9

How to be a Cognitivist about Practical Reason

Jacob Ross

Cognitivism about practical reason is the view that intentions involve beliefs, and that the rational requirements on intentions can be explained in terms of the rational requirements on the beliefs that figure in intentions. In particular, cognitivists about practical reason have sought to provide cognitive explanations of two basic requirements of practical rationality: a *consistency* requirement, according to which it is rationally impermissible to have intentions that are jointly inconsistent with one's beliefs, and a *means–end coherence* requirement, according to which, to a first approximation, it is rationally impermissible to intend an end while failing to intend what one regards as a necessary means to this end. In order for the cognitivist to explain these requirements, she must arrive at an account of the beliefs that figure in intentions, on the basis of which she can show that any agent who violates these requirements of pratical rationality must have beliefs that violate the requirements of theoretical rationality. Providing such an account, however, turns out to be no easy● task. ● Q1

This paper will be divided into three parts. In the first, I will lay out some general constraints that a theory of intentions must satisfy if it is to figure in a cognitivist explanation of the requirements of intention consistency and means–end coherence. In the second part, I will consider the standard cognitivist account of intentions, according to which the intention to ϕ involves the belief that one will ϕ because of that very intention. I will show that this account of intention faces a number of serious problems. And in

I am grateful for very helpful comments on an earlier draft of this paper from Stephen Finlay and from Gideon Yaffe. I have also benefited greatly from discussions with Jeff King and David Manley, from the students in my Practical Reason seminar, held at the University of Southern California in the spring of 2007, and from participants in the 4th Annual Metaethics Workshop. My greatest debt is to Mark Schroeder, for countless extremely illuminating discussions.

the third part, I will discuss an account according to which the beliefs that figure in intentions must be defined not only in terms of their content, but also in terms of the other attitudes on which they are based. I will argue that this is the kind of account of intentions that the cognitivist about practical reason should adopt.

1. GENERAL CONSTRAINTS ON A COGNITIVIST ACCOUNT OF INTENTIONS

1.1. The Intention Consistency Requirement and the Strong Belief Thesis

Some cognitivists hold that intentions are *identical* with a certain kind of belief. Minimally, cognitivists hold that intentions *involve* beliefs. And generally, cognitivists maintain that the intention to ϕ involves the belief that one will ϕ. We may call this latter claim the *Strong Belief Thesis*.[1] One of the primary motivations for the Strong Belief Thesis is that it allows for a cognitivist explanation of the consistency requirement on intentions. This is the requirement that one's intentions be logically consistent not only with one another but also with the totality of one's beliefs.

In order to state this requirement precisely, it will be useful to introduce a term. When someone intends to ϕ, let us say that the *propositional content* of her intention is the proposition that she ϕs. We can now state the requirement of *Intention Consistency* as follows:

> One ought rationally to be such that the set consisting of all propositional contents of one's beliefs, as well as all the propositional contents of one's intentions, is logically consistent.

We can easily provide a cognitivist explanation of this requirement so long as we assume the Strong Belief Thesis. For this thesis implies that believing that p is a necessary condition for having an intention whose propositional content is p. And it is widely accepted that there is a requirement of *Belief Consistency*, which can be stated thus:

> One ought rationally to be such the set of all the propositions one believes is logically consistent.

[1] See Michael Bratman, "Intention, Belief, Practical, Theoretical," forthcoming in Jens Timmerman, John Skorupski, and Simon Robertson, eds., *Spheres of Reason*. Bratman calls this thesis the "strong belief requirement," though I will reserve the latter expression for a normative principle, which I will discuss below.

But if believing that *p* is a necessary condition for having an intention whose propositional content is *p*, then any proposition that is the content of one's intentions will also be the content of one's beliefs. Therefore the set that includes all the propositions that one believes will be identical with the set that includes all the propositions that are the contents either of one's intentions or of one's beliefs. And so, if it is rationally required that the former set be logically consistent, then it is likewise rationally required that the latter set be logically consistent.

Furthermore, it does not appear to be possible to provide a purely cognitivist explanation of the consistency requirement on intentions on the basis of any assumption that is weaker than the Strong Belief Thesis. For if the necessary condition for intending that *p* is having some doxastic attitude that is weaker than belief in *p*, then the doxastic attitudes involved in having inconsistent intentions may themselves be fully consistent. There is no inconsistency, after all, in having a high degree of confidence in each proposition in a set of jointly inconsistent propositions, as the lottery paradox illustrates. Thus, it seems that in order to explain the consistency requirement on intentions in terms of a demand of cognitive consistency, the doxastic attitude involved in intention can be nothing weaker than all-out belief. One might still hold that in order to intend that *p*, one need not believe *that p*, but that it is instead sufficient to believe some weaker proposition. Jay Wallace, for instance, has proposed a cognitivist theory according to which the belief involved in intending to ϕ is the belief that is, in some relevant sense, possible that one ϕs.[2] However, if the belief involved in intending that *p* is weaker than the belief that *p*, then it is possible to have inconsistent intentions without having inconsistent beliefs.[3]

For suppose the content of the belief constitutive of intending to ϕ is weaker than the proposition that one will ϕ, and suppose the content of the belief constitutive of intending not to ϕ is weaker than the proposition that one will not ϕ. Let *p* be the proposition that one will ϕ, and let *q* be the weaker proposition that is the content of the intention constitutive of intending to ϕ. Since *q* is weaker than *p*, *q* is equivalent to ●(*p* or (*q* and not *p*). Hence, the belief constitutive of intending to ϕ will be the belief that (*p* or (*q* and not *p*). But since (*p* or (*q* and not *p*) is consistent with (not *p*), one can consistently have the belief constitutive of intending

[2] Jay Wallace, "Normativity, Commitment, and Instrumental Reason," *Philosophers' Imprint* 1/3 (December 2001).

[3] The following argument is a generalization of an argument Bratman makes in "Intention, Belief, Practical, Theoretical".

to ϕ while believing that not p. A fortiori, since we are assuming that the belief constitutive of intending not to ϕ is weaker than the belief that not p, (p or (q and not p) is consistent with the belief constitutive of intending not to ϕ. Hence, it follows from our assumptions that one can intend to ϕ, and intend not to ϕ, without having inconsistent beliefs. And so, on these assumptions, a violation of intention consistency need not involve any cognitive inconsistency. And so, if the cognitivist is to explain the consistency requirement, he must not hold that the belief constitutive of intending an action is weaker than the belief that one will perform this action.

One might hold that the cognitivist could explain the Belief Consistency Requirement not on the basis of the Strong Belief *Thesis*, but instead on the basis of *Strong Belief Requirement*:

> One ought rationally to be such that (if one intends to do something, one believes one will do it).

While the Strong Belief Thesis is a metaphysical principle, according to which believing one will ϕ is a necessary condition for intending to ϕ, the Strong Belief Requirement is a normative principle, according to which intending to ϕ without believing that one will ϕ, though metaphysically possible, is irrational. On the basis of the Strong Belief Requirement, we can give a straightforward explanation of the Consistency Requirement. For if rationality requires that one believe one will do whatever one intends to do, then rationality will require that the set of propositional contents of one's intentions be a subset of the set of propositions one believes. Hence, if rationality requires that the set of propositions one believes be logically consistent, then rationality will also require that the propositional contents of one's intentions be consistent with one's beliefs. Hence, from the Belief Consistency Requirement, together with the Strong Belief Requirement, we can derive the Intention Consistency Requirement.

Thus, we might explain the Consistency Requirement by appealing not to the Strong Belief Thesis, but instead to the Strong Belief Requirement. This strategy, however, is not available to the cognitivist about practical reason. For the cognitivist about practical reason aims to explain certain requirements of practical rationality, including the Consistency Requirement, purely on the basis of requirements of theoretical rationality. And while the Strong Belief Requirement may be a requirement of rationality, one cannot plausibly hold that it is a requirement of *theoretical* rationality. For intending to ϕ without believing that one will ϕ, if it is indeed possible, would not seem necessarily to involve any irrationality in one's beliefs. If, for example, one does not believe that ϕ will phi because one has compelling evidence that one will not ϕ, and yet one intends to ϕ

nonetheless, then it will be one's intentions, and not one's beliefs, that are irrational. Thus, violations of the Strong Belief Requirement, if they can occur at all, needn't involve theoretical irrationality.

Hence, since the cognitivist aims to explain the Intention Consistency Requirement in terms of the requirement of theoretical rationality, it seems she has no choice but to accept the Strong Belief Thesis. In the next section, I will raise what appears to be a serious problem for this thesis, and I will indicate how I believe the cognitivist should respond.

1.2. A Problem for the Strong Belief Thesis, and How the Cognitivist Should Respond

Suppose you are an unlucky train passenger. On a thousand past occasions, a strange sequence of events have occurred: you have been the only passenger onboard a train, the train has been approaching an oncoming train, and the conductor has suddenly died of cardiac arrest, forcing you to take control of the train in order to avoid a collision. On each of these occasions, you have arrived at a junction prior to colliding with the oncoming train, and you have had the options of turning the train you are on either to the left or to the right. On five hundred of these past occasions, you chose to turn the train to the left, and on the remaining five hundred occasions you chose to turn it to the right. There has not, however, been any discernible pattern within this sequence of choices. And so you form the justified belief that you choose at random, and that you have an equal propensity to send the train to the left or to send it to the right.

Today you are once again the only passenger onboard a train, whose conductor is Casey Jones. You are very familiar with Casey, and with this particular route: you know that at River Junction, Casey either turns the train to the left or to the right, and that he turns in both directions with equal frequency, and without any discernible pattern. Hence, on this particular occasion, you are 50 percent confident in the following proposition:

R The train will turn to the right at River Junction

Today Casey appears to be in poor health, and so you think it quite possible that he will die during today's voyage. Moreover, because of your unfortunate track record, you regard the following proposition as a genuine epistemic possibility:

D Casey dies right before the train reaches River Junction, at which point there is another train approaching, and there are only two equally good ways to avoid a fatal collision, namely to turn the train to the left or to turn the train to the right.

In this case, conditional on D, how confident should you be in R? Since you justifiedly believe that you have an equal propensity to send the train to the left or to the right in the kind of circumstance under consideration, you should believe that if D were true, you would have a 50 percent chance of sending the train to the left, and a 50 percent chance of sending it to the right. Hence, by the Principal Principle, your credence in R conditional on D should be one-half. And so your credence in R conditional on D should be equal to your unconditional credence in R.

Now according to a plausible evidentialist view of reasons for belief, only something that makes it more likely that *p* is true (relative to a given agent's epistemic situation) can be a reason for this agent to believe that *p* is true. But in the situation under consideration, D does not make it any more likely that R is true. Hence, it seems that D is not a reason to believe R.

We can arrive at a similar conclusion by another route. For, plausibly, a rational agent updates her beliefs by conditionalizing on her total evidence. Hence, when one learns, and learns only, that D is true, one should conditionalize on D, and so one's new credence in R should be equal to one's prior credence in R conditional on D. In other words, upon learning that D is true, and in the absence of any further evidence, one's credence in R should remain one half. And so one's credence in R should not be anything close to unity. But if, in response to learning that D is true, one's credence in R should be one half rather than unity, then it seems that D cannot be a sufficient reason to believe that R is true.

