it is possible that both statutory schemes are unjust or that both are just, all things considered. Say we accept (as I am for the purposes of this paper) the following claim: There are good normative reasons for proscribing the behavior of those who try to act criminally. One can appeal to that claim in support of one’s laws proscribing attempts only if one’s statutory definition of attempt casts its net around all and only the acts of trying to act criminally. Since Colorado’s definition of attempt catches much more in its net than Indiana’s, both schemes cannot be justified on the grounds that through them the state serves the just goal of proscribing the behavior of those who try to act criminally. However, there may be other good reasons for casting the net as Colorado does; in such a case, and on the assumption that Indiana law supplies a definition of attempt to act that aligns with trying to act, both laws would be just. Or, if neither scheme defines attempt in a way that aligns with trying, and there are not good reasons to proscribe what the two schemes do proscribe, then both schemes would be unjust. For our purposes here, however, it is helpful to abstract away from other reasons that one might adduce for proscribing what either state proscribes and make only the normative assumption that there are good reasons to proscribe trying to act criminally. Nothing prevents a state from defining attempt in any way it likes; but only a definition which aligns attempt with trying can, in justifying its law, take advantage of the appealing idea that we don’t want people
to try to act criminally. Our question, then, is what trying is and whether or not it requires an intention to act.

Before moving ahead, it’s important to head off a pair of possible misunderstandings. The first possible misunderstanding concerns what, precisely, I will be arguing for. I am not arguing, merely, that trying to act requires some kind of intention. I am not claiming, for instance, that one cannot try to kill without, say, intending to pull a trigger (or swing a knife, or drop poison in a coffee cup, etc.). Rather, I am claiming that one cannot try to kill unless one intends to kill. Trying to act requires intending the completed act, not merely something distinct from, even if in the neighborhood of, the completed act. So, I will be arguing that even in cases in which, for instance, the completed crime requires an intention to do something, although not an intention to complete the crime, it’s not possible to attempt the crime without intent to complete it. What this implies is that a statutory scheme “in between” Colorado’s and Indiana’s which required some intention for attempt, but not necessarily an intention to complete the crime, would, I will argue, fail for the same reasons as Colorado’s: it draws the net too widely relative to the aim of criminalizing trying to act criminally. In any event, so as to avoid excessively complicated expressions, when the term “intention” is used here without a specification of what it is an intention to do, it should be understood to refer to an intention to complete the act in question.

The second possible misunderstanding concerns the justifiability of holding someone responsible for an act that he did not intend to commit. One might reason like so: If the jury is right that Jones did not intend to kill anyone (although he knew he was likely to), then he did not kill intentionally. But if he did not kill intentionally, then he is not responsible for killing (although he is responsible for those things that he did do intentionally, such as injuring and frightening people). So, the mistake in both Indiana and Colorado law that gives rise to the puzzling legal state of affairs described is not a mistake in those states’ laws proscribing
attempts, but in their murder laws. Their murder laws proscribe unintentional killings, and that’s unjust. Colorado, in order to maintain a coherent legal code, then doubles the mistake by proscribing “attempts” that do not involve intention, and so don’t promise, even if completed, to end in intentional, and thus responsible, action.

This bit of reasoning is flawed, for it rests on a pair of claims, both of which are initially appealing, but at least one of which is likely to be false: (1) An agent can act intentionally only if she intends to so act, and (2) An agent is responsible for her action only if the act was intentional. (1) admits of counterexamples in which agents intend something distinct from, but “close” to, what they end up doing and so do the act intentionally. For instance, say that Jones had intended to kill someone, but had no idea who would be at the party. Did he intentionally kill Williams? It seems so, even though he did not intend to kill Williams, whom he may not have even known to exist, much less expected to be at the party that night. Further, arguably, although not intending to kill anyone, in knowing that he might, and being willing to live with or even welcome that result, Jones thereby killed intentionally. It’s possible, that is, that together with his intention to injure and frighten, his foreknowledge was sufficient to make his act of killing intentional, despite never having been intended.

How many counterexamples (2) admits of will depend on what one’s view of (1) is. If, for instance, you deny that Jones killed Williams intentionally, in the actual case in which he had no intention at all to kill, you might take his act of killing as an example of one for which an agent is responsible despite its being unintentional. Be that as it may, however, everyone must accept that, for instance, people are responsible for results springing from negligence, despite the fact that negligence necessarily involves an absence of any mental representation, much less any foreknowledge, of the results that figure in the description of the agent’s act. It would be an absurdity to suggest that the man who kills a pedestrian after failing to refill his brake fluid necessarily killed the pedestrian intentionally; precisely what makes the case the peculiar tragedy
it is for not just the pedestrian, but also the driver, is that the killing is \textit{unintentional}, and yet one for which the driver very well may never, and ought never, forgive himself.

In any event, in what follows, I will be assuming that both (1) and (2) are false and thus that both Indiana and Colorado have defensible murder laws, although only one of them has a defensible view about attempted crimes. Since, as I’ll argue, trying to act requires intent to act, it is Indiana that has the correct view of attempt.

Section 1 describes and criticize an argument of Michael Bratman’s that implies, if it’s successful, that one can try to act without intending to act. The argument depends on the theory of intention developed by Bratman in his book \textit{Intention, Plans and Practical Reason}, as well as in a number of important papers on the topic.\textsuperscript{7} Bratman calls the theory “the planning theory of intention”. Section 1 begins by examining the planning theory of intention and offering small, but important, modifications to it. This is followed by an argument for the claim that the planning theory, suitably modified, does not in fact imply, as Bratman claims it does, that it is possible for an agent to try without intending. Section 2 goes one step further by employing the planning theory as a premise in an argument intended to show that, in fact, one can try only if one intends. If this argument succeeds, then not only does the planning theory fail to imply what Bratman takes it to imply, it implies the exact opposite. The conclusion of the paper draws out some implications for the law of the view of trying and intention argued for in the preceding sections.

1.

The planning theory of intention starts with the following important insight: what makes the mental state of intention distinct from other practical mental states—mental states, that is, that play a role in deliberation and in the motivation and guidance of action—are the norms of
rationality that govern it. Desires, for instance, are also practical mental states. But an agent has not failed to live up to the standards of rationality when she, for instance, has both a desire and an aversion for one and the same thing. By contrast an agent who intends something and also intends to prevent it from occurring is guilty of irrationality. To put Bratman’s insight another way, we might say that intentions are in an important way belief-like (although Bratman does not think they are to be equated with beliefs). Beliefs, too, are governed by norms of rationality that do not govern other mental states like, for instance, hopes. Someone might hope that something occurs and hope that it doesn’t without being guilty of irrationality; it makes perfect sense to see a possible future state of affairs as both attractive and unattractive, depending on how you look at it. But one who believes that something will occur, and that it won’t, is irrational; that’s just not a set of mental states that it could make sense to have. What is distinctive about both intention and belief is that the having of such a mental state qualifies an agent to be assessed by certain norms of rationality appropriate to the mental state in question.

