

## CHAPTER 2

# *Intention, belief, and instrumental rationality\**

Michael E. Bratman

### I TWO APPROACHES TO INSTRUMENTAL RATIONALITY

Suppose I intend end E, believe that a necessary means to E is M, and believe that M requires that I intend M. My attitudes concerning E and M engage a basic requirement of practical rationality, a requirement that, barring a change in my cited beliefs, I either intend M or give up intending E.<sup>1</sup> Call this the *Instrumental Rationality* requirement – for short, the *IR* requirement.<sup>2</sup>

\* This essay is a sequel to my “Intention, Belief, Practical, Theoretical,” in Simon Robertson, ed., *Spheres of Reason* (Oxford: Oxford University Press, forthcoming). Some of the ideas developed here are also in that earlier essay, but I hope in this present chapter to go somewhat beyond that earlier work. I do, however, see my overall argument in favor of (to use terminology to be introduced in the main text) the practical commitment view, in contrast with cognitivism, as drawing on both of these essays (as well as on the basic account presented in my *Intention, Plans, and Practical Reason* [Cambridge, MA: Harvard University Press, 1987; reissued by CSLI Publications, 1999]). The present essay was motivated in part by Kieran Setiya’s “Cognitivism about Instrumental Reason: Response to Bratman” (2005), which constituted his very thoughtful and helpful comments on “Intention, Belief, Practical, Theoretical,” at the Conference on Practical Reason, University of Maryland, April 2005. My present chapter has also benefited from conversation with John Perry, Jennifer Morton, and Sarah Paul, detailed comments from Gideon Yaffe and John Broome on earlier drafts, and very helpful comments from George Wilson and from the editors of this volume.

<sup>1</sup> Concerning the need for the belief that M requires that I intend M, see Robert Binkley, “A Theory of Practical Reason,” *The Philosophical Review*, 74 (1965): 423–448, at 443. The language of “requirement” comes from John Broome; see “Normative Requirements,” *Ratio*, 12 (1999): 398–419.

<sup>2</sup> In other work I have focused on what I have called a requirement of means-end coherence of one’s intentions and plans. I see IR as a central aspect of that requirement, though the requirement of means-end coherence goes beyond IR, strictly speaking, in requiring that an agent fill in her plans with one or another sufficient means when what is needed is that the agent settle on some such means or other. And the requirement of means-end coherence allows for delay in filling in plans with means when there remains sufficient time. (For some of these complexities see *Intention, Plans, and Practical Reason*: 31–35.) But IR is at the heart of the requirement of means-end coherence, and it will simplify my discussion here to focus on it. Note that both IR and the requirement of means-end coherence specifically concern means-end rationality with respect to *intended* ends; there remain further issues about means-end rationality concerning things one wants, prefers, or values.

Suppose now that I believe that E, and I also believe that E will only occur if M. My beliefs engage a basic demand of theoretical rationality, a demand that, roughly, either there be a change in at least one of these two beliefs or I believe M. Call this the *Belief-Closure* requirement – for short, the BC requirement. BC, note, is not a consistency demand on my beliefs: failure to add the further belief that M need not involve inconsistency in the way that adding a belief that not-M would. Nevertheless, something like BC seems a basic rationality constraint on belief.<sup>3</sup>

Both IR and BC express constraints on the coherence of the agent's relevant attitudes; and these constraints are aspects of the normal rational functioning, in the psychic economy of believing-and-intending agents, of the cited attitudes. The intentions and beliefs of such agents will tend to be responsive to these constraints. But the requirements differ in important ways. IR is engaged only if I intend E; whereas BC is engaged if I believe E, whether or not I intend E. And a central way of meeting the demands of IR involves intending M; whereas a corresponding way of meeting the demands of BC involves, rather, believing M. Further, if we ask why these principles – IR and BC – are, indeed, aspects of the rational functioning of the cited attitudes, we arrive, I believe, at importantly different answers. Roughly: In the case of BC we will appeal, I think, to something like a general need for coherence of one's beliefs if one is to understand the world. In the case of IR we will appeal, I think, to something like a general need to intend necessary means if one is to be an effective agent and if one is to have a practical standpoint that has the kind of efficacy characteristic of self-government.<sup>4</sup>