However, R is a sufficient reason to turn the train to the right at River Junction (just as it is also a sufficient reason to turn the train to the left at River Junction). But assuming one can only turn the train to the right at River Junction if one intends to do so, it seems that R must also be a sufficient reason to intend to turn the train to the right at River Junction. Now suppose that the strong belief thesis were true. In this case, the belief that one will turn the train to the right at River Junction is a constituent of the intention to turn the train to the right at River Junction. Hence, any sufficient reason to intend to turn the train to the right at River Junction must also be a sufficient reason to believe that one will turn the train to the right at River Junction. But any sufficient reason to believe that one will turn the train to the right at River Junction must also be a sufficient reason to believe that the train will turn to the right at River Junction. Hence, if the Strong Belief Thesis is true, then D must be a sufficient reason to believe that the train will turn to the right at River Junction, that is, to believe R. We have seen, however, that there is strong reason to deny that

D is a sufficient reason to believe R. Hence, there is strong reason to deny the Strong Belief Thesis.

How should the cognitivist respond to this objection? I believe she should take issue with the following inference

(1) R is a sufficient reason to turn the train to the right.
(2) One can only turn the train to the right by intending to turn the train to the right.
(3) R is a sufficient reason to intend to turn the train to the right.

This inference seems to rely on the following principle:

> If X is sufficient reason to ϕ, and one can only ϕ by ψ-ing, then X is a sufficient reason to ψ.

But the above principle only applies to actions. It does not apply to things that are not subject to the will. Suppose, for example, that the fact that it is a beautiful Sunday afternoon is a sufficient reason to go for a walk. And suppose that I can only go for a walk if I do so by converting glucose into adenosine triphosphate. We cannot conclude that the fact that it is a sunny day is a sufficient reason to convert glucose into adenosine triphosphate. Indeed, the fact that it is a beautiful Sunday afternoon could not be a reason for me to convert glucose into adenosine triphosphate, since the latter process is not an action.

But if the above principle applies only to actions, then we cannot apply it to intentions unless we regard intentions as actions. And the since the cognitivist holds that intentions involve beliefs, and since there is good reason to deny that beliefs are actions, the cognitivist has good reason to deny that intentions are actions, and hence that the above principle applies to intentions. So the cognitivist has good reason to deny the inference from (1) and (2) to (3).

Thus, the counterargument to the Strong Belief Thesis that we have considered in this section relies on a principle which the cognitivist can reasonably reject. There are, however, other objections to the Strong Belief Thesis. And since, as I have argued above, the cognitivist has little choice but to accept the Strong Belief Thesis, anyone who is persuaded by these objections should reject cognitivism, at least in its pure form. There may, however, be interesting views in the neighborhood of pure cognitivism that do not involve the Strong Belief Thesis. In Appendix A, I discuss one such view, a view according to which intending to ϕ involves not believing that one will ϕ, but rather accepting that one will ϕ from the practical point of view.

1.3. Means–End Coherence and the Non-Cognitive Conditions of Intention

The simplest formulation of the means–end coherence requirement, which we may call the *Strong Means–End Coherence Requirement*, is this:

SME One ought rationally to be such that (if one intends to ψ and one believes that ϕ-ing is a necessary means to ψ–ing, then one intends to ϕ).

In other words, it states that anyone who intends to ψ and believes that ϕ-ing is a necessary means to ψ-ing is thereby rationally required to intend to ϕ. Now in order to explain the consistency requirement on intentions, it is sufficient to find a *necessary* condition for intention, such as the condition that anyone who intends to ϕ must believe that she will ϕ. However, in order to explain the means–end coherence requirement, we must also find a sufficient condition for intention. For if all we knew were a necessary condition for intending to ϕ, then although we might be able to show that a belief–intention pair rationally requires the satisfaction of this necessary condition, this would not amount to showing that this pair of attitudes rationally requires intending to ϕ.

(A note on terminology: for brevity, I will say that an agent is *rationally required* to A whenever she has a set, S, of attitudes such that she could not rationally fail to A while retaining this set of attitudes. Thus, in saying that an agent is rationally required to A, I do not mean to imply that the agent in question could not rationally fail to A *simpliciter*, but only that she could not rationally fail to A while keeping her other attitudes constant.)

Although cognitivists must hold that intentions involve beliefs, and some cognitivists maintain that intentions are *identical* with a certain kind of belief, cognitivists needn't make this stronger claim, since they may hold that intentions also involve a non-cognitive component, such as a desire or disposition. However, there are significant constraints on the type of non-cognitive component the cognitivist can posit. In particular, the cognitivist must hold that the non-cognitive component of the intention to ϕ is a condition that is present whenever one is required, by the means–end coherence requirement, to intend to ϕ—let us call such a condition a *non-cognitive background condition* of the intention to ϕ. For suppose the cognitivist denies this. Then she must hold that someone could violate the means–end coherence requirement purely in virtue of failing to have the non-cognitive component of the required intention. That is, she must hold that there could be a case in which someone intends to ψ, believes that ϕ-ing is a necessary means to ψ-ing, and has the belief-component of the intention to ϕ, but fails to intend to ϕ because she lacks the non-cognitive component of the intention to ϕ. Since such an agent would violate the means–end coherence requirement, any cognitivist explanation of this requirement would need to imply that any such agent would violate a requirement of theoretical rationality. But this implication is implausible. For while the requirements of theoretical rationality may require that an agent who has certain beliefs, or other attitudes, have certain beliefs, but such a requirement does not require that an agent who has certain belief or other attitudes have certain non-cognitive attitudes, dispositions, or the like. Hence, if it is possible for an agent to be required, by means–end coherence, to intend to ϕ, and yet to lack the non-cognitive component of the intention to ϕ, then it will be impossible to give a cognitivist explanation of the means–end coherence requirement. Therefore, the cognitivist must deny that this is possible, and so she must claim that the non-cognitive component of the intention to ϕ is what I have called a background condition.

But if the non-cognitive component of the intention to ϕ is a background condition in this sense, then the belief component of the intention to ϕ cannot be simply the belief that one will ϕ. For it is possible to satisfy any background conditions of the intention to ϕ, and to believe that one will ϕ, without intending to ϕ.

Consider the following case. Barry the banker has received a shipment of a million dollars. Right now the money is on the counter where any thief could easily take it. Barry intends to protect the money, and he believes that locking the money in the safe is a necessary means to doing so. Normally, at this point in the day, he would engage in some simple instrumental reasoning, and form the intention to lock the money in the safe. But today, before he does so, Robbie the robber enters, disguised as Marvin the Martian. Robbie declares "I have come from the twenty-third-and-a-half century to give you a copy of your biography." Barry is very keen to read his biography, and he immediately turns to today's date, where he reads "Barry locks the money in the safe." Being the gullible individual he is, he believes every word that he reads, and thus believes that he will lock the money in the safe. And yet, enthralled in reading about his own future, he doesn't make any intentions, and in particular, he doesn't form the intention to lock the money in the safe. And while Barry is engrossed in his biography, Robbie makes off with the money.

Surely this sort of case is possible. And in this case, though Robbie does not intend to lock the money in the bank, he does believe that he will, and he satisfies the background conditions for intending to lock the money in the safe (since these are present whenever one is required by means–end coherence to intend to lock the money in the safe). Thus, more is involved in the intention to ϕ than the belief that one will ϕ plus the non-cognitive background conditions of this intention. And since, as we have seen, the cognitivist must hold that the only non-cognitive condition of the intention to ϕ is a background condition, it follows that the cognitivist must hold that the intention to ϕ involves some cognitive condition beyond the belief that one will ϕ. And, as we will now see, there are tight constraints on what this cognitive condition can plausibly be held to be.

1.4. The Problem of Mere Recognition

In the simplest kind of instrumental reasoning, we begin in a state in which we intend some end, to ψ, and in which we believe that ϕ-ing is a necessary means to ψ-ing, and we then form the intention to ϕ. Thus, we begin in a situation in which we have other attitudes which require, by means–end coherence, a further intention that we lack, and we then form this required intention. I might, for example, intend to drink a beer, and believe that going to the fridge is a necessary means to drinking a beer. Since I am required, by means–end coherence, to intend to go to the fridge, I must satisfy the non-cognitive background conditions of this intention. And so it follows that when, at the outset of the process of instrumental reasoning, I have not yet formed the intention to go to the fridge, I must lack only the cognitive component of this intention. And so my forming this intention requires, and requires only, that I form this cognitive component. More generally, the cognitivist must hold that, in the simplest cases of instrumental reasoning, forming the instrumental intention consists entirely in forming the belief component of this intention.

And suppose that, at time t, I intend to drink a beer, and I believe that going to the fridge is a necessary means to doing so, and hence I satisfy the non-cognitive background conditions for this intention. Let p be the propositional content of the belief component of the intention to go to the fridge. And suppose that, at t, I have compelling reason to believe that p is true, and I form the belief that p purely on the basis of this compelling evidence. In this case, my belief that p would have been formed in a rational manner. But in my situation, my forming the belief that p is tantamount to forming the intention to go to the fridge. Hence, if I can rationally form the belief that p on the basis of compelling evidence, then I can rationally form the intention to go to the fridge on the basis of this same evidence. Now at least in the case where p is true, to come to believe that p on the basis of compelling evidence is to *recognize* that p. And so, if it is possible to form the belief constitutive of intending to ϕ on the basis of compelling evidence that p is true, then it will be possible for a mere process of recognition to constitute practical reasoning. But it would seem that mere recognition can never constitute practical reasoning. Certainly, recognition can play an important role in practical reasoning, since very often one of the things we do in reasoning our way to the intention to ϕ is to recognize practical reasons to ϕ. But as cases of akrasia illustrate, the recognition of these reasons is only part of practical reasoning, and the formation of the intention to ϕ, though causally influenced by such recognition, is a distinct event from this recognition. The formation of intentions, it seems, is not a matter of merely recognizing that something is true, but rather of resolving to make something true.

Since, therefore, the cognitivist must hold that, at least in the simplest cases of instrumental reasoning, forming the instrumental intention consists entirely in forming the belief component of this intention, it follows that, in order to avoid the conclusion that practical reasoning can consist in mere recognition, the cognitivist must hold that the belief component of the instrumental intention cannot be formed on the basis of compelling evidence. And so it seems that if the belief component of the intention to go to the fridge is the belief that *p*, the cognitivist must hold that *p* is a proposition that one cannot come to believe on the basis of compelling evidence.