Bratman calls his theory “the planning theory of intention” because the norms of rationality that govern those with intentions are grounded in the aims of planning agency. Planning serves certain roles in helping us to reach our ends. In particular, it serves a coordinating role, helping us both to mesh what we do with what others do so as to reach joint goals, and to mesh what we do now with what we do later, and did before, so as to reach individual goals that require time-slices of the agent to work together. Thus, stated in its most general form, the planning theory of intention is comprised of the following pair of claims:

(A) An agent who has an intention is subject to assessment by certain distinctive norms of rationality.
(B) The norms of rationality governing those with intentions meet the following condition: by following these norms the agent’s intentions support the forms of inter- and intrapersonal coordination that are available to planning agents.

Notice that as stated the planning theory of intention could be filled out in a variety of different ways, depending on the precise specification of the norms governing those with intentions. In fact, Bratman’s argument for the claim that one can try without intending relies on his view of the particular norms of rationality governing intention, namely what he has called norms of “strong consistency”, “means-end coherence” and “agglomerativity”.

The strong consistency requirement is this: an agent cannot rationally intend to do A if he believes that he will not do A. This is not to be confused with the view that if you intend to do A, then rationality requires you to believe that you will. Bratman's condition is weaker than this, for under it you could intend to do A, fail to believe that you will, and also fail to believe that you won't, without falling into irrationality. Someone who held the view that an agent with an intention to A is not rational unless he also believes that he will A might still subscribe to the planning theory of intention, for he might accept both claims (A) and (B). However, he would not subscribe to the particular version of the planning theory espoused by Bratman according to which only the absence of belief in failure is required to support the forms of inter- and intrapersonal coordination available to planning agents.

The strong consistency requirement has attracted more attention in the literature than the other two requirements and is, therefore, fairly well-understood. The requirement of means-end coherence has been much less carefully examined. It imposes at least two distinct limitations on intention. First, it requires that an agent cannot choose just anything as his means to the achievement of a particular end and maintain rationality; rather, the means must fit the end appropriately or must, at least, do so relative to the agent’s beliefs. The requirement of means-
end coherence, that is, places an agent under rational pressure to select an action as a means to his end only if he believes it to be a possible means to that end. Further, an agent’s intentions are means-end coherent only if he intends all those acts that are believed by him to be necessary means to other acts he intends. This isn’t to say that rationality requires that, for every action an agent intends, he also intends a means to that action; our plans are often left open without being worthy of rational criticism. However, if the agent believes an act to be required if he is to do as he intends, then rationality requires him to intend that act. In fact, the requirement of means-end coherence might impose the stricter requirement that an agent intend all those acts that rationality requires him to believe to be necessary means to his ends, whether or not he actually believes them to be. Imagine an agent who has ample evidence that curing his disease requires amputation, evidence sufficient to convince any rational agent, and who yet refuses to believe it; and imagine that this agent refuses to intend to amputate despite his intention to cure the disease. He might have not just irrational beliefs, but irrational intentions, intentions that fail to be means-end coherent despite the fact that he intends all the acts that he actually believes to be necessary means to the other acts he intends.

Bratman is not as explicit as one might like as to what pressure is imposed by the agglomerativity condition. He says, “Given the role of intentions in coordination, there is rational pressure for the agent to put his various intentions together into a larger intention.” Under one natural construal of this statement, Bratman is requiring more of rational agents than he ought. Say Bratman takes the agglomerativity condition to be as follows:

Agglomerativity 1: An agent cannot rationally intend to A and intend to B unless he also intends both A and B.
Under this construal, the agent must actually formulate an intention to do both A and B in order to rationally intend A and intend B; he must “put his various intentions together into a larger intention”. The problem with this formulation is that it demands mental labor from agents that they often have no need to perform, given their aims. For instance, imagine that I intend to go to Los Angeles tomorrow and to London one year from tomorrow; I have both plans. Does rationality require that I have an intention to do both things? Before answering this question, it is worth pausing momentarily to consider how a conjunctive intention to do both A and B differs from a pair of intentions, one to do A and another to do B. To see the difference, notice that the content of every intention is a single action, even if the action it favors is composed of other actions. Someone who has a single intention to go to LA tomorrow and to go to London a year from tomorrow intends a single act, the act of going-to-LA-tomorrow-and-to-London-a-year-from-tomorrow. What makes this complex event a single action? Perhaps nothing more than the fact that it is the object of a single intention that plays a role in bringing about the behavior in question. Still, the intention favors a single complex act. Thus, under the formulation on offer, the agglomerativity condition requires that we conceive of everything we intend to do as one complex action consisting of many smaller actions and that we intend that grand act. Now return to the question of whether or not rationality requires me to have the conjunctive intention to go to LA tomorrow and to London in a year. Formulating such an intention requires some mental work from me; this alone doesn’t make it irrational, for there may be goods achieved through this work. But are there? No, for I don’t believe and have no good reason to believe, that I will have any trouble at all doing both things; a year is more than enough time to get from LA to London. In short, the mental work of formulating the conjunctive intention is work for which I will not be paid by being brought closer to the achievement of any of my ends. How could rationality require me to do work of that sort?
Still, Bratman is right that there is rational pressure on occasion to “put our intentions together into larger intentions”. This is required of agents who believe that they cannot do all of the particular things they intend, for instance. After all, there’s no tension between an intention to A and a belief that I can’t both A and B. Nor is there tension between that belief and the intention to B. I intend, for instance, to spend the afternoon writing and believe that I can’t do both that and go to the movies; no problem there, for I don’t intend to go to the movies. Still, someone who intends to A and intends to B and believes he can’t do both is guilty of some form of irrationality. What follows is that the belief places pressure on the two intentions despite not placing it on either one individually. Thus there must be some principle of rationality that requires the agent to compare his beliefs not with each of his intentions individually, but with them as a set; it is this idea that motivates the thought that the rational agent “puts his intentions together”. I suggest that this condition ought to be understood like so:

**Agglomerativity 2**: It is irrational to intend A and intend B if the intention to do both A and B would itself be irrational according to some other principle of rational intention.

This formulation of the agglomerativity condition does not require that the agent actually formulate an intention to both A and B if she intends each individually; it doesn't require that the agent put all of her perfectly compatible intentions into larger conjunctive intentions. Hence, it avoids the difficulties encountered by Agglomerativity 1, for it doesn’t require mental work from agents who have no need to perform it. However, Agglomerativity 2 does require that agents avoid having sets of intentions that cannot be combined into conjunctive intentions while maintaining rationality on independent grounds, and so it places agents under pressure to combine their individual intentions, and thus make conflicts evident to themselves, when there is doubt as to the rationality of the conjunctive intentions in question.
Bratman’s position, then, is that what makes a practical mental state an intention is that by virtue of having it the agent is subject to assessment by the norms of strong consistency, means-end coherence and (I will assume) Agglomerativity 2. (From here forward when referring to “the agglomerativity condition” I am referring to Agglomerativity 2.) This view lies at the center of Bratman’s argument for the claim that an agent can try to do something without intending to do it. The argument starts with reflection on examples with the following structure: Consider a person who wants either to A or to B, but not to do both. In order to do this, he decides that he’ll try to do each knowing that his chances of doing both are very low, but that his chances of doing one or the other are much higher if he tries to do each than if he invests his energies in one course of conduct at the expense of the other. There are many examples of this sort. Every year highschool seniors take steps towards attending Yale and take steps towards attending Harvard—they apply to both, pay exorbitant application fees to each, travel for interviews, exert relatives to contact old friends in the admissions offices, etc.—knowing full well that they cannot attend both, and that, even if they could, to do so would be a sure way to fail to realize the benefits of attending either. Bratman’s example is simpler: imagine that there are two video games in which the player is to maneuver missiles towards a target. Further imagine that the two games are wired in such a way that if one hits the target of game 1, then both games shut down. This is similarly true with game 2. Further, were one to get very close to hitting the two targets simultaneously, both games would shut down before either target was hit. An agent who knows all this sits down to play both games simultaneously on the reasoning that if she divides her energies between both games she is more likely to hit one target or the other.