These last claims about what lies behind IR, on the one hand, and BC, on the other, are, of course, sketchy; and I cannot pursue these matters in detail here. I do think, though, that there is here a general and plausible idea. This is the idea that these stories will differ, and that one will cite

<sup>3</sup> There are important issues, in understanding BC, about what Gilbert Harman calls “clutter avoidance”: as Harman emphasizes, we do not suppose one must add *all* beliefs entailed by other beliefs one already has. See Gilbert Harman, *Change in View* (Cambridge, MA: MIT Press, 1986): 12. These issues do not arise in the same way for IR. Given that the agent believes that intending M is itself necessary for E, forming that intention will not be mere “clutter.” I return to this matter briefly below in n. 21.

<sup>4</sup> That is, the concerns and commitments that constitute the practical standpoint with which one identifies need to be ones that are effectively in control of one's intentional conduct, if one is to be self-governing. One aspect of such effective control will be conforming, in general, to IR when one's beliefs about what is required are accurate and when the intended ends are elements in, or in other ways endorsed by, the practical standpoint with which one identifies. I try to deepen this connection between central norms on intention and self-government in “Intention, Belief, Practical, Theoretical.”

basic theoretical concerns – with understanding, for example – and the other will cite basic practical concerns – with effective agency and self-governance, for example. When we put this very general idea together with our initial observations about the differences between IR and BC – the focus on intention, in the one case, and belief in the other – we are led to the view that these are importantly different demands of rationality – in one case practical, in the other case theoretical – though these demands will, of course, significantly interact in many cases.

There are, however, philosophical pressures that have led a number of philosophers to draw principles along the lines of IR and BC much more closely together. Their idea, roughly, is to see IR, or something close to it, as, at bottom, a special case of the theoretical requirement expressed in BC, or something close to it, together perhaps with some further principle of theoretical rationality. There are different versions of this idea, as we shall see. But what they share is the idea that IR is, at bottom, a theoretical demand on beliefs. This is *cognitivism* about instrumental rationality.<sup>5</sup>

Cognitivism about instrumental rationality identifies what had seemed to be a basic element of practical rationality with theoretical rationality. It need not, however, say that all demands of practical reason are, at bottom, demands of theoretical reason.<sup>6</sup> So, for example, I see John Broome, Wayne Davis, Gilbert Harman, Kieran Setiya, J. David Velleman, and R. Jay Wallace as, in different ways, cognitivists about instrumental rationality;<sup>7</sup> but whereas Velleman is, quite broadly, a cognitivist about practical reason – he sees practical reason as grounded in a theoretical

<sup>5</sup> More precisely, this is cognitivism about that aspect of instrumental rationality that IR, and closely related principles, aims to capture. For this use of the term “cognitivism” (in contrast with its standard use in meta-ethics) see my “Cognitivism about Practical Reason,” as reprinted in my *Faces of Intention* (New York: Cambridge University Press, 1999), and my “Intention, Belief, Practical, Theoretical.” Kieran Setiya also uses this term in this way in “Cognitivism about Instrumental Reason,” *Ethics*, 117 (2007): 649–673.

<sup>6</sup> Setiya makes this point in “Cognitivism about Instrumental Reason: Response to Bratman.”

<sup>7</sup> See John Broome, “The Unity of Reasoning?,” in Jens Timmerman, John Skorupski, and Simon Robertson, eds., *Spheres of Reason* (Oxford: Oxford University Press, forthcoming); Wayne Davis, “A Causal Theory of Intending,” *American Philosophical Quarterly*, 21 (1984): 43–54; Gilbert Harman, “Practical Reasoning,” reprinted in Gilbert Harman, *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press, 1999): 46–74; Setiya, “Cognitivism about Instrumental Reason”; J. David Velleman, *Practical Reflection* (Princeton, NJ: Princeton University Press, 1989), *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), and “What Good is a Will?,” in Anton Leist and Holger Baumann, eds., *Action in Context* (Berlin and New York: de Gruyter and Mouton, 2007); R. Jay Wallace, “Normativity, Commitment, and Instrumental Reason,” *Philosophers’ Imprint*, 1 (3) (2001). Setiya’s cognitivism about instrumental rationality goes directly by way of his version of BC. My strategy in this essay of focusing on the role of BC in such cognitivism follows Setiya in this respect.