One possibility is that *p* could be a proposition for which there can never be compelling evidence. The problem with this view, however, is that while there are many propositions for which there could never be compelling evidence (e.g. "Caesar did and did not cross the Rubicon") it is doubtful that any such proposition can be rationally believed, and if a proposition cannot be rationally believed, then it cannot be a component of a rational intention. One might hold that there are certain propositions for which the evidence can be at most merely sufficient, but can never be compelling. But if there were any such propositions, then it would seem that belief in them would be, at most, rationally permissible, and never rationally required, and hence, having formed the belief in such a proposition, one would always be at liberty to subsequently withhold one's assent. And so if practical reasoning consisted in the formation of beliefs for which there could be sufficient evidence, but for which there could never be compelling evidence, then it would seem that, having formed an intention on the basis of practical reasoning, one would always be at liberty to rationally withdraw the belief component of the intention, thereby withdrawing the FN:4 intention.[4]

[4] One might object that I am here assuming a form of evidentialism which no cognitivist would accept: I am assuming that the only reasons for belief are evidential reasons. But so long as one understands a reason for belief a consideration that can figure in the reasoning whereby this belief is formed, this is a very reasonable assumption. So-called practical reasons for a belief, such as the fact that having the belief in question would serve one's interests, can figure in reasoning whereby one forms the intention to

Fortunately, in order to maintain that the belief component of an intention cannot be formed on the basis of compelling evidence, the cognitivist needn't hold that the belief in question is the belief in a proposition for which there can never be compelling evidence. For she might instead maintain that although one can have compelling evidence for the belief that constitutes intention, one cannot form this belief on the basis of compelling evidence, since one never has this evidence *prior to the existence of the belief in question*. And the way this might be true is that the propositional content of the belief in question might entail that one has this very belief. For, arguably, someone who lacks a given belief can never have compelling evidence for a proposition that entails that she has this belief. Although one might have plenty of evidence for the claim that believes that *p* even if one does not, it is arguable that such evidence can never be decisive. After all, one can always ask oneself whether *p* is true, and if one answers negatively, or suspends judgment, then one can be fairly confident that one does not believe that *p*, however much other evidence there may be for the claim that one does so believe. Hence, if the content of the belief-component of a given intention is a proposition that entails that one believes this very proposition, then, at least arguably, this belief could never be formed on the basis of compelling evidence. But once one has formed this belief, one may then have compelling evidence that the proposition believed is true, and so one may then have good reason to retain this belief.

It seems, therefore, that the cognitivist has strong reason to adopt the *Self-Referential Belief Thesis*:

> Every intention involves a belief that entails that one has this very belief.

For this thesis enables her to maintain both that the formation of intentions can never consist entirely in the formation of a belief on the basis of compelling evidence, and that, having formed an intention, one may have good reason to retain this intention. And many cognitivists have indeed proposed accounts of intention that entail the Self-Referential Belief Thesis. So let us now turn our attention to these accounts.

2. SELF-REFERENTIAL BELIEF ACCOUNTS OF INTENTION

In his seminal paper "Practical Reasoning," Gilbert Harman proposed that the intention to ϕ involves the belief that one will ϕ because of that very

acquire the belief in question, but they cannot figure in reasoning whereby one forms the belief in question.

intention. Similarly, in "Cognitivism about Instrumental Reason," Kieran Setiya argues that an intention to ϕ is a belief that can be expressed thus: "I will ϕ in part because of this very intention."[5] And David Velleman, in *Practical Reflection*, provides several alternative characterizations of the content of the beliefs that constitute intentions, all of which involve such direct self-reference. Sometimes he characterizes the intention to ϕ as a belief that can be expressed simply as "I'll ϕ *herewith*," while on other occasions he identifies this intention with a more complicated belief, expressed, for instance, as "because I have such and such motives for getting myself to ϕ, and I know that I have such motives, I am *hereby* reinforcing these predispositions to the point where I'll ϕ."[6]

On these views of intention, intentions involve beliefs that refer to themselves, or to the intentions of which they are essential constituents, directly, so that their proper expression involves indexicals or demonstratives such as "*hereby*," "*herewith*," or "*this* very intention." Hence these views entail not only the Self-Referential Belief Thesis, but more specifically the *Direct Self-Reference Thesis*:

> Every intention involves a belief that refers to itself directly, in the sense that its expression requires an indexical or demonstrative that refers directly to the belief in question, or to the intention in which it figures.

In the next section, I will show that the Direct Self-Reference Thesis faces a number of serious difficulties, but that many of these difficulties can be avoided if we move to an indirect version of the Self-Reference Thesis. Later I will show that there are further problems that are faced by both versions of the Direct Self-Reference Thesis.

2.1. Direct and Indirect Self-Reference Accounts of Intention

There are several problems with the view that intentions involve directly self-referential beliefs. For one thing, it is questionable whether there are any such beliefs. For it is plausible that anything that can be believed can be doubted, and hence that at any time, one can believe a given proposition only if one could alternately suspend judgment concerning this proposition. Now let B_p be any directly self-referential belief, and let p be its content. It follows that p involves direct reference to B_p. Hence, at the time at which

[5] Kieran Setiya, "Cognitivism about Instrumental Reason" (forthcoming in *Ethics*); see also *Reasons without Rationalism* (Princeton: Princeton University Press, 2007).

[6] See J. David Velleman, *Practical Reflection* (Princeton: Princeton University Press, 1989), 86–8.

one first believes p, one could not have any attitude toward p unless B_p exists. And so at this time, one could only suspend judgment concerning p if one simultaneously believed that p, which is impossible. Thus, if it were possible, at this time, to believe that p, then it would be possible to believe something which one could not possibly doubt, which contradicts our initial assumption.

A further reason for doubting that there are such things as directly self-referential beliefs is that it is hard to see how such beliefs could be assigned identity conditions. For if someone has a belief B_1 at one time, and a belief B_2 at a later time, then a necessary condition for these being numerically the same belief is that they have the same content. Thus, we can only specify the identity conditions for a belief if we can independently specify the belief's content. But if a belief involves direct reference then we can only specify the content of this belief if we can independently specify the identity conditions of the objects to which it directly refers. Therefore, if a belief involves direct reference to itself, then we can only specify its content if we can independently specify its own identity conditions. But since we can only specify its identity conditions if we can independently specify its content, it follows that we cannot specify its identity conditions. Hence, if there were such things as directly self-referential beliefs, then either they must lack identity conditions, or they must have identity conditions that are ineffable. And neither of these possibilities appears to be very plausible.

Moreover, even if directly self-referential beliefs were possible, they could not be formed by any valid inference. For a conclusion that involves direct reference to a given object, x, can only follow from premises that likewise involve direct reference to x. Thus, if the belief B makes direct reference to itself, then it can follow only from premises that likewise make direct reference to B. But this is possible only if B already exists at the time when one believes the premises. And if B cannot follow from any premises that precede its existence, then B cannot be formed by any valid inference.

The cognitivist need not regard this last argument as a decisive objection to the view that intentions consist in directly self-referential beliefs, for the cognitivist may hold that practical reasoning does not consist in valid inferences. Indeed, Harman, Velleman, and Setiya all reject the view that practical reasoning can be expressed in the form of a logical inference. Still, if there is to be such a thing as practical reasoning, then it must be possible to have reasons for forming intentions. And hence if intentions are directly self-referential beliefs, then it must be possible to have reasons for acquiring such beliefs. It seems, however, that there can be no such reasons.

For, in general, one can only have reason to A if one can consider the question as to whether to A, and can hence see various considerations as bearing on this question. And one can have reason to acquire a directly self-referential belief only if there is some proposition, p, such that one can have reason to come to believe that p, and such that, in coming to believe that p in an appropriate manner, one would thereby acquire a self-referential belief. It follows that one can have reason to acquire a directly self-referential belief only if there is some proposition, p, such that one can consider the question as to whether to come to believe that p, and such that, in coming to believe that p in an appropriate manner, one would thereby form a directly self-referential belief. But it will only be true that, in coming to believe that p in an appropriate manner, one forms a directly self-referential belief, if p is a proposition that refers directly to the belief that one would thereby form. Yet prior to having formed a belief, there is no proposition one can consider that refers directly to this belief. Therefore, prior to forming a self-referential belief, there will be no proposition that one could consider coming to believe, such that in coming to believe this proposition in an appropriate manner one would thereby acquire a self-referential belief. Therefore, there can be no reason for acquiring a directly self-referential belief. And so if intentions involve directly self-referential beliefs, the there can be no reason to form an intention. Hence, it will be impossible to reason one's way to an intention.

Fortunately, however, we can solve all these problems if we modify the account of intention under consideration. On what we may call the *indirect self reference account*, intentions involve beliefs that refer to themselves not indexically, but by means of descriptions.[7] On a simple version of such an account, the intention to ϕ involves a belief that can be expressed thus: "I will ϕ because of my current intention to ϕ." For the sake of simplicity, I will focus, in what follows, on this simple version of the indirect self-reference account of intentions, though my arguments apply more generally.

Any account of the intention to ϕ as involving a belief of the form "I will ϕ because of my intention to ϕ" clearly involves circularity. Hence, it can hardly qualify as a reductive analysis of intention. Such an account is no more circular, however, than Setiya's account according to which to intend to ϕ is to have a belief that can be expressed as "*I will ϕ because of this very intention.*" Further, such circularity is not a serious problem for the cognitivist. For the ultimate aim of the cognitivist is not to elucidate the concept of intention, but rather to explain the rational requirements

[7] In "Intentions and Self-referential Content" (*Philosophical Papers* 24 (1995), 151–66), Tomis Kapitan draws a distinction that is similar to the one I am drawing here between direct and indirect self-reference accounts of intention.

to which intentions are subject. And the cognitivist may achieve this aim by positing certain conditions for intending to ϕ, even if these conditions cannot be understood without an independent grasp of the notion of intention.

The indirect self-reference account of intentions solves several of the problems we have seen for the standard cognitivist account of intentions as consisting in directly self-referential beliefs. First, unlike directly self-referential beliefs, indirectly self-referential beliefs have a content that can be represented without the belief in question ever having existed. For this reason, the existence of indirectly self-referential beliefs is compatible with the claim that whenever one can believe a proposition, it is possible instead to suspend judgment concerning this proposition. Second, it is possible to specify the content of an indirectly self-referential belief without having independently specified the identity conditions of this belief, and for this reasons, indirectly self-referential beliefs can have specifiable identity conditions. Third, indirectly self-referential beliefs have a content that can be entailed by beliefs that do not involve direct reference to these beliefs. And so indirectly self-referential beliefs can be formed via valid inferences. And fourth, as we will now see, the indirect self-reference account of intentions does a better job that the direct self-reference account at explaining the means–end coherence requirement.

2.2. Explaining the Weak Means–End Coherence Requirement

Kieran Setiya proposes a cognitivist explanation for a means–end coherence requirement in "Cognitivism about Instrumental Reason." Here he endorses, and attempts to explain, a principle that we may call the *Weak Means–End Coherence Requirement*:

WME If a fully rational agent intends to ψ, and believes that she will ψ only if she ϕs-because-one-now intends-to-ϕ, then she intends to ϕ.

Setiya's explanation of this principle involves the following *Modus Ponens Requirement*:

MP If a fully rational agent believes that p, and believes that (if p then q), then she believes that q.

Since, on Setiya's account of intentions, intending to ψ involves believing that one will ψ, it follows from this account that if someone intends to ψ, and believes that she will ψ only if she ϕs-because-she-now-intends-to-ϕ, then she is required, by the Modus Ponens Requirement, to believe that she will ϕ-because-she-now-intends-to-ϕ. And so, in order to explain WME, Setiya need only claim that anyone who believes she will ϕ-because-she-now-intends-to-ϕ is rationally required to intend to ϕ. And this claim follows from the following principle:

X One ought rationally never to falsely believe that one intends to ϕ.