Given Bratman’s rationality conditions on intention, these are examples in which an agent is trying to do each of two things, but does not intend to do either. To see this, start by asking if Bratman’s game-player intends to hit target 1? If the answer is "yes", then by symmetry she must also intend to hit target 2. However, if she intends to hit target 1 and she
intends to hit target 2, and yet she believes that she will not hit both targets, she must have irrational intentions, for (by agglomerativity) an intention to hit both targets would have to be rational for her to maintain rationality; but it is not, for (by strong consistency) she cannot believe that she will not hit both targets given this pair of intentions. However, it is simply manifest that she is trying to hit target 1 and that she is trying to hit target 2. Hence one can try without intending.

The argument is not easy to reject. Notice that neither of two natural ways to reject the argument—by rejecting the strong consistency requirement, or by rejecting the agglomerativity condition—are, in the end, attractive. To see why, recall that the whole point of the rationality conditions on intention is their connection to the forms of coordination that are available to a planning agent. An agent who intends to do things that she flatly believes she will fail to do frequently faces a variety of insurmountable obstacles to the maintenance of both inter- and intrapersonal coordination. In general, the problem is that we can count on people to do as they intend; but we cannot count on people to do as they believe they will not do. For instance, imagine that I intend to pick up my daughter at daycare at 4:00 today, but believe that I will not; perhaps I expect myself to be hopelessly absorbed in my own work come this afternoon. When my wife calls to ask if she should plan to pick up our daughter, what should I tell her? On the one hand, she can count on me to do it (I intend to) and, on the other, she can count on me not to (I believe I won’t). The outcome that we both want—our daughter gets picked up and only one of us makes the trip to the daycare center—would be made more likely if I were to live up to the standards of a rational agent; but given my belief and intention this outcome is no more likely than either of the suboptimal outcomes (we both go to the daycare center, or neither of us does). What follows is that my intention is less than fully rational and it is so precisely because I believe I will not do as I intend. We can just as easily generate cases in which violation of the strong consistency requirement undermines intrapersonal coordination: imagine, for instance,
that I am wondering what to do with myself at 5:00 today. Can I count on having already picked up my daughter? Yes and no, depending on whether I consult my present intentions or my present beliefs.

Obeying the agglomerativity condition seems no less essential to the achievement of the forms of inter- and intrapersonal coordination that are the hallmark of planning agency, and for the same reasons. The fact that an agent intends to A gives us reason to count on her to do so; similarly when she intends to B. But if she believes that she won’t do both, then that gives us reason to think that she will fail to do at least one or the other, and she might do neither. What, in such a case, can we count on her to do? There’s no saying, and so the forms of coordination that require that we can count on agents to act in certain ways will be undermined when agents violate the agglomerativity condition. This is similarly the case if the conjoined intention—the intention to A and B—violates the requirement of means-end coherence by, say, requiring a means that the agent specifically intends not to do. In such a case, the intention not to undertake the necessary means allows the agent and others to count on her not to do it, while the pair of intentions, the intention to A and the intention to B, allows the agent and others to count on her to do so. Again, without being able to count on the agent either to act or not to, coordination is undermined.

Of course, we can think up examples in which an agent does better by violating, rather than obeying, the requirements of strong consistency, agglomerativity or even means-end coherence. A case in point is Gregory Kavka’s “toxin puzzle” in which an eccentric billionaire offers to give you a large sum of money today now to intend to drink a sickness-inducing toxin tomorrow that you know you will have no reason to drink when the time comes. Assuming that you believe yourself, in general, to do only what you prospectively see yourself as having reason to do, you have good reason to believe that you won’t drink the toxin; you’ll already have the
money, so drinking will promise sickness and no reward. Still, you do better—you get the money—if you nonetheless intend to.

Notice that the kinds of cases in which inter- or intrapersonal coordination are undermined when agents violate the strong consistency requirement appear much more common than cases in which violation of that requirement actually furthers such coordination. It is an empirical matter whether this appearance is accurate. But imagine that it were not. Imagine a world in which people are constantly being rewarded when they intend to do things they believe they will not do. Imagine a world, that is, in which the very forms of inter- and intrapersonal coordination that planning helps us to achieve are achieved by those agents who violate the requirement of strong consistency and, for whatever reasons, are not achieved by those who obey it. Imagine further that the violators of the strong consistency requirement are not merely lucky to achieve coordination any more than the man who intends to drink the toxin while believing he will not is merely lucky to get the money; rather, the violators of the strong consistency requirement manage to put themselves into just the mental states that they need to be in, that it makes sense to be in in this world, if coordination is to be achieved. Who are the rational ones in this world, those who follow the demand of strong consistency, or those who violate it?

Which answer to this question one is drawn to is a function of one’s conception of instrumental rationality. According to one conception of the instrumental rationality of a person’s collection of mental states—what I will call the “causal conception”—the rational person is the one whose mental states conform to the model that there is good reason to believe will in fact serve her ends, given the causal structure of the world in which she lives. Under this conception, the violators of the strong consistency requirement are the rational agents, for there is good reason to believe that in the world in which they live it is through such violation that inter- and intrapersonal coordination are to be achieved. According, however, to another conception of rationality—what I will call the “attitudinal conception”—the rational person is the
one whose mental states conform to a model dictated by the characteristic function or purpose of those mental states. To justify a claim to the effect that a particular set of mental states is not attitudinally rational to have we point to the purpose that such mental states are to serve and the way in which a set of mental states of the sort in question tend to frustrate that purpose. To clarify with an example: it appears, just by reflecting on the notion of coordination, that coordination requires people to be able to count on themselves and others to act in certain ways; further, it appears, just by reflecting on the notion of “counting-on”, that we cannot count on agents to act in certain ways if some of their authoritative mental states, such as their intentions, represent them as acting that way, and others, such as their beliefs, as not acting that way. On the attitudinal conception of rationality, that is, the mental states in question impose constraints on agents quite independently of any facts about what will and will not earn the agent goods in the particular circumstances in which he finds himself; the constraints are imposed simply by virtue of the nature of the attitudes in question. Those attitudes, by their nature, serve a particular purpose (such as inter- and intrapersonal coordination), and this purpose dictates the conditions under which such attitudes are rational or irrational; it dictates, that is, the norms of rationality governing agents who have the attitudes in question. Under the attitudinal conception, it is the agents who obey the strong consistency requirement, in our hypothetical world, who are the rational ones, even though, as it happens in that world, they do not achieve any of the forms of coordination with themselves and others that intentions are to make possible. An agent who obeys the strong consistency requirement in this hypothetical world is like a person who takes steps to heal his ailing appendix, rather than have it removed, even when healing it is more painful and promises no special benefit. It’s true that, given the way of the world, a healthy appendix will never serve its natural function. Still, from the point of view of that function, the thing to do is to heal the organ rather than remove it. Similarly, from the point
of view of the function of intention, the thing to do is to obey the strong consistency requirement, even if, given the way of our hypothetical world, this is to follow a less than optimal course.