concern with self-knowledge and self-understanding – the cognitivism defended by the others is more limited: it extends at least to IR but does not purport to extend to all of practical reason. I do think that it would be difficult to be a cognitivist about IR but not about certain other rationality requirements on intention. This is clearest with respect to the consistency requirement on intentions. This is the requirement that one's overall set of intentions be consistent as well as consistent with one's beliefs: one needs to be able to put one's various intentions together – to agglomerate them – into an overall plan that is internally consistent and consistent with one's beliefs.<sup>8</sup> A marriage of cognitivism about IR with a rejection of cognitivism about these consistency demands is likely to be unstable, since the alternative story about the consistency demands will threaten to spill over to a story about IR. It is a hard question what a cognitivist about IR should say about norms of cross-temporal stability of intention. In any case, my primary focus here will be on cognitivism about IR.

Here is one way to look at it. Suppose you now prefer A at some later time to its envisaged alternatives. This is not yet to intend A. After all, while you do now prefer A you may still see the issue as not yet settled and be engaged right now in further deliberation about whether to A. Now suppose that this further deliberation does issue in an intention to A. In what, precisely, does this transition from a preference to an intention consist? In particular, in now intending to A you come to be under the rational pressure of IR: you now need, roughly, to intend known necessary means, or give up your intention to A. In contrast, the mere preference for A did not, by itself, engage IR. What about the transition from preference to intention explains why IR is newly engaged?<sup>9</sup>

A cognitivist about IR will see the transition from preference to intention as at least in part a matter of belief. And a cognitivist about

<sup>8</sup> Note that this demand for consistency is not just that each intention have a consistent content; it includes, as well, the demand that one be able to agglomerate one's various intentions into an overall intention that has a consistent content. I will sometimes emphasize this way in which this demand for consistency involves an implicit agglomerativity demand; but sometimes, for ease of exposition, I will simply speak of consistency, leaving the idea of agglomeration implicit. As this formulation of the consistency demand suggests, there are puzzles here that parallel puzzles about consistency of belief and the "preface paradox"; but I put these aside here.

<sup>9</sup> Cf. R. Jay Wallace: "there must be something about the attitude of intending to do *x* that goes beyond the attitude of desiring that one do *x*, in a way that brings a distinctively rational requirement into play" ("Normativity, Commitment, and Instrumental Reason": 18). (Though note that I allude to preference whereas Wallace appeals to desire.) As my remarks below indicate, though, I think we need to be careful here, since the step from preference to an attitude that engages something like IR might be only a step to a "settled objective" and not a step all the way to intention, strictly speaking.

IR will appeal to the belief aspect of this transition to explain why IR is newly engaged. In contrast, on the view I would like to defend – and the view suggested by our preliminary remarks about the differences between IR and BC – the step from preference to intention is a practical step, a step to a distinctive kind of practical commitment that is not itself a belief. And it is that, not belief, that is at the heart of the new applicability of IR. So we can contrast cognitivism with such a *practical commitment* view of instrumental rationality.

A basic idea that underlies the kind of practical commitment view I would want to defend is that intentions are elements of a planning system, one that has fundamental roles in the coordination and control of action. Planning agents like us normally need to intend means if we are to achieve our intended ends. We are not gods who can simply and effectively will “let there be E!” (If I were simply to will “let there be light!” it would not work: I need to flip the switch.) And insofar as one fails to intend means intending which is necessary for intended ends, this planning system will fail to be effective.<sup>10</sup> Further, insofar as one’s intentions are inconsistent with each other and/or with one’s beliefs, this planning system will fail in its coordinating role, a role that is at the heart of the cross-temporal effectiveness of that system. So, in general, conformity to norms of consistency and means-end rationality are – at least for non-divine planning agents with reliable beliefs about the world – conditions for the successful operation of this system of coordinated control. Further, a full-blown planning agent will not just happen to conform to such norms: she will think in ways that are at least implicitly guided by these norms; and this will be part of the explanation of the successful functioning of her planning agency. In particular, she will be responsive in her thinking to the need to intend means intending which is needed for her intended ends; and she will be responsive in her thinking to a demand for consistency of intentions and beliefs. Norms of means-end rationality and consistency will be, for her, internal norms.<sup>11</sup> Her intention in favor of E is a practical commitment in part in the sense that it engages these internal norms.