WME follows from the conjunction of principle X and the Modus Ponens Requirement. And so if the latter two principles are genuine requirements of theoretical rationality, then Setiya will have succeeded in deriving WME purely on the basis of requirements of theoretical rationality. Note, further, that if we accept these two principles, then we can explain WME on the basis of any account of intentions whatsoever that includes the Strong Belief Thesis. For the only assumption about the nature of intentions that figures in this argument is the assumption that anyone who intends to ψ believes that she will ψ.

However, many would question whether principle X is a genuine requirement of rationality, let alone a genuine requirement of theoretical rationality.[8] Fortunately, if we move from the direct to the indirect self-reference account of intentions, we can explain WME without invoking principle X.

Suppose an agent intends to ψ, and believes that he will only ψ if he ϕs-because-he-now-intends-to-ϕ. Since he is required, by means–end-coherence, to intend to ϕ, it follows that he satisfies the non-cognitive background condition of the intention to ϕ. Hence, in order to show that he is rationally required to intend to ϕ, it will suffice to show that he is rationally required to have the belief component of the intention to ϕ. And this, according to the indirect self-reference view, is the belief that he will ϕ-because-he-now-intends-to-ϕ. But since he intends to ψ, and hence, on the direct self-reference view, believes that he will ψ, and since he also believes that he will ψ only if he ϕs-because-he-now-intends-to-ϕ, it follows from the Modus Ponens requirement that he is rationally required to believe that he will ϕ-because-he-now-intends-to-ϕ. And this, according to the direct self-reference account of intentions, is precisely the cognitive component of the intention to ϕ. Hence, in intending ψ, and in believing that he will only ψ if he ϕs-because-he-now-intends-to-ϕ, he is rationally required to intend to ϕ. And so, if we assume the indirect self-reference account of intentions, we can explain WME purely on the basis of the the Modus Ponens requirement, without invoking principle X.

Another advantage of the indirect self-reference account is that it enables us to understand instances of instrumental reasoning as valid inferences.

[8] For criticisms of this principle, see Michael Bratman's "Intention, Belief, Practical, Theoretical".

For suppose one begins with the intention to ψ and the belief that one will only ψ if one ϕs-because-one-now-intends-to-ϕ—and hence one satisfies the background conditions for intending to ϕ—and on the basis of this intention–belief pair one forms the intention to ϕ. On the indirect self-reference account of intentions, this process of reasoning is equivalent to the following inference:

(1) I will ψ because I now intend to ψ.
(2) I will only ψ if I ϕ because I now intend to ϕ.

---

(3) I will ϕ because I now intend to ϕ.

And this is a valid inference.

It seems, therefore, that the indirect self-referential belief account of intentions is highly successful. For in addition to solving the problems for the direct self-reference view we discussed earlier, it also does a better job at explaining both the means–end coherence requirement and instrumental reasoning. But we aren't out of the woods yet. For while the indirect self-reference view provides a good explanation of WME, the latter, as I shall now argue, is too weak to count as the proper formulation of the requirement of means–end coherence.

2.3. Why the Weak Means–End Coherence Requirement is Too Weak

I will now argue that since there are often circumstances in which it would be rational to intend to ϕ, but in which it would not be rational to believe that we will ϕ only if we so intend, WME cannot be the proper formulation of the means–end coherence requirement.

Consider a case in which I am the designated driver for a party that I will be attending this evening. I intend to drive home safely, and I believe that remaining sober is a necessary means to my driving home safely. Suppose I know that it is very difficult for me to resist the temptation to drink alcohol, and so if I want to ensure that I remain sober at the party, I must form the firm intention to do so before I arrive at the party. Suppose, however, that I recognize that there is a chance that I might remain sober even if I don't form this prior intention prior to going to the party: there might not be any alcohol at the party, or there may be no alcohol that appeals to me, or I may I may find that I am unusually resistant to its tempting influence. In this case, it seems clear that I am rationally required to intend to remain sober at the party. But this conclusion does not follow from:

WME One ought rationally to be such that (if one intends to ϕ, and believes that one will ϕ only if one ϕs-because-one-now-intends-to-ψ, then one intends to ψ).

For WME would imply that I am rationally required to intend to stay sober only if I believe that (I will drive home safely only if I remain-sober-because-I-now-intend-to-remain-sober). But in the case in question, I do not have this belief.

The case just considered involves a prior intention directed toward a future action. But similar problems arise for concurrent intentions concerning our present actions. Consider a case in which Daria intends to win a game of darts. Her opponent is doing very badly, and so all Daria needs in order to win the game is to hit one of the three central rings of the five-ring dart board. She is a very good darts player, and so she knows that if she intends to hit one of the three central rings, she will do so. Of course, if she merely had the general intention to hit the dartboard, without specifically intending to hit one of the three central rings, she might nonetheless hit one of these central rings. Let us suppose that she is agnostic on the issue of whether she would hit one of the three central rings if she had no specific intention to do so: she thinks there is a reasonable chance that she would, but that there is also a reasonable chance that she would not. In this case, it would seem that, given Daria's other beliefs and intentions, she cannot rationally fail to intend to hit one of the central three rings. But again, this requirement does not follow from WME. For Daria does not believe that she will hit one of the three central rings only if she does so because she intends to do so.

It seems, therefore, that WME is too weak to count as the proper formulation of the requirement of means–end coherence. However, as I will now argue, the Strong Means–End Coherence requirement is too strong. Fortunately there is a third formulation of this requirement, the *Moderate Means–End Coherence Requirement*, which, like Baby Bear, is just right. And this formulation, it will turn out, can be explained on the basis of the indirect self-referential account of intention.

2.4. Explaining the Moderate Means–End Coherence Requirement

Recall that according to the Strong Means–End Coherence Requirement, one cannot rationally intend an end without intending what one believes to be a necessary means to this end. This formulation appears to be overly strong. For when we intend to carry out some action in the future, there may be a large number of intermediate actions that we believe to be necessary means to this future action, but to have an intention, in the present, to carry out each one of these intermediary actions would seem to involve superfluous mental clutter. Thus, I intend to bicycle to campus tomorrow afternoon. And I believe that the following are all necessary means to my doing so: walking to my front door; turning the handle of my front door; opening my front door; walking to my bicycle; removing my keys from my pocket; inserting my bicycle key into my bicycle lock; turning the key; etc., etc. But surely I am not rationally required to already have all these intentions. It would suffice for me to form these intentions tomorrow afternoon.

And there is a further reason to deny that SME is the proper formulation of the means end coherence requirement. Suppose I am in prison, and I intend to keep my mind occupied tomorrow by counting the cracks in the wall of my prison cell. And suppose I believe that remaining in my prison cell is a necessary means to counting these cracks. In this case, it follows from SME that I am rationally required to intend to remain in my prison cell. But this hardly seems plausible. For I know that I have no choice but to remain in my prison cell. Hence, whether to remain in my prison cell is not a possible object of rational deliberation. But it would seem that I cannot be rationally required to have an intention that I could not form by way of rational deliberation. And so it would seem that I cannot be rationally required to intend to remain in my prison cell, contrary to SME.

In the above example, the necessary means is something I believe I will never be able to choose. But other counterexamples can be given that do not involve this feature. Suppose it is now Monday. I have a crystal ball that enables me to see everything that I will do on Tuesday, but that does not reveal any events after Tuesday. Suppose, further, that tonight I will forget everything I learned from the crystal ball, and so tomorrow I will no longer have foreknowledge of all my actions. Suppose that today, as I gaze into the crystal ball, and I observe that tomorrow I will pick up my suit from the dry cleaners. After observing this, I might deliberate concerning what to do on Wednesday. Since I know my sister's wedding is on Wednesday, I may form the intention to wear my suit to her wedding. And I may believe that picking up my suit from the dry cleaners is a necessary means to wearing my suit to her wedding. But since I already know, on the basis of compelling evidence, that I will be picking up my suit from the dry cleaners, I am not in a position to deliberate concerning whether to pick up my suit from the dry cleaners. For deliberation concerning whether to ϕ must proceed from a state of uncertainty concerning whether one will ϕ, and this is ruled out in my present case. In the words of Isaac Levi, prediction crowds out deliberation. And since I am not in a position to rationally deliberate concerning whether to pick up my suit from the dry cleaners, it seems I cannot be rationally required to intend to do so.

One possible response to this problem is to say that one is never under a rational requirement to intend to ϕ if one has sufficient evidential reason to believe that one will ϕ. But the cognitivist should not welcome this proposal. For, at least very plausibly, a belief is only rational if it is theoretically or epistemically rational, and a belief is only theoretically or epistemically rational if there is sufficient evidential reason for this belief. Hence, if intending to ϕ involves believing that one will ϕ, then whenever one lacks sufficient evidential reason to believe that one will ϕ, it would be irrational to intend to ϕ. And so, if one accepts the current proposal, and holds that one can only be rationally required to intend to ϕ when one lacks sufficient evidential reason to believe one will ϕ, then must conclude that one can only be rationally required to intend to ϕ when intending to ϕ would be irrational. And this is hardly a desirable conclusion.

There is another problem with this proposed response. For when one is about to perform an action, one is generally in a position to know that one intends to perform this action, and that one is unlikely to change one's mind or to fail in the performance of this action. Thus, when one is about to perform an action, then one generally has sufficient evidential reason to believe that one will perform this action. And so if this proposed response were correct, and one is never under a rational requirement to intend to ϕ if one has sufficient evidential reason to believe that one will ϕ, then it will very seldom be true that one is rationally required to perform any action that one is about to perform.

What we should say, therefore, is not that one is never rationally required to ϕ when one has sufficient evidential reason to believe that one will ϕ, but rather that one is never rationally required to intend to ϕ when, *apart from one's intention to* ϕ, one has sufficient evidential reason to believe one will ϕ. And, if we adopt this proposal, then we must reformulate the means–end coherence requirement. According to the *Moderate Means End Coherence Requirement*,

MME One ought rationally to be such that (if one intends to ψ, and one believes that ϕ-ing is a necessary means to ψ-ing, and if, apart from the intention to ϕ, one would have insufficient evidence for the belief that one will ϕ, then one intends to ϕ).

Note that this principle applies in many cases in which the Weak Means–End Coherence Requirement does not apply, and in which, intuitively, a principle of means–end coherence ought to apply. Consider once more the case of Daria who intends to win the game of darts, and who believes that hitting one of the three central rings of the dartboard is a necessary means to winning the game. Since she does not have the belief that she will win the game only if she hits one of the central rings *because she now intends to do so*, WME does not imply that she is rationally required to hit one of the three central rings. However, under normal circumstances, MME will imply that she is rationally required to hit one of these rings. For Daria recognizes that unless she intends to hit one of the central three rings, there is a good chance that she will not hit one of these rings. Hence, conditional on her not having the intention to hit one of the central three rings, she has little confidence that she will hit one of these rings. And so if she does not believe that she intends to hit one of these rings, she should not believe that she will hit one of them. And if she does not have sufficient reason to believe that she intends to hit one of these rings, she will not have sufficient reason to believe that she will hit one of them. But at least under normal circumstances, unless she intends to hit one of these rings, she will not have sufficient reason to believe that she intends to do so. Consequently, apart from the intention to hit one of these rings, she would not have sufficient reason to believe that she will hit one of them. Therefore, in this case, the conditions for the applicability of MME are met.