As a further illustration of the attitudinal conception of rationality, notice that that conception can be used to explain the peculiarity of Moore’s paradox. The man who asserts “p, but I don’t believe it” might be accurately reporting his own beliefs: he might believe both p and that he does not believe that p. Nobody denies, that is, that this is possible. Nor is there any logical contradiction in the contents of the two reported beliefs: one, the belief that p, concerns an extra-mental fact; the other, the belief that he does not believe that p, concerns a fact about his mind. So what’s peculiar about the man’s statement? The answer might be that the following is a norm of attitudinal rationality governing belief: an agent’s belief that p is rational only if she does not also believe that she does not believe that p. Whether or not this really is a norm of the attitudinal rationality of belief will depend on what beliefs are for and whether or not there is a link to be found between obeying this norm and the function or aim of belief. Perhaps, for instance, the point of belief is to help us to reach knowledge and thus to motivate us to search for justifications for the things that we do believe. But such a search for justification requires self-conscious awareness of what one’s beliefs are, and it is precisely this which is undermined by the presence of beliefs that one lacks the beliefs one actually has. Be that as it may, the point worth emphasizing is only that the notion of attitudinal rationality might be precisely what guides our intuition in reaction to Moore’s paradox, an intuition that something has gone wrong in the statement, even if it is far from clear, at first glance, precisely what.

Return to Bratman’s argument, seemingly showing that one can try to act without intending to do so. The argument depends crucially on the claim that the game player, and others like her, would have irrational intentions were she to intend to hit target 1 and intend to hit target 2, given that she believes that she will not hit both. But would her intentions be causally or attitudinally irrational? They would not be causally irrational: the game player has good
reason to believe that the having of both intentions is a good way, in the circumstances, to reach her goal of hitting exactly one target. The same point goes for any similar example: by intending to go to Harvard and intending to go to Yale, the highschool senior increases his chances of going to one or the other. However, the game player and the highschool senior have attitudinally irrational intentions and for just the reasons that Bratman suggests, reasons that have been rehearsed above: the point of intentions is to provide for inter- and intrapersonal coordination, and the norms of strong consistency and agglomeration are linked to such coordination through the notion of counting on someone to act, an essential element in coordination that is undermined when agents violate strong consistency and agglomeration. So once we see that there are two different sorts of rationality informing our intuitions about which hypothetical agents are rational and which not, the assertion that the agent in the example is rational is seen to be equivocal: she is causally rational and attitudinally irrational. So, although she is manifestly trying to hit target 1 and trying to hit target 2, it does not follow that one can try without intending; the game-player very well might have the two intentions in question.

To see the point in sum, it might be helpful to state the heart of Bratman’s argument in syllogistic form like so, where it is assumed that the game player is trying to hit target 1 and trying to hit target 2 and believes that she will not hit both:

1. The game player is rational.
2. If she intends to hit target 1 and intends to hit target 2, then she is irrational.
\[\therefore\] She does not intend to hit target 1 and intend to hit target 2.

What I have suggested is that this is “a syllogism with two middle terms” and so its conclusion doesn’t follow from its premises. The term “rational” in premise 1 refers to causal rationality, while the term “irrational” in premise 2 refers to attitudinal rationality.
Bratman can resist this argument by insisting that the game-player is attitudinally rational or, in other words, by insisting that “rational” in premise 1 refers to attitudinal rationality. However, it would be a mistake to support this assertion merely by appeal to intuition. Our intuitions about rationality are not terribly fine-grained and are influenced by various irrelevant factors such as the way in which an example is described. Say, for instance, that one tells the story of two game players who happen to be named Hesperus and Phospherus. Hesperus is trying to hit target 1, Phospherus target 2. Each knows that both targets can’t be hit and each is extremely confident in both his own abilities and his superiority to the other. Further, they are working together, aiming to make it the case that between them one target is hit, by having each of them do their utmost to hit their respective targets. They assume, and let’s assume they are right, that the chances of them hitting the two targets simultaneously are inordinately slim, and that the chances of one or the other hitting a target would reduce drastically were each to do anything less than his utmost to hit his assigned target. I think it would seem to most people that each is rational and that each intends to hit his respective target. Why does anything change when it turns out that Hesperus and Phospherus are one and the same person?

Further, and more importantly, it is possible to support the claim that the game player in Bratman’s example is attitudinally irrational with reasons, and not just with intuitions. Where we have the violation of a principle of rationality there is a call for explanation, a need to say how what is before us makes sense. Imagine that you are given only the following pieces of information: the game player is trying to hit target 1 and trying to hit target 2; he wants to hit at least one target; he believes that he cannot hit both. Now if the player’s behavior is to make sense, we need an answer to the following question: “Why is he trying to hit each target, rather than just trying to hit one or the other?” The answer, of course, is that he thinks it more likely that he’ll hit one if he tries to hit each than if he tries to hit only one. But the answer isn’t the point; the point is that the question needs answering if we are to find the game player’s conduct
intelligible. Further, and importantly, we wouldn’t need an answer to that question if the game player did not believe that he can’t hit both targets. What this shows is that given that the agent in the example is trying to hit each target, he is under rational pressure not to believe that he can’t hit both and, importantly, that pressure goes away only by imagining alterations in his mental states, such as absence of the relevant belief. The presence of this pressure needs explanation and it is provided by the presence in the game player of the intentions to hit the targets, for where there is intention there are the pressures under which the game player labors.

Notice that we can’t explain the pressure to explain himself by appeal to different intentions from the intentions to hit each target. If, for instance, the game player merely intended to try to hit target 1 and intended to try to hit target 2, there would be no pressure to explain, for he believes that he can do both of those things, just as he believes that he can take all the steps short of succeeding in his efforts to hit both targets. The same would go for any other retreat in the content of the intentions associated with his tryings. So long as he’s trying to hit each target while believing that he can’t hit both, he owes us an explanation; that debt can’t be explained by appeal to any intentions other than the intentions to each target. (There remains the possibility, to be headed off in the next section of this paper, that the need for explanation is to be explained by the tryings themselves, independently of any associated intentions. But for now, let that pass.)