Thus, MME appears to be a plausible candidate for being the proper formulation of the means–end coherence requirement. And luckily, it is not difficult to provide a cognitivist explanation of MME. All we need to assume is the Strong Belief Thesis. For suppose this thesis is true, and that the antecedent of MME is satisfied. In this case, one will believe that one will ψ, and one will also believe that ϕ-ing is a necessary means to ψ-ing, and so one will be rationally required to believe that one will ϕ. Further, if one satisfies the antecedent of this conditional, then apart from the intention to ϕ, one would lack sufficient evidence for the belief that one will ϕ. And, at least plausibly, if one lacks sufficient evidence for the belief that one will ϕ, then one cannot rationally believe that one will ϕ. It follows that, apart from the intention to ϕ, one cannot rationally believe that one will ϕ. Therefore, if one is rationally required to believe that one will ϕ, then one is rationally required to intend to ϕ. But we have seen that if the antecedent of the conditional is satisfied, then one is rationally required to believe that one will ϕ. Therefore, if the antecedent is satisfied, then one is rationally required to intend to ϕ. In other words, one ought rationally to be such that if one satisfies the antecedent of the conditional, one satisfies the consequent. And so the Moderate Means–End Coherence Requirement is true, and can be explained purely on the basis of requirements of theoretical rationality, assuming only the Strong Belief Thesis.

So things are looking rosy. For if we accept the Strong Belief Thesis, we can explain not only the requirement of Intention Consistency, but also the Moderate Means–End Consistency Requirement, which is arguably the proper formulation of the means–end coherence requirement. And if we accept the Self-Referential Belief Thesis, then we can also avoid the problem of mere recognition that we discussed in Section 1.4. And while the direct version of the Self-Referential Belief Thesis faced a number of difficulties, we saw how these difficulties can be avoided by moving to the indirect version. But alas, two serious problems still lie ahead. In the next section, I will argue that the standard self-referential belief accounts of intention have very implausible implications concerning the circumstances in which intentions can be rational. And in Section 2.6, I will argue that any view of intentions that involves the Self-Referential Belief Thesis will be incompatible with a coherent understanding of instrumental reasoning.

2.5. The Problem of Causal Overdetermination

We have seen that there is reason to doubt the cognitivist view that intending to ϕ involves believing that one will ϕ. But the self-referential views of intention that we are now considering have much stronger implications concerning the beliefs involved in intention. First, they imply that a necessary condition for intending to ϕ is that one believe that one intends to ϕ, or at least that one believe that one has the belief-component of this intention. Second, they imply that a necessary condition for intending to ϕ is believing that one's intention to ϕ will play a role in causing one to ϕ. Are these implications plausible?

Concerning the first implication, many authors have pointed out that young children, and even some animals (e.g. chimpanzees) appear to have intentions, and yet it seems implausible to ascribe to them second-order beliefs about their attitudes. Of course, one might argue that young children and animals really do have these beliefs but in an implicit or inarticulate manner; or one might argue that young children and animals don't really have intentions; or one might argue that while they may have intentions of a sort, these intentions are not subject to rational requirements, and so the cognitivist account of intentions need not apply to them. Each of these maneuvers, however, has a cost, and so there is reason to prefer an account of intentions that does not require us to make them.

Far more problematic is the implication that in order to ϕ one must believe that one will ϕ because one now intends to ϕ. For it seems that in many cases in which intentions are called for, the belief that one will perform the intended action because one now intends to do so would not be a rational belief.

FN:9 Harman considers cases of this kind in "Practical Reasoning."[9] Suppose Judy has the option of going to a party, but decides to stay home instead. In this case, Harman argues, it might not be rational for Judy to believe that she will stay home *because she so intends*. For even if she had no intention to stay home, she might do so anyway out of habit. Harman calls intentions of this kind "negative intentions."

Not all cognitivists are persuaded by such cases. Kieran Setiya argues as follows: "[Harman's] 'negative' intentions are causes, too. It is just that the action they cause is over-determined: it would have happened without them." Hence we can accept the claim that anyone who intends to ϕ believes that she will ϕ because she so intends, "so long as we reject, FN:10 or qualify, the counterfactual test for causation."[10] Thus, even if, in the absence of the intention to stay home, Judy would have stayed home out of habit, it does not follow that her intention to stay home can't cause her to stay home. For her intention to stay home might preempt her habit in causing her to stay home, and might thereby be the genuine cause of her staying home.

The problem with this response, however, is that pre-emption cuts both ways. While it is true that the one's current intention to ϕ might preempt other potential causes of one's ϕ-ing, it is likewise true that other potential causes of one's ϕ-ing might preempt one's current intention to ϕ. Thus, now, at noon, I intend to brush my teeth before going to bed at midnight, but I am aware that I may brush my teeth at midnight not because of my having this intention at noon, but rather out of habit. Similarly, right now, in the summer, I intend to wear long sleeves in the winter, but I recognize that when winter comes, I may wear long sleeves not because of any intention I formed in the summer, but because I will then want to stay warm, and I will believe that wearing long sleeves is a necessary means to staying warm. Or being the designated driver, I may intend not to drink at the party I will be attending this evening, while recognizing that there is a possibility that I will refrain from drinking at the party not because of my current intention not to drink at the party, but because when I arrive at the party I will find that the only available beverages are ones that I find revolting.

Thus, a prior intention to perform a given action may be pre-empted by an independent motivation, at the time in question, to perform the action in

[9] Gilbert Harman, "Practical Reasoning," *Review of Metaphysics*, 29 (1976), 431–63, reprinted in his *Reasoning, Meaning, and Mind* (Oxford: Clarendon Press, 1999), 46–74. Harman's discussion of the Judy example occurs on pp. 53–4 (all page references to Harman's paper will refer to the reprint).

[10] "Cognitivism about Instrumental Reason," footnote 30.

question, such as a habit, or an emotional response, or the recognition of a decisive reason. Further, a prior intention can be pre-empted by subsequent deliberation. I may now intend to retire when I'm 65, while recognizing that what causes me to retire when I'm 65 may be not be my current intention to retire when I'm 65, but rather an intention to retire when I'm 65 that I form at a later stage in life when I reopen the question as to when to retire.

Finally, one can have an intention while recognizing that one may fulfill this intention not because one has this intention, but rather because of circumstances completely outside of one's control. Suppose that Ed intends to do anything Jane asks him to do on her birthday. He might have this intention while recognizing that Jane may not ask him to do anything on her birthday. But if she doesn't ask him to do anything, then although it will trivially be true that Ed does everything Jane asks him to do on her birthday, this will not be true because of Ed's intention.

Thus, in a wide variety of cases where it appears to be rational to intend to ϕ, it would not be rational to believe that one will ϕ because one now intends to ϕ. And so if, as the standard cognitivist accounts of intentions imply, anyone who intends to ϕ believes that she will ϕ because she so intends, then it follows that these apparently rational intentions are in fact irrational. This, therefore, is a very strong reason to reject such accounts of intention.

2.6. The Problem of Practical Reasoning

We saw earlier that, on the indirect self-reference account, intentions can be formed by way of valid inferences. However, on this view of intentions, they cannot be formed by way of *sound* inferences. For on this account, the intention to ϕ involves the belief that one will ϕ because one now intends to ϕ. Hence, the intention to ϕ involves a belief that can only be true if one intends to ϕ. But any premises from which one form the intention to ϕ by way of a sound inference must be premises that entail the belief component of the intention to ϕ, and so they must entail that one intends to ϕ. Now if these premises are true, then one already intends to ϕ, and so one cannot form the intention to ϕ on the basis of these premises. And if, on the other hand, these premises are not all true, then they cannot provide the basis for a sound inference. Either way, it will be impossible to arrive at the intention to ϕ by way of a sound inference.

But this conclusion, like the conclusion that on the direct self-reference view intentions cannot be formed by way of valid inference, need not trouble the cognitivist, since, as before, she may deny that practical reasoning takes the form of a logical inference. But the indirect self-reference account of intentions has a much more troubling implication, since it implies that no one who understands what intentions are could ever form the intention to ϕ by way of a conscious process of reasoning. For if an intention involves a belief, then any reasoning in which one forms this intention must be reasoning in which one forms the belief it involves. And reasoning leading to the belief that p is reasoning in which one is guided by the question *whether p*, and in which one arrives at an affirmative answer to this question. Hence, if the intention to ϕ involves the belief that p, then any self-conscious process of reasoning wherein one forms the intention to ϕ must be a self-conscious process of reasoning wherein one arrives at an affirmative answer to the question *whether p*. Hence, if the intention to ϕ involves the belief that one will ϕ because one now intends to ϕ, then one can self-consciously form the intention to ϕ only if one can self-consciously arrive at an affirmative answer to the following question:

Q1 Will I ϕ because I now intend to ϕ?

And if one understands what intentions are, then one will regard Q1 as equivalent to:

Q2 Will I ϕ because I of an attitude that involves the belief that (I will ϕ because I now intend to ϕ)?

But if one is asking question Q2, then one cannot already have made up one's mind concerning the answer. And if one is asking question Q2 self-consciously, then one will recognize that one has not yet made up one's mind concerning the answer. Thus, one will recognize that one does not believe that one will ϕ because one now intends to ϕ. And so the answer to Q2 will be obvious: "no!" And if one understands that the view of intentions under consideration is correct, and hence one regards Q1 as equivalent to Q2, then the answer to Q1 will be equally obvious: "no!" Hence one will be unable to self-consciously arrive at an affirmative answer to Q1. And so it follows, on the account under consideration, that one will be unable form the intention to ϕ by way of a self-conscious process of reasoning.[11]

FN:11

[11] Perhaps the problem derives from the temporal indexical, "now." There would be no difficulty, after all, in arriving, by way of self-conscious reasoning, at the conclusion "I will ϕ because I intend at t to ϕ," so long as one does not believe that the time is now t. And so if the cognitivist held that the belief-component of the intention to ϕ is not the belief that (one will ϕ because one now intends to do so) but rather the belief that (one will ϕ because one intends at t to do so) for some t other than the present, then the cognitivist could avoid the problem just described. But she would do so at significant cost. For she would no longer have a view of intentions according to which intention formation cannot consist in mere recognition. For one can certainly arrive, on the basis

We have seen that in order to explain the means–end coherence requirement, the cognitivist must give sufficient conditions for intending a means. These sufficient conditions must include a cognitive condition. And if it also includes a non-cognitive condition, the latter must take the form of a background condition. And this implies that, in instrumental reasoning, the formation of the instrumental intention consists in the formation of the cognitive component of this intention. And this implies, in turn, that if this cognitive component is a belief that can be formed on the basis of compelling evidential reasons, then the formation of intentions can consist in mere recognition. The cognitivist may be able to avoid the conclusion that the formation of intentions can consist in mere recognition, so long as she adopts the self-referential belief thesis. But if she adopts this thesis, she will run into all the problems we have encountered in the last three sections.