Bratman is right that we assent to the claim that the agent in the example is rational, but he is wrong about why we assent. We assent not because the game-player is under no pressure to explain why trying to hit each target makes sense in the face of the belief that he can’t hit both but because we see that, although he has crafted his set of mental states in a way that violates norms of attitudinal rationality, he has done so because doing so is the best way to reach his goals in the circumstances. In other words, our assent to the game-player’s rationality derives from our allowing causal rationality to trump attitudinal, in this case, in our all-out assessment of
the agent’s rationality. Whether it would do so across the board is unclear, but it doesn’t matter for our purposes. What matters is that the pressure to explain himself under which the game-player labors indicates that, whatever we think of his rationality all told, he is not attitudinally rational.

It’s important to see that there is a deflationary move that could be made here, and would be attractive to someone skeptical of the very idea of attitudinal rationality. We might say that the only kind of rationality is causal rationality. This is not to say that there isn’t appeal to principles like strong consistency and agglomerativity. However, we might understand such principles not as exceptionless norms of attitudinal rationality, as I, following Bratman, have suggested, but, instead, as rules of thumb specifying how, generally, one’s mental states ought to be ordered if they are to help one to coordinate one’s behavior inter- and intrapersonally. Under this deflationary line of thought, what we learn from the game player example is merely that the norms of strong consistency and agglomerativity admit of exceptions: there are rare cases when it makes sense to violate them. Where I have painted the pressure to explain himself under which the game player labors as speaking to the violation, on his part, of the norms of attitudinal rationality governing intention, the advocate of this deflationary line of thought would suggest that this pressure arises merely from the fact that the game player is not obeying a rule that, in most circumstances, it makes sense to follow; he explains himself to us simply by showing that his circumstances are not of the usual sort, and so are not covered by the rules in question.

The advocate of this deflationary line of thought might support his position by directly attacking the very coherence of the notion of attitudinal rationality. In fact, the notion seems to have a very soft spot: what gives a particular mental state its “aim” or “function”? If the answer were to appeal to, say, the fact that, characteristically, mental states of the relevant sort cause certain attractive results—intentions, typically, cause coordination, for instance—then it might seem that there is no more to attitudinal rationality than causal; the conditions under which a
mental state can be reasonably expected to bring about a certain result are determinative of the norms of rationality that the mental state is to meet. But if the actual causal role of the mental state is not what determines its aim or function, then what does? What makes it the case that one mental state is to play a role in producing inter- and intrapersonal coordination and another is not?

Someone who conceives of the principles of strong consistency and agglomerativity as mere rules of thumb and who supports the position with this sort of attack on the very notion of attitudinal rationality is in partial agreement with me, for such a person accepts that Bratman’s argument fails to show that it is possible to try to act without intending to. However, such a person goes on to insist on something that I deny, namely that Bratman’s game player is rational in every meaningful sense. Under this view, the game player seems irrational at first glance in the same way that someone who harms another seems to be doing something that falls short of maximizing utility. In this last case, we can dispel that impression by indicating how causing harm will result in more utility given the agent’s very special circumstances; analogously, we can indicate how the game player is actually more likely, rather than less, to achieve intrapersonal coordination by violating the principle of agglomerativity given his very special circumstances.

The notion of attitudinal rationality, as a kind of rationality distinct from causal, will, in the next section of this paper, serve as a premise in a positive argument for the claim that one tries to act only if one intends to. So, the deflationary position just sketched is not available to me, despite the fact that both its advocate and I deny that Bratman’s argument shows that claim to be false. What I can hope to do here, however, is merely to acknowledge a debt that, unfortunately, I will not pay fully in this paper: a full defense of the argument to be offered below would require defense of the notion of attitudinal rationality. However, one thing can be said here in defense of that notion: an analogy to biological functions at least gives one pause in
acquiescing to the deflationary claim that principles like agglomerativity and strong consistency are nothing more than rules of thumb. Eyes have the function of helping us to detect the properties of objects that are at some distance from our bodies, and not in immediate contact with them. A large part of the reason for believing this is that, in fact, our eyes cause us to detect remote properties. There is a meaningful sense in which one hasn’t done what’s healthy for one’s self when one gouges one’s eyes out, even if, for some strange reason, this results in an improvement in one’s ability to detect the remote properties of objects. We might say that the function of one’s eyes places one under pressure not to gouge them out, even when gouging them out will help one to realize the very goods that eyes have the function of providing. In similar fashion, the thought goes, the function of one’s intentions places one under pressure not to violate strong consistency or agglomerativity, even when doing so will help one to realize the very goods that intentions have the function of providing. In short, the notion of attitudinal rationality, as something distinct from causal, is no less appealing than a natural and appealing conception of health as linked to the biological function of one’s organs.

In summary, then, we can ask two different questions about the rationality of an agent’s mental states: (1) Do her mental states conform to patterns that fit with their aims or functions? (That is, is she attitudinally rational?) and (2) Do her mental states conform to patterns that there is good reason to believe will further her ends, under the circumstances? (That is, is she causally rational?) There is little reason to think that reflection on the second of these questions can tell us very much about intention or trying. After all, there is so much variation in the circumstances that agents face that there is little reason to expect there to be anything in common among the sets of mental states possessed by the causally rational or the causally irrational. By contrast, we can ask ourselves whether or not an agent’s mental states make her assessible by the norms of attitudinal irrationality appropriate to intention and thus determine if she has an intention, rather than some other mental state. Thus, we have a tool for assessing whether or not an agent has an
intention that is far better than the usual appeal to linguistic or conceptual intuition. Instead of asking, of a particular agent, “Would we say that she intends to A?” we can ask “Would we take her to be irrational were she to fail to meet the demands of strong consistency, means-end coherence and agglomerativity?” What we’ve learned from discussion of Bratman’s views is that it is important that this question is confined to consideration of attitudinal rationality, to the exclusion of causal; if we allow consideration of both we end up concluding that the game-player and others like her lack intention, a conclusion which is unwarranted for the reasons already given.

2.

Arguing against an argument against a claim is not the same, of course, as arguing for the claim. So far it’s been argued only that the planning theory of intention ought not to be construed in such a way that it implies the falsity of the claim that an agent tries to act only if he intends to. This section goes farther by arguing that, in fact, when construed as described in section 1, the planning theory actually implies the truth of that claim. That is, when we take seriously the idea that to have an intention is to be subject to the norms of strong consistency, means-end coherence and agglomerativity, understood as above and understood only as norms of attitudinal rationality, it follows that one tries only if one intends.

At the end of the last section it was suggested that one of the virtues of the planning theory of intention is that it supplies a tool for systematically determining if an agent or hypothetical agent has an intention that does not require appeal to the vagaries of linguistic or conceptual intuition. We need only determine if the agent is subject to the distinctive norms of attitudinal rationality that govern intention, namely strong consistency, means-end coherence and agglomerativity (construed as Agglomerativity 2). However, we now need to look more closely
at this claim. In particular, we need criteria for determining whether or not an agent really is to be assessed by the norms of attitudinal rationality governing intention. Such criteria, however, are not as difficult to produce as one might think. Notice that whenever a person seems to violate a norm of rationality, of whatever sort, he owes an explanation that shows how, at least by his own lights and perhaps by ours as well, his seemingly incongruous mental states or behavior actually make sense. When an agent seems to violate a norm of attitudinal rationality he can meet this explanatory demand, and thus show that, in fact, he is not in violation of the norm, in one of two ways. He can show that either (a) he actually lacks the relevant attitude on the basis of which the norm applies to him, or (and more importantly for our purposes) (b) the violation of the norm of attitudinal rationality actually, in the circumstances, is a means to some end that he possesses and he is, therefore, causally rational.

For instance, if a person announces that he intends to win the lottery we first want to know if he believes it to be fixed, or if he has peculiar beliefs about probability that don’t imply, as ours do, that he won’t. When we find out that, in fact, he, like us, believes that he will not, then we expect some further explanation, for the agent appears to be in violation of the norm of strong consistency. He can, at this point, suggest that he doesn’t really intend to win, but merely to buy a ticket and keep a positive attitude, say. This is to adopt the first approach in explanation of his apparent violation of the norm; it amounts to saying that the norm doesn’t actually apply to him at all, that he is not to be assessed by it. Alternatively, the agent can give us an explanation of how having the intention to win, in the face of the belief that he will not, is causally rational. He might say, for instance, that he’s influenced by social-psychological research that shows that those who intend to win, whether or not maintaining attitudinal rationality, are much more likely to win. He might say that he really doesn’t know why this is—he can’t think of any explanation himself—but he badly wants to win and so sees the having of the intention to win, in the circumstances, as worth the pain of attitudinal irrationality. This is to explain his behavior by
appeal to its causal rationality, its likelihood to serve his ends, given the circumstances, even in the face of its attitudinal rationality.

The question, then, of whether or not an agent who is trying to A also intends to A has reduced to this: Does the agent owe an explanation if his mental states fail to conform to the norms of attitudinal rationality governing intention? Would he owe an explanation if, say, he is trying to A while believing he will fail that either indicates that he is not, despite appearances, trying to A at all, or that trying to A, in the circumstances, and in violation of the norm of strong consistency, is a good way to reach his ends? I believe he would. To see the argument for this claim, consider, first, why we might think that an agent who is trying to act while believing he will fail owes us an explanation of the relevant sort. The argument will extend to the other two norms of attitudinal rationality governing intention.

There are a wide variety of circumstances in which agents try to do things that they believe they will not do and whom we would not pronounce irrational. In a recent paper, Carl Ginet has identified three kinds of case. First: An agent might believe she can’t do something, but also believe that she’s far from infallible in the assessment of her own abilities. She might then try to do the thing she thinks she can’t in order to find out if she’s right. She might do this because she has strong reason to succeed in acting, but she also might do it because she has reasons for wanting to know her own abilities even though she has no reason to succeed in acting. Second: An agent might believe she can’t do something but be challenged by someone else who claims she can. To prove this other person wrong, she might try to do that thing. Such an agent might lack a reason to succeed in acting and might also lack a reason to know her own abilities; she might only, that is, be trying to act in order to prove another wrong. It is worth looking at the example that Ginet uses to illustrate the third sort of case. He writes,
Suppose Sam knows that [a] log is far too heavy for him to move by pushing on it. But Sam is a soldier under the command of Rudy, and Rudy has ordered Sam to try to move the log by pushing on it. Sam would very much prefer that Rudy not find out that he (Sam) cannot move the log by pushing on it, but the penalty for refusing to obey the command would, in Sam’s view, be a worse consequence than Rudy’s knowing that Sam cannot move the log by pushing on it. So Sam pushes on the log as hard as he can in order to comply with Rudy’s command to try to move it by pushing on it, but very much regretting…that by so doing he will reveal to Rudy that he cannot move it by pushing on it.13

By making it part of the example that Sam does not want Rudy to know that he can’t move the log, Ginet distinguishes this case from the second sort of case in which the agent hopes to show another person that he can’t do the thing he is trying to do; Sam’s reason for trying, in Ginet’s example, is his desire to comply with the command. Notice that Ginet has provided a blueprint here for generating an arbitrarily large number of categories of agents who try to act while believing they can’t. Abstractly, Ginet’s example is constructed as follows: Assume that S believes that he cannot A and that he has no reason to successfully A. Assume also that agents sometimes have reasons of types R1, R2…Rn for trying to A that are not also reasons for successfully A-ing (i.e. sometimes they want to see if they can A, despite their belief that they can’t; sometimes they want to prove to someone else that they can’t; etc.). Now imagine that, for some reason, S is positively averse to trying to A for reasons of any of these n types (in Ginet’s example, instead of wanting another to know that he can’t A, he wants that other person not to know that). But he has a stronger reason for trying to A nonetheless (in Ginet’s example, he wants to avoid whatever penalty would befall him were he to fail to obey the command). Using this blueprint, we can generate an arbitrary number of distinct types of example in which an
agent has reason to try to A but, by the nature of the construction, lacks any of the previously identified possible reasons for doing so. The point can be put another way: Since anything can cause anything else, trying to act can cause anything one likes. In order to generate an example of an agent who tries to A while believing he can’t and maintains rationality, we merely set things up so that his trying to act will cause something that he wants more than he wants to avoid any of the other things that trying to act will cause.

The point is that in all of Ginet’s examples and in all of the possible examples that one might construct by following in Ginet’s footsteps, we pronounce in favor of the agent’s rationality in trying to act in the face of belief in failure only because we see how, in light of his ends, it makes sense for him to do so. These are pronouncements of causal rationality, not of attitudinal. They amount to the giving of the second sort of explanation identified above for deviance from the norm of strong consistency; instead of denying that the agents in question are really trying to act, we show how, in light of their circumstances, trying to act in the face of belief in failure is a good way to reach their ends. What follows is that we are yet to find an example of an agent who tries to act while believing he can’t and maintains attitudinal rationality.

To say this much isn’t to say that such agents are attitudinally irrational, but only that our reasons for thinking them rational derive from our recognition of their causal rationality. We need more in order to support the point I’m after, namely that those who try to act while believing they will fail are attitudinally irrational. But to support this point we need an argument for thinking that there’s a link between compliance with the norm of strong consistency and the point or purpose of trying. We found such a link between that norm and the purpose of intending (namely, inter- and intrapersonal coordination) through the connection between coordination and the need to count on those with whom action is being coordinated. Is there a similar link in the case of trying? Yes. To see this, start by noting that the function of intention and the function of
trying must be quite closely intertwined, for the function of trying to act is, in part, to support the function of intending to act. It is by trying that we translate our intentions into behavior. But it is behavior that must be coordinated if we are to reach our individual and joint ends; plans can mesh as much as we like, but until we act on them we don’t get where we want to go. Plans by themselves don’t give us coordination, although they are essential to it; we also need to act on our plans, and it is by trying to act that we do so. Just as intending is a universal means to coordinating our behavior, trying to act is a universal means to acting, and thus to coordinating our behavior (among other things). If we can’t count on ourselves or others to try to act, then we can’t count on them to act; and if we can’t count on them to act, we can’t coordinate our behaviors with them in ways that are essential for reaching our ends.\(^{14}\)

This isn’t yet an argument; it just amounts to the observation that the roles of intending and trying are not so easily disambiguated. However, we are now in position to produce the needed argument: Trying to act serves all the ends that action serves. In fact, it’s function is to enable agents to achieve whatever ends are achieved through action. Now if trying to act is to serve this purpose then we must be able to count on ourselves to act when we try; that, after all, is why we try. But if the norms governing those who intend apply to them because we need to count on them to act, the very same norms apply to those who try. So, if you try, you are assessible by the very norms that apply to you when you intend. Since assessibility by such norms is what makes an agent’s mental state one of intending, it follows that if you try to act you also intend to.