Besides the constraints just given, are there any independent reasons to accept the Self-Referential Belief Thesis? Some cognitivists argue for the Self-Referential Belief thesis on the basis of what we may call the *Self-Referential Intention Thesis*:

> Anyone who intends ϕ intends to (ϕ because of her current intention to ϕ).

In Appendix B, I argue that none of the standard arguments for the Self-Referential Intention Thesis is sound. I also argue that there is good reason to reject this thesis. Hence, arguments for the Self-Referential Belief Thesis that are based on the self-referential Intention Thesis are without force.

And so the cognitivist appears to be faced with a dilemma. There are a number of serious problems with the Strong Belief Thesis, that is, and so if the cognitivist claims that in order to satisfy the cognitive conditions for intending to ϕ, one must believe that one intends to ϕ, then the cognitivist is in hot water. But if, on the other hand, she claims that in order to satisfy the cognitive condition for intending to ϕ, it is sufficient to have a belief with some specified content, a content that does not entail that one intends to ϕ, then she will be committed to the conclusion that merely recognizing the truth of a proposition on the basis of compelling evidence can constitute forming an intention.

But it may be that the cognitivist can avoid both horns of this dilemma. That is, she may be able both to reject the view that intending to ϕ involves believing that one intends to ϕ, and to reject the view that believing some

of compelling evidence, at the belief that one will ϕ because, at some other time, one intends to ϕ. And so if such belief were the cognitive component of the intention to ϕ, one could form the intention to ϕ merely by recognizing the truth of this belief.

particular proposition that does not entail that one intends to ϕ is sufficient for having the belief component of the intention to ϕ. The way to reject both of these claims is to deny that having the belief component of the intention to ϕ is simply a matter of having a belief with an appropriate content. An alternative is that the belief component of the intention to ϕ must be a belief that not only has the right content, but that also has the right kind of basis.

3. AN ALTERNATIVE ACCOUNT OF INTENTIONS

3.1. The Basis of Intentions

Gilbert Harman has proposed that an intention to ϕ is an idea concerning the future which

(i) represents itself as causing it to be the case, or as guaranteeing, that one shall ϕ; and
(ii) is arrive at and maintained by practical reasoning.

We have seen that the kind of self-referential view expressed in (i) faces a number of problems, and in Appendix B I discuss further problems with this kind of view. But the second part of Harman's view may be more promising. Perhaps we can distinguish between intentions and mere predictions of our future actions on the ground that the beliefs involved in the these two cases are arrived at by way of different kinds of reasoning. Perhaps Harman is on the right track in saying that an intention is "an idea of the future arrived at and maintained by practical reasoning."[12] FN:12

This suggestion cannot be accepted exactly as it stands, for it seems that there could be intentions that are not arrived at and maintained by practical reasoning. Suppose I intend to go to the fridge. I arrived at this intention via practical reasoning, proceeding from the intention, or perhaps the mere desire, to drink orange juice, and the belief that going to the fridge is a necessary means to drinking orange juice. It seems possible, if extremely improbable, that a swamp man could spontaneously materialize in front of my fridge with a psychology very similar to my own, and in particular, with a similar intention to go to the fridge. In this case, his intention would not have been arrived at by practical reasoning. Furthermore, once one has formed an intention, it does not seem to be always necessary to engage in practical reasoning simply in order to maintain this intention. Thus it

[12] "Practical Reasoning," 63.

would seem that the swamp man's intention to go to the fridge might be neither arrived at, nor maintained, by practical reasoning.

It appears, therefore, that we cannot understand intentions as beliefs about the future formed and maintained by way of practical reasoning. Perhaps, however, we can understand them as beliefs that are directly based on practical reasons. Let us say that a given attitude, *A*, belonging to an agent, *s*, is *directly based on practical reasons* just in case *s*'s disposition to retain attitude *A* is directly explained by the fact that *s* has a set of attitudes from which she *could* arrive at attitude *A* by way of practical reasoning. Given this definition, we may propose that the intention to ϕ consists in the non-cognitive background condition for intending to ϕ, whatever that may be, plus a belief that one will ϕ that is directly based on practical reasons.

Thus, although the swamp man's belief that he will go to the fridge was not arrived at by practical reasoning, this intention is at least partly based, in the relevant sense, on his intention to drink orange juice and his belief that going to the fridge is a necessary means to drinking orange juice. For this pair of attitudes partly explains his disposition to retain his belief that he will go to the fridge, since apart from this pair of attitudes he would be less reluctant to form some other plan of action, and hence to cease believing that he will go to the fridge. This must be true, since we are assuming that he is a psychological duplicate of me, and in my case the corresponding belief–intention pair play the corresponding explanatory role in my psychology. And this pair of attitudes is one on the basis of which it is possible to arrive at the belief that one will go to the fridge by way of practical reasoning. Thus, since the swamp-man has a belief that he will go to the fridge that is based on a set of attitudes of the appropriate kind, he satisfies the cognitive condition for intending to go to the fridge.

So when is it true that a given belief could be arrived at on the basis of a given set of beliefs by way of practical reasoning? Clearly, the set consisting in the intention to ψ and the belief that ϕ-ing is a necessary means to ψ-ing is a set of attitudes on the basis of which one could arrive, by way of practical reasoning, at the intention to ϕ. And so if the intention to ϕ involves the belief that one will ϕ, it is a set on the basis of which one could arrive, by way of practical reasoning, at the belief that one will ϕ. But equally clearly, it is not the only set of attitudes of this kind, since not all practical reasoning takes this simple, instrumental form. And in order for the view of intentions under consideration to be complete, it would need to specify which processes of reasoning count as practical reasoning and which do not—it could not, without vicious circularity, simply say that a process of reasoning counts as practical reasoning just in case it issues in an intention, since it defines intentions in terms of the kinds of reasoning from which they issue. Of course, even once one has specified which processes of reasoning count as practical reasoning, there will still remain some circularity, since some of these forms of reasoning will proceed from sets of attitudes that include intentions. But this kind of circularity is no more vicious than that found in standard functionalist accounts of mental attitudes, on which attitudes of a given type are defined in terms of the ways in which they causally interact with other attitudes, including attitudes of the same type.

3.2. Evaluating Our Account of Intentions

We must now determine whether the account of intentions we have proposed can explain the requirements of intention consistency and of means–end coherence, and whether it can do so while avoiding the pitfalls of the other accounts of intention we have considered.

Clearly, like any account of intentions that entails the Strong Belief Thesis, the present account can explain the Intention Consistency Requirement. Similarly, like any account of intentions that entails the Strong Belief Thesis, this account, in conjunction with principle X, can explain the Weak Means–End Coherence Requirement. And, much more importantly, like any account that entails the Strong Belief Thesis, this account can explain the Moderate Means–End Coherence requirement.

Can it provide a cognitivist explanation of the Strong Means–End Coherence Requirement? It seems it cannot. For any such explanation would need to show that it is a requirement of theoretical rationality that:

> Anyone who intends to ψ, and believes that ϕ-ing is a necessary means to ψ-ing, must believe that she will ϕ *on the basis of a set of attitudes from which she could arrive, by practical reasoning, at the belief that she will* ϕ.

And while it is clear, on the current account, that any such agent would be rationally required to believe that she will ϕ (since this belief follows straightforwardly from her other beliefs), it is unclear why she should have to believe this on the basis of a set of attitudes of a specified kind. For she might have other sufficient reasons to have this belief, and if this is the case then it would seem perfectly theoretically rational for her to have this belief entirely on the basis of these other sufficient reasons.

Suppose, for example, that I am driving home, and I know that I will soon be approaching a stop sign. And suppose I intend always to obey the traffic regulations, and I believe that a necessary means to my doing so is that I stop at the upcoming stop sign. In this case, I will have a belief–intention pair on the basis of which I could form the belief that I will stop at the stop sign by way of practical reasoning. Suppose, however, that I have abundant evidential reason for believing that I will stop at the stop sign. Suppose, for example, that in my long driving history, I have never failed to stop at a stop sign. In this case, I have a surplus of attitudes on the basis of which I could form the belief that I will stop at the stop sign. And it would seem that I could be perfectly theoretically rational without basing my belief that I will stop at the stop sign on all these attitudes: I could, with perfect theoretical rationality, believe that I will stop at the upcoming stop sign purely on the basis my knowledge of my own perfect track record.

In response to a similar difficulty, Harman has suggested that it is a requirement of rationality that one never form a belief by way of theoretical reasoning that one could form by way of practical reasoning. Such a principle would imply that, in the case just described, I could not rationally form the belief that I will stop at the stop sign by theoretical reasoning, on the basis of my knowledge of my track record, since I am in a position to form this belief by practical reasoning, on the basis of my intention always to obey the traffic regulations. However, even if the proposed principle were a requirement of rationality, it surely isn't a requirement of theoretical rationality: if I were to form the belief that I will stop at the stop sign on the basis of the available evidence, I would not thereby be theoretically irrational. And so the cognitivist cannot appeal to such a principle in explaining the means–end coherence requirement.

It seems, however, that there is no reason to look for supplementary principles by which we could explain the Strong Means–End Coherence Requirement. For, as we saw in Section 2.5, there is reason to believe that SME is considerably too strong. And so we should regard that fact that our theory does not entail SME as a virtue of this theory, not as a vice. Since it does explain the Moderate Means–End Coherence Requirement, and since the latter appears to be a formulation of the right level of strength, our account appears to provide an adequate cognitivist explanation of the requirement of means–end coherence.

In this respect, however, it does not differ from any other account that entails the Strong Belief Thesis. But our theory has a number of additional virtues. First, like the self-referential belief accounts of intentions, it avoids the implication that the recognition of the truth of a proposition on the basis of compelling evidence can constitute the formation of an intention. This view can avoid this conclusion, not by claiming that the belief component of the intention to ϕ is a belief with a special content, but rather by claiming that the belief in question must be based on a set of attitudes by which it could be formed by way of practical reasoning. So long as the view includes an account of practical reasoning according to which purely evidential reasons can never be an adequate basis for such reasoning, the view will entail that intentions can never be based on purely evidential reasons. And so the view under consideration shares the principal virtue of the self-referential belief account of intentions. But it avoids many of the problems of the latter account.

First, because, on this view, to intend to ϕ one needn't believe that one will ϕ because of this very intention, this view does not imply that, in cases in which we think the causal efficacy of our intention may be preempted by other factors, we cannot rationally intend to ϕ. Thus, this view does not suffer from the problem of causal overdetermination discussed in Section 2.3.

Second, this view allows for a very satisfying account of instrumental reasoning. Suppose I intend to ψ and believe that ϕ-ing is a necessary means to ϕ-ing. On the view we are considering, I will be in a position to reason my way to the intention to ψ. For on the view under consideration, in intending to ψ, I will believe that I will ψ, and from this belief, together with the belief that ϕ-ing is a necessary means to ψ-ing, I can infer that I will ψ. The belief that I thereby form will be based on a pair of attitudes of the right kind (namely an intention and an instrumental belief), and so in having this belief I will satisfy the cognitive condition for intending to ϕ. And since I am required, by means–end coherence, to intend to ϕ, I must antecedently satisfy the non-cognitive conditions for intending to ϕ, whatever they may be. And so since, in forming the belief that I will ϕ, I come to satisfy the cognitive condition for intending to ϕ, I will come to satisfy all the conditions for intending to ϕ, and will thus acquire the intention to ϕ. Hence, I will arrive at the intention to ϕ by way of a valid inference.

And third, on the present view, it is possible to form the intention to ϕ by way of an inference that is not only valid, but sound. For on the present view, the belief involved in the intention to ϕ is simply the belief that one will ϕ, not the belief that one will ϕ because one so intends. And so the relevant belief is one that can follow from a set of premises that do not entail that one has this belief. Hence it can follow deductively from a set of premises all of which are already true before this belief is formed. And since the belief component of the intention to ϕ does not imply that one has this belief, it can be formed by way of a self-conscious process of reasoning.

I conclude, therefore, that if one is to be a cognitivist about practical reason, then this is the account of intentions one should accept. And so we have sketched an answer to the question of *how* to be a cognitivist about practical reason. There remains, of course, the question of *whether* to be a cognitivist about practical reason, a question whose answer goes beyond the scope of this paper.

APPENDIX A: INTENTIONS AND PRACTICAL ACCEPTANCE

Even if we reject the Strong Belief Thesis, we might still hold that a related thesis is true. That is, we might hold that while intending to ϕ needn't involve believing that one will ϕ, it does involve planning on the basis of the supposition that one will ϕ, or in other words, taking it for granted, in the context of practical reasoning, that one will ϕ. Let us use the term "acceptance," or "acceptance from the practical point of view," to refer to this attitude of taking a proposition for granted in the context of practical reasoning. We may then formulate this suggestion as the *Strong Acceptance Thesis*:

> A necessary condition for intending to ϕ is *accepting* that one will ϕ in all contexts of practical reasoning.

The view that intending to ϕ involves accepting that one will ϕ, and that rational requirements on intentions can be explained in terms of rational requirements on the attitude of acceptance, should not be described as a purely cognitivist view. For the attitude we are calling acceptance is not a purely cognitive attitude: it does not aim at truth, accuracy, verisimilitude, knowledge, understanding, or any other cognitive aim. More generally, practical acceptance does not have a mind-to-world direction of fit: it does not aim to represent the world as it really is. It aims rather at the effective guidance of action, and for this reason it can be rational to accept a proposition from the practical point of view even when one lacks sufficient reason to believe that it is true, and even when one has sufficient reason to believe that it is false. Thus, in some cases, we can take one theory for granted in the context of practical reasoning (say, classical mechanics) which we do not believe to be true, because the theory we believe to be true (say, quantum mechanics) is too complicated or unwieldy to employ in the context of practical reasoning, and we can serve our interests better by treating the simpler theory as true.[13] In some cases, we can rationally take things for granted in the context of practical reasoning because doing so puts us in a beneficial frame of mind. Hence, one can rationally take it for granted, in certain contexts of practical reasoning, that the client one is representing is innocent, or that one's spouse is faithful, or that one will recover from one's illness, even though one does not have sufficient reason to believe these claims. And in some cases, we can rationally take a proposition for granted in the contest of practical reasoning without believing that it is true

FN:13

[13] Someone may object that what we take for granted is not that classical mechanics is true, but rather that classical mechanics is approximately true. But this is not right. For reasoning on the supposition that something is approximately true is very different from, and far more complicated than, reasoning on the supposition that it is true. For example, on the supposition that *p* and *q* are true, we can infer any conclusion that is entailed by the conjunction of *p* and *q*. But from the supposition that *p* and *q* are approximately true, we cannot infer that the propositions that are entailed by their conjunction are approximately true.

because we know that, if the proposition is false, it will make little difference to how we act.[14]

FN:14

But while practical acceptance should not be described as a cognitive attitude, we might describe it as a quasi-cognitive attitude. For one thing, while it does not, in general, have a mind-to-world direction of fit, it doesn't in general have a world-to-mind direction of fit either: someone who accepts classical mechanics in the context of practical reasoning does not thereby have a tendency to bring it about that classical mechanics is true. Further, like belief, to accept a proposition from the practical point of view is in some sense to treat it as true. Like belief, acceptance seems to be governed by a norm of consistency, as it seems that in the context of practical reasoning, one cannot rationally accept inconsistent propositions. Similarly, like belief, it seems to be governed by a principle of closure, in the sense that there is rational pressure to accept, or take for granted, the propositions that follow from other propositions that one accepts. But if we assume that intention involves acceptance, and that acceptance is governed by analogues of all the norms on belief that figure in standard cognitivist explanations of the norms on intention, then we might expect to be able to give an explanation of these norms on intention that has all virtues of the standard cognitivist explanation, but without invoking the Strong Belief Thesis. And since acceptance appears to be a quasi-cognitive attitude, it would seem that such an explanation should count, if not as a purely cognitivist explanation of the requirements of practical rationality under consideration, at least as a quasi-cognitivist explanation—or, to borrow an expression from Oliver Roy, who proposes an explanation of this kind, a "hybrid cognitivist explanation."[15]

FN:15

We saw that the intention consistency requirement can be explained in terms of the Strong Belief Thesis in conjunction with *Belief Consistency*:

> One ought rationally to be such the set of all the propositions one believes is logically consistent.

But if we move from the Strong Belief Thesis to the Strong Acceptance Thesis, then we must move from the requirement of Belief Consistency to *Acceptance Consistency*:

> One ought rationally to be such that the set of all the propositions one accepts is jointly consistent with one's beliefs.

For if Acceptance Consistency is true, and if a necessary condition for intending to ϕ is accepting that one will ϕ, then one ought to be such that the set of propositional contents of one's intentions is jointly consistent with one's beliefs. Hence, Acceptance Consistency, in conjunction with the Strong Acceptance Thesis, entails Intention Consistency.

[14] I discuss this last type of case in "Rejecting Ethical Deflationism" (*Ethics* 116, July 2006) and in my dissertation, "Rational Acceptance and Practical Reason" (2006, Rutgers University).

[15] For an explanation along these lines, worked out in much more detail than I have provided here, see ch. 6 of Olivier Roy's *Thinking before Acting: Intention, Logic, Rational Choice* (Amsterdam: Institute for Logic Language and Computation, 2008).

Unfortunately, there are problems with this account. For one cannot plausibly accept the Strong Acceptance Thesis while denying the Strong Belief Thesis. For suppose the strong belief thesis is false, and hence that one can intend to ϕ without believing one will ϕ. Suppose, in particular, that Michael intends to stop at the bookstore on his way home from work, but he does not quite believe that he will do so, since he knows he might forget. And suppose he is offered a bet, called bet X, with the following payoff structure: if he accepts bet X, and he goes to the bookstore, then he will make a profit of 25 cents, but if he accepts bet X and fails to stop to the bookstore, then he will be tortured for the rest of his life. It seems clear that, in this case, Michael might well not take it for granted that he will stop at the bookstore in the context of deciding whether to take bet X. Indeed, if Michael is rational, then he will not take this for granted, since taking this for granted would license taking bet X, and taking bet X would be a rational choice only for an agent who, unlike Michael, is fully confident that he will stop at the bookstore. Hence, it would seem that, if Michael can intend to stop at the bookstore without being fully confident that he will do so, then he can also intend to stop at the bookstore without taking it for granted that he will do so whenever he is engaging in practical reasoning. More generally, if it is possible for an agent to intend to ϕ without believing that she will ϕ, then it is possible for an agent to intend to ϕ without taking it for granted that she will ϕ in every context of practical reasoning. In other words, if the Strong Belief Thesis is false, then the Strong Acceptance Thesis must also be false.

And so the latter thesis must, at the very least, be weakened. One possible weakening, endorsed by Olivier Roy, is the *Moderate Acceptance Thesis*:

> A necessary condition for intending to ϕ is accepting that one will ϕ in every context of practical reasoning in which the intention to ϕ is relevant.

But the above example is as much a counterexample to the Moderate Acceptance Thesis as it is to the Strong Acceptance Thesis. For in the context of deciding whether to accept bet X, Michael's intention to stop at the bookstore is clearly relevant. And yet, if we deny the Strong Belief Thesis, then we should allow that Michael could intend to go to the bookstore, without accepting the proposition that he will do so in the context of deciding whether to take bet X.

We might try weakening the thesis still further, by the *Weak Acceptance Thesis*:

> A necessary condition for intending to ϕ is accepting that one will ϕ in *some* contexts of practical reasoning.

Unfortunately, however, this weaker requirement does not suffice to explain the requirement of intention consistency. For if all that is involved in intending to ϕ is accepting that one will ϕ *in some context or other*, then an agent who has two contradictory intentions may accept that she will ϕ in one context of practical reasoning, and she may accept that she will not ϕ in a distinct context of practical reasoning. But when contradictory acceptances are compartmentalized in this manner, then they need not involve any irrationality. It would not be irrational, for example, if an agent took it for granted that space is Euclidean in the context of planning a trip to Kalamazoo, and if this same agent took it for granted that the space is non-Euclidean in the context of planning a trip to Mercury. Nor would it be irrational for a lawyer to take it for granted that her client is an upstanding citizen when she is defending him in court, and yet to take it for granted that he is a scroundrel when she encounters him in non-professional contexts.[16]

Thus, if the cognitivist aims to provide a cognitivist explanation of the Intention Consistency Requirement, she has little to gain from moving from the claim that intention involves belief to the claim that intention involves practical acceptance. For if the latter claim is given a weak formulation, and states only that intending to ϕ involves accepting that one will ϕ in some contexts of practical reasoning, then it will not suffice to explain the Intention Consistency Requirement. And if, on the other hand, this claim is given a stronger formulation, and states that intending to ϕ involves accepting that one will ϕ in every context of practical reasoning, or in every relevant context of practical reasoning, then this claim will be implausible apart from the Strong Belief Thesis.

What about the means–end coherence requirement? Can it be explained on the basis of a quasi-cognitivist view of intentions according to which intending to ϕ involves accepting that one will ϕ? Recall that when we were considering straightforward cognitivist views according to which intending to ϕ involved the *belief* that one will ϕ, the most promising attempts to explain the Means–End Coherence Requirement appealed to *Modus Ponens Requirement*:

> One ought rationally to be such that (if one believes that p and one believes that (if p then q) then one believes that q).

This principle, together with the Strong Belief Thesis, entails that if one believes that one will ψ, and believes that one will ψ only if one ϕs, then one is rationally required to believe that one will ϕ. Now if we move from the Strong Belief Thesis to the Strong Acceptance Thesis, then we will need to explain why anyone who intends to ψ, and hence accepts that she will do so, and believes that ϕ-ing is a necessary means to ψ-ing, is rationally required to accept that she will ϕ. And to explain this, we will need something like the principle *Hybrid Modus Ponens*:

> One ought rationally to be such that (if one accepts that p and one believes that (if p then q) then one accepts that q).