The argument just offered can be put more carefully:

1. If an agent has a mental state the content of which is to-A, and the function of that mental state requires that the agent can be counted on to A, then she is on those grounds
assessible by the norms of strong consistency, means-end coherence and agglomerativity.

\[ [(p \& q) \rightarrow r] \]

2. If an agent has a mental state the content of which is to-A, and she is on those grounds assessible by the norms of strong consistency, means-end coherence and agglomerativity, then she has an intention to A. \[ (p \& r) \rightarrow s \]

3. (from 1 & 2) If an agent has a mental state the content of which is to-A, and the function of that mental state requires that the agent can be counted on to A, then she has an intention to A. \[ (p \& q) \rightarrow s \]

4. If an agent is trying to A, then she has a mental state the content of which is to-A, and the function of that mental state is to make it the case that she A’s. \[ t \rightarrow (p \& u) \]

5. If the function of a mental state is to make it the case that an agent A’s, then the function of that mental state requires that the agent can be counted on to A. \[ u \rightarrow q \]

6. (from 4 & 5) If an agent is trying to A, then she has a mental state the content of which is to-A and the function of that mental state requires that the agent can be counted on to A. \[ t \rightarrow (p \& q) \]

\[ \therefore (from \ 6 \ & \ 3) \] If an agent is trying to A, then she has an intention to A. \[ t \rightarrow s \]
The stretch of argument of premises 1-3 employs the driving ideas behind the planning theory of intention. The stretch of argument from premise 4 to the conclusion extends those ideas to determine what mental states the trying agent must have.

It is worth making a few comments about this argument, by way of clarification. First, notice that the argument does not imply, nor do I believe it’s the case, that one cannot act without trying to act. Just as one can achieve coordination without intention, one can act without trying to. Nonetheless, the point of trying is to act, just as the point of intending is coordination.

Second, as Ginet’s examples, and the other examples that we can generate following his lead, show, it is perfectly possible for an agent to serve some other end, besides acting, by trying, an end that is achieved just by trying, independently of action. This, however, is no objection to the argument just offered for in this respect trying is just like intending: as the toxin puzzle shows, intending can serve some other purpose besides inter- and intrapersonal coordination. Just as this fact is compatible with the fact that the point of intending is to achieve coordination, the fact that Ginet’s examples illustrate is compatible with trying’s function being to bring about action.

Third, to say that we must “count on” people to act as they intend or as they are trying to act, is not to say that action is, or need be, guaranteed if trying and intending are to serve their respective purposes. We might count on ourselves to act as we try in just the way that we count on ourselves and others to do as we intend; it’s always possible that people won’t do what we are counting on them to do, but there is nonetheless a presumption of action where there is trying or intending. There’s another way to put the point: Unless it’s hopeless, it always makes sense to advise those who hope to coordinate their behavior to make plans. This is true even though we all recognize that sometimes the best laid plans don’t get us coordination, and even though we all recognize that sometimes people achieve coordination fortuitously and without planning. Similarly, unless it’s hopeless, it always makes sense to advise someone who hopes to act to try.
This is true even though we all recognize that sometimes we fail despite our best efforts, and sometimes succeed effortlessly. The reason that these are both good pieces of advice, despite the fact that they don’t identify certain routes to success, is that what is being advised is to use the tool that you can count on to work; planning is the tool for coordinating, and trying is the tool for acting; we can count on those tools to work, even though they don’t always.

This point should go some way to answering concerns that might be raised about premise 5, the claim that the function of trying to act requires counting on the trying agent to act. It might be objected that the sense in which the aim of coordination requires that we can count on agents to do as they intend is quite different from the sense in which the aim of acting requires that we can count on agents to do as they are trying to do. Our chances of coordination are low when we can’t count on agents to do as they intend. But there is no comparable problem when it comes to acting; it’s not as though an agent’s chances of acting are lower if the agent can’t be counted on to do as he tries. Counting-on serves a clear instrumental role in generating coordination, but no comparable instrumental role in generating action. While this is correct, it doesn’t serve to undermine the argument, for there is another sense in which counting on agents to do as they act is necessary for accomplishing the characteristic aim of trying: it’s what provides reason to try. Counting-on is necessary to make trying worthwhile (when trying isn’t serving some purpose distinct from its characteristic purpose); similarly, counting-on is necessary to make intending worthwhile (when intending isn’t serving some purpose distinct from its characteristic purpose). So, while counting-on is necessary to the purpose of intending for different reasons from those for which it is necessary to the purpose of trying, it is nonetheless necessary for both, and that’s all that the argument requires.

In any event, we have now reached the result at which this paper aimed: In trying to act, and thereby employing a mental tool the function of which can be realized only if we can count on the agent to act, the agent becomes subject to the very norms of attitudinal rationality.
that govern those with intentions. Since, as the planning theory of intention has shown us, to have an intention is to have a practical mental state that makes one appropriately assessible by such norms, it follows that if one tries to act, one intends to act.

As noted at the beginning of this paper, this result does not show, uncontrovertibly, that there is no more to the mental side of trying to act than intending. But notice how suggestive the argument offered here is of that result. If what individuates types of practical mental state are the norms of attitudinal rationality to which those with those mental states are subject, and if it is the case, as has been argued, that those who try are subject to the norms governing those who intend, then it is hard to see what grounds one could have for distinguishing the mental component of trying from intention. This isn’t to say that there are no possible grounds: it’s possible that there will turn out to be additional norms of attitudinal rationality governing those who try that do not govern those who intend. But in the absence of a description of such norms or even an argument for thinking there are any, it seems likely that the mental side of trying is just intention.

**Conclusion**

Return, finally, to the law. As I said at the beginning, the result reached here has an immediate practical implication. It provides an answer to the following question: Which of two statutes proscribing attempted crimes is defensible, one (like Indiana’s) that makes intent an element of attempt even when the completed crime does not require intent, or one (like Colorado’s) that requires for attempt only the mental states required for the completed crime? The answer is the first, for without intent nothing that a person does could be a case of trying to act, regardless of the question of what mental states one must have in order to successfully act. Since laws against attempts are laws against trying to do things that we take, for their own
reasons, to be worthy of legal proscription, any just law governing attempt will have intent to act as an element.

The proposition that one tries to act only if one intends to act implies that completed crimes can be individuated more finely than attempts. That is, it is often the case that completed crimes C1 and C2 are of different types, despite the fact that an attempt to perform C1 and an attempt to perform C2 are of the same type. For instance, imagine that two drivers, S1 and S2, are drag racing. As they speed down the road, both know that they are risking the death of passing pedestrians. But imagine that S1 actually intends to kill a pedestrian, while S2 does not. It’s not that S1 has his mind on two things, winning the race and killing a pedestrian, as he drives down the road; rather, he, like S2, is only thinking about winning the race, but, unlike S2 he went out drag racing with the aim of killing a pedestrian. Now imagine that during the course of the race, each driver kills a pedestrian. Under many statutory schemes they have committed different crimes, for many statutory schemes individuate completed crimes by appeal to intent: to kill with intent is a purposive homicide, while to kill knowing there’s a risk that you might is merely reckless homicide. But what could the difference be between attempts to commit these two crimes? In order to attempt either crime, a person would have to intend to kill, for trying necessarily requires intent, as has been argued here. Thus, although completed crimes can be of different types when they differ only with respect to intent, attempted crimes cannot differ only in this respect, for every attempt includes intent.

We might think that this shows that it is possible to attempt only crimes that it is possible to complete with intent, unlike reckless homicide where intent undermines its recklessness. Better, however, is to see the result as showing, instead, that attempts are always attempts to commit some unspecified crime in a particular category, where the categories are defined broadly enough to include all of those grades that differ from one another only by the accompanying mental states of the defendant. There are attempts to kill, but there aren’t
attempts to commit negligent, reckless or purposive homicide; there’s no such thing as an attempt to commit a particular grade of homicide, but only an attempt to commit homicide, period. Thus, even if both of our drag racers are not guided by an intent to kill at the final moment before crashing into their unfortunate victims—imagine that both have genuinely lost control of their vehicles at that moment—one, namely the one who went out racing with an intention to kill, has done all that is required to be guilty of an attempted homicide. Both kill, and both killings are the result of recklessness on the part of the drivers; but one of these homicides was attempted, while the other was not.

There is a further practical upshot of the view that, where members of a category of crime are individuated by mental state, it is only categories of crime that are attempted. In many states, and under the Model Penal Code §5.05(1)\textsuperscript{15}, the sentence for an attempt is tied to the sentence for the completed crime. Commonly, for instance, although not under the Model Penal Code, the sentence for an attempted crime is half the sentence for the completed crime. But what has been shown here is that this practice lacks an adequate rationale, for sentencing of completed crimes is often a function of the mental states that are elements of those crimes. Reckless homicides, for instance, are given much less stiff penalties than intended homicides. If attempts can’t be individuated with reference to specific completed crimes, but only with reference to categories of completed crimes, then sentences for attempts ought not to be tied to the sentences for specific crimes; such crimes often belong to the same category (i.e. homicide) while differing greatly from one another in penalty.

While it is right to say that trying lies at the basis of the connection between the mind and the body exhibited in action, what’s been shown here is that there is little reason to think that there is more to that connection than intention in the mind, behavior in the body, and a connection, probably causal, between the two. It has been shown here, also, that the law can
learn much from what may seem to be no more than a philosophical fact (in the pejorative sense of the term “philosophical”) about this relationship between the mind and body.¹

¹ As Alfred Mele has pointed out to me in conversation, the action component of trying might very well be mental. Someone who tries to do a bit of arithmetic in his head might do nothing non-mental. What is true, however, is that there’s no trying without both a spring of action and something sprung. Ordinarily, although not always, the something is a bit of physical behavior.


³ Sometimes a court will say that an agent who intends X knowing that it is likely to cause Y also intends Y. That is, courts will sometimes use the term “intend” to refer to something that is not intention in the ordinary sense. In fact, such a usage is parasitic on the ordinary usage for when courts use the term this way they also recognize that the sense in which the defendant intends X is quite different from the sense in which he “intends” Y. However, in this paper, the terms “intend” and “intention” will be used only in their ordinary senses.

⁴ Jones v. State, 689 N.E.2d 722 (Ind. 1997)


One of the central questions under discussion in this literature is this: Is the strong consistency requirement violable? That is, is it possible to intend to act while believing you will not, as Bratman assumes? There are difficult questions here linked, at bottom, to difficult questions about the distinction between logical and metaphysical possibility. On the one hand, it seems that, since there is no logical contradiction in the idea of any particular combination of mental states present in a particular mind no matter what their content—just as there’s no logical contradiction in the idea of a library containing books that contradict each other—there is no impossibility in it. However, if there could be metaphysical impossibility without logical, then this inference does not follow; perhaps the nature of certain mental states is such as to exclude others from the mind. Even if there is such a thing as metaphysical impossibility, however, it doesn’t follow that it is metaphysically impossible to intend while believing in failure. In any event, I will be assuming that Bratman is right: the pressure not to believe you will fail that is imposed by an intention is normative rather than metaphysical. The same goes, I assume, for the other rationality requirements on intention.

9 Bratman 1987, p. 194.

10 Of course, this is just the sort of thing that those who take the strong consistency requirement to be inviolable would deny to be possible.

11 Newcombe’s paradox places in tension not the causal and the attitudinal conceptions of instrumental rationality, but, instead, two different forms of the causal conception. One-boxers take the past financial success of one-boxers to show that there is good reason to think that in the circumstances they are more likely to reach their ends by taking one box. Two-boxers, by contrast, think that given that the decisions they make cannot cause events that preceded those
decisions, they do best by taking both boxes. The disagreement, then, is about what one has reason to think will lead to the best outcome, given the odd circumstances of the example, and not a disagreement of the sort that can be found between those accepting the attitudinal and those accepting the causal conception of rationality.

12 Carl Ginet, “Trying to Act” in Freedom and Determinism, J. K. Campbell, M. O’Rourke and D. Shier (eds.), MIT Press: Cambridge, pp. 89-102, 2004. Ginet takes all three cases to illustrate that an agent can try to do something without intending to do it. However, Ginet is simply assuming that an agent does not intend to do what he believes he will not do, and so concludes that the people in all of his examples lack the intention to act. Thus, the argument to follow, if it succeeds, refutes this aspect of Ginet’s position.

13 Ginet 2004, pp. 95-6

14 Bratman is well-known for having shown that intention has “two faces”: one having to do with intentional action, the other having to do with the mental states involved in planning. Bratman takes trying, like aiming, to contribute to intentional action. However, what is being suggested here is that trying, like intending, has a second face, as well, for in addition to playing a role in generating intentional action, it is essential to the kind of coordination that is the hallmark of planning agency.

15 Model Penal Code §5.05(1) reads, “[A]ttempt[s]…are crimes of the same grade and degree as the most serious offense that is attempted…An attempt…to commit a capital crime or a felony of the first degree is a felony of the second degree.” Hence, the Model Penal Code determines the sentence for an attempt to commit a noncapital offense by reference to the crime that the agent aims to complete; it equates the sentence for attempt with that sentence. But it pegs the sentence for an attempt to commit a capital crime, also, to the sentence for the completed crime, giving the defendant, in such cases, a discount on his punishment.
Thanks to Michael Bratman, Steve Finlay, Andrei Marmor, Al Mele, Kadri Vihvelin, Jonathan Weil and participants in a recent seminar in philosophy of law at USC for comments on an earlier draft.