However, Hybrid Modus Ponens appears to be false. For suppose I am convinced that Einstein was right, and that space is non-Euclidean. Indeed, suppose I believe the following material condition: if space is Euclidean, then I am a monkey's uncle. Suppose, however, that in a context of practical reasoning in which I am planning a trip to Kalamazoo, I accept, or take it for granted, that space is Euclidean. This would not seem to involve any irrationality. But in this case, Hybrid Modus Ponens implies that in planning my trip to Kalamazoo, I am rationally required to accept

[16] For a discussion of such contextual variations in acceptance, see Micheal Bratman's "Practical Reasoning and Acceptance in a Context"

that I am a monkey's uncle, or in other words, that the only way I can be fully rational, given the beliefs and acceptances just described, is to take it for granted, in this context, that I am a monkey's uncle. And this, of course, is not a very plausible implication.

Even if this problem could be solved, there is a remaining problem. For it is not plausible that accepting that one will ϕ, along with whatever non-cognitive background conditions there may be for intending to ϕ, is sufficient for intending to ϕ. Recall the example from Section 2.1 in which Barry the banker intends to protect the million dollars, and he believes that locking the money in the safe is a necessary means to protecting the money. Recall that he also believes that he will lock the money in the safe, but he believes this simply but because he has read that he will do so in his putative biography he was given by Robby the robber. In this case, Barry clearly satisfies the non-cognitive background conditions for intending to lock the money in the safe. We might further suppose that Barry is so convinced by the statements he reads in the biography that he accepts them all from the practical point of view. Thus, he takes it for granted, in practical reasoning, that he will lock the money in the safe. Even so, such acceptance will not amount to the intention to lock the money in the bank. And so while Barry is taking it for granted that he will lock the money in the safe, Robby will be able to run off with the money.

It seems, therefore, that the quasi-cognitivist view of intentions, like the pure cognitivist view, fails to provide a satisfactory account either of the consistency requirement or of the means–end coherence requirement.

APPENDIX B: THE SELF-REFERENTIAL INTENTION THESIS

In "Practical Reasoning," Harman argues for the following two theses:

> *Strong Belief Thesis* (**SBT**): Anyone who intends to ϕ believes that she will ϕ.
>
> *Self-Referential Intention Thesis* (**SRI**): Anyone who intends ϕ intends to (ϕ because of her current intention to ϕ).

Together, these two theses entail the Strong Belief Thesis. For it follows from these two theses that anyone who intends to ϕ believes that she will (ϕ because of her current intention to ϕ). And so it follows from these theses that the intention to ϕ involves a belief that refers to this very intention. But if a belief is a component of an intention, then in referring to this intention, it will refer, at least indirectly, to itself. And so it follows from the above theses that intentions involve self-referential beliefs.

Since I discussed the Strong Belief Thesis in Section 1, here I will focus on the Self-Referential Intention Thesis. I will begin by criticizing the standard arguments in favor of this thesis, and I will then present an argument against it.

The first argument for SRI to consider, presented by Harman, can be stated as FN:17 follows:[17]

A1. One cannot arrive at the intention to ϕ by way of practical reasoning if one believes that one's intention to ϕ would not result in one's ϕ-ing.

A2. The only natural way to explain A1 is to assume that every intention to ϕ is an intention to ϕ-because-of-that-very-intention.

A3. Hence we can conclude, by inference to the best explanation, that anyone who intends to ϕ intends to (ϕ because of her current intention to ϕ).

I believe that the first premise of this argument is true, but that the second is false. There is another natural way to explain A1, namely this: intending to ϕ is an attitude one has *in order to* ϕ. And there is a *negative* causal belief condition on the in-order-to relation: one cannot ψ in order to ϕ if one believes that ψ-ing would not result in one's ϕ-ing. For example, one cannot practice in order to win a contest if one believes that practicing would not result in one's winning the contest. However, there is no corresponding *positive* causal belief condition on the in-order-to relation: it is not the case that one can only ψ in order to ϕ if one believes that one's ψ-ing would result in one's ϕ-ing. For example, it is not the case that one can only practice in order to win the contest if one believes that one's practicing would result in one's winning the contest; one might, after all, be uncertain as to whether practicing would have this result. Thus, if the correct explanation of A1 is the fact that intending to ϕ is something one does in order to ϕ, then there will be a negative causal belief condition on intending to ϕ: it will be impossible to intend to ϕ while believing that one's intending to ϕ will not result in one's ϕ-ing. There need not, however, be any positive causal belief condition on intending to ϕ, and so it may be possible to intend to ϕ without believing that one's intention to ϕ will result in one's ϕ-ing. Thus, if we assume that intending to ϕ is something one does in order to ϕ, which is an assumption that Harman himself makes elsewhere, then we can give a very straightforward explanation of A1 that does not commit us to A2.

A second, and closely related, argument for SRI, presented by Kieran Setiya, can FN:18 be stated as follows:[18]

B1. One cannot coherently intend to (ϕ, but not because of one's current intention to ϕ).

B2. The only natural way to explain B1 is to assume that every intention to ϕ is an intention to ϕ-because-of-that-very-intention.

B3. Hence we can conclude, by inference to the best explanation, that anyone who intends to ϕ intends to (ϕ because of her current intention to ϕ).

But again, the second premise is false, because there is an alternative explanation of B1. For B1 follows from a negative causal belief condition on intentions, discussed

[17] See Harman's "Desired Desires," in Ray Frey and Chris Morris, eds., *Value, Welfare, and Morality* (Cambridge: Cambridge University Press: 1993), 138–57.

[18] See his "Cognitivism about Instrumental Reason" *Ethics* 117 (July 2007), 649–73.

above, in conjunction with the strong belief thesis. That is, B1 follows from these two premises:

NC. One cannot intend to ϕ while believing that one's current intention to ϕ will not result in one's ϕ-ing

SB. Anyone who intends to ϕ believes that she will ϕ.

For by SB, if one intends to (ϕ, but not because of one's current intention to ϕ), then one must believe that one's current intention to ϕ will not result in one's ϕ-ing. And hence it follows, by NC, that one cannot intend to ϕ. But anyone who intends to (ϕ, but not because of one's current intention to ϕ) must intend to ϕ. Therefore, if NC and SB are true, then it is impossible to intend to (ϕ, but not because of one's current intention to ϕ). And so one can explain B1 without appealing to B2.

Of course, one might reject the above explanation of B1 on the grounds that one rejects the strong belief thesis, SB. But, as we have seen, this move is not available to the cognitivist about practical reason.

FN:19 A third argument for RI, again from Harman, can be presented as follows:[19]

C1. Intending to ϕ is something that one does in order to ϕ, and that one recognizes that one does in order to ϕ.

C2. But if one intends to ϕ, and recognizes that one is ψ-ing in order to ϕ), then one intends to (ϕ because one now ψs).

C3. Therefore, if one intends to ϕ, and recognizes that one (intends to ϕ in order to ϕ), then one intends to (ϕ because one now intends to ϕ).

C4. Therefore, in intending to ϕ, one intends to (ϕ because one now intends to ϕ). That is, anyone who intends to ϕ intends to (ϕ because of her current intention to ϕ).

Once again, the second premise is false. Suppose I will be competing in a musical contest. Suppose I know that my competitors are incompetent, and so I will win the contest so long as I practice adequately. I don't know, however, what instrument I will be required to play: it may be the harpsichord, or it may be the kazoo. So I practice playing both. In this case I intend to win the contest, and I recognize that I am practicing playing the harpsichord in order to win the contest, but I do not intend to (win the contest because I am practicing playing the harpsichord). After all, I recognize that I may not play the harpsichord in the contest at all, and so my current practicing may not play any causal role in my winning the contest. Hence, it would seem that one could intend to ϕ, and recognize that one intends to ϕ in order to ϕ, without intending to (ϕ because one intends to ϕ).

FN:20 A final standard argument for RI, presented by John Searle, can stated thus:[20]

D1. The content of a propositional attitude is the proposition that is true just in case that mental state is satisfied.

[19] See his "Practical Reasoning," *Review of Metaphysics* 29 (1976), 431–63.

[20] See his *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

D2. The intention to ϕ is a propositional attitude that is satisfied only if one ϕ-s because of that very intention.

D3. Therefore it must be part of the content of the intention to ϕ that one ϕ-s because of that very intention.

D4. Therefore, anyone who intends to ϕ intends to (ϕ because of her current intention to ϕ).

FN:21 Once again, the second premise can plausibly be rejected.[21] Suppose that now, at noon, I intend to brush my teeth tonight before going to bed. And suppose I do brush my teeth tonight before going to bed. In such a case, we would normally say that my prospective intention to brush my teeth is satisfied. And if we learned that that in the evening I brushed my teeth out of habit, and that the prospective intention I had formed at noon to brush my teeth in the evening played no causal role in bringing about my evening tooth-brushing activity, we would not normally say that my prospective intention was therefore unsatisfied. Thus, one can plausibly maintain that an intention to ϕ can be satisfied even if it is not the case that one ϕ-s because of this intention.

Now it may be that prospective intentions (intentions we have prior to the time of the intended action) refer to concurrent intentions (intentions we have at the same time as the intended action). That is, it may be that whenever we have a prior intention to ϕ, we intend that at some future time we *ϕ because we then have a concurrent intention to be ϕ-ing*. In this case, the prior intention to ϕ will refer to a concurrent intention ϕ, and will only be satisfied if one ϕ-s because of a concurrent intention to ϕ. But the former intention still needn't be self-referential, since it may be that the only intention that figures in the content of the prospective intention is the later, concurrent intention, and not the prospective intention itself.

Thus, the standard arguments for SRI do not appear to be successful. Moreover, there is strong reason to reject SRI. For it seems that if we have reason to intend to ϕ, then we have reason to perform actions that are necessary in order to guarantee that we ϕ. Therefore, if intending to ϕ involved intending to ϕ because of this very intention, then anyone who has reason to intend to ϕ has reason to act in ways that are necessary to guarantee not that his ϕ-ing result from his current intention to ϕ. But this does not seem to be the case. Suppose, for example, that it is noon, and I intend to brush my teeth at midnight. Suppose I believe that if I leave my toothbrush in its usual place, I may brush my teeth not because of my current intention. In this case, moving my toothbrush to an unusual location would be a necessary means to ensuring that I brush my teeth because of my prior intention, and not merely out of habit. And yet it would seem that I could have reason to intend to brush my teeth tonight without having reason to move my toothbrush to an unusual location. Hence, it seems that in intending, at noon, to brush my teeth tonight, I needn't intend to brush my teeth tonight because of my current intention to do so.

[21] Alfred Mele discusses this argument in "Are Intentions Self-referential?" (*Philosophical Studies* 52 (1987) 309–29. His diagnosis is that we should reject the first premise and retain the second.

Queries in Chapter 0

Q1. This Chapter title is not matching with the title mentioned in toc and NTP.

Q2. In this paragraph closing Paraenthesis is Missing.