There is a complexity here, however. IR is tied to the normal, successful functioning of the planning system as a system that effectively controls action in pursuit of intended ends. A norm of consistency of intention is

<sup>10</sup> Here I am in agreement with Joseph Raz, “The Myth of Instrumental Rationality,” *Journal of Ethics and Social Philosophy*, 1 (1), [www.jesp.org](http://www.jesp.org), (2005): 17.

<sup>11</sup> For this terminology see *Intention, Plans, and Practical Reason*: 109.

to the normal, successful functioning of the planning system as a coordinating system. Intentions are elements of a planning system whose central roles are those of effective coordinating control. However, we are also capable of commitments to ends that we do not treat as subject in precisely the same way to the sorts of coordinating pressures that impose a demand for agglomeration and consistency. I might try to get into Harvard Law, and try to get into Stanford Law, while knowing that these law schools coordinate admissions and so that, while I have a shot at getting into each, it is not possible to get into both.<sup>12</sup> As I see it, I do not, strictly speaking, intend to get into Harvard Law, or intend to get into Stanford Law. This is because if I intended one I would, by the symmetry of the case, intend the other; so I would thereby have intentions that violate rational demands for agglomeration and consistency. Nevertheless, I do have getting into Harvard Law as what Hugh McCann calls a “settled objective”;<sup>13</sup> and similarly concerning getting into Stanford Law.

Such examples seem to show that, while each settled objective needs to be internally consistent, not all settled objectives engage agglomeration and consistency demands in precisely the way characteristic of intention. In some cases a settled objective is not embedded in the standard intention-like way within the overall coordinating role of one’s planning system, though the pursuit of the settled objective does impose more localized coordination pressures. (I need, for example, first to get the Harvard application, then fill it out; if I try to proceed in the reverse order it won’t work. And I need to ensure that my plan for filling out the application meshes with my other plans for the day.) The conclusion I draw is that not all settled objectives are intentions, strictly speaking<sup>14</sup> (though they may be associated with intentions – say, to get into some law school or other, or intentions about means to the objective). Nevertheless, settled objectives do engage a requirement that is something like IR: after all, in the envisaged case, I am under rational pressure to settle on means to get into Harvard Law, pressure I would not be under if I merely desired to get into that school. Rational pressures of means-end rationality can be engaged even if one’s attitude does not engage pressures of agglomeration and consistency in precisely the way characteristic of intention.

<sup>12</sup> This example is modeled on the video games example I discuss in *Intention, Plans, and Practical Reason*: chapter 8.

<sup>13</sup> Hugh McCann, “Settled Objectives and Rational Constraints,” *American Philosophical Quarterly*, 28 (1991): 25–36.

<sup>14</sup> It is in this conclusion that I disagree with McCann, though I will not here try to respond to his arguments against this conclusion.

Now, it seems that certain non-planning agents can still have settled objectives. Perhaps a squirrel can have getting the nuts as a settled objective even though it does not have sufficient structure in its thinking about the future to be a planning agent.<sup>15</sup> However, when a planning agent like us has certain settled objectives that are not, strictly speaking, intentions, pressures of means-end rationality engaged by such settled objectives will normally be met by forming *intentions* about means. Even if my commitment to getting into Harvard Law is a settled objective but not an intention, my sub-plans for pursuing this end will normally be intended, strictly speaking. This is because those sub-plans still need to mesh with my sub-plans for getting into Stanford Law. After all, the plan is for one and the same agent – namely, me – to carry out both sub-plans even though they are in pursuit of objectives that are, given the special features of the case, not co-possible.<sup>16</sup>

An implication of these reflections on settled objectives is that the two different norms on intention highlighted here – IR and a norm of consistency – have a slightly different status. A norm of means-end rationality that is similar to IR will be engaged by settled objectives, whereas the cited norm of agglomeration and consistency will not be engaged in the same way by settled objectives as by intentions, strictly speaking. But, having noted this complexity, I will focus primarily on the account of IR for intentions, since it is here that cognitivism about instrumental rationality has its best chance.

As I have said, I envisage an account of IR that ties it to the proper functioning of the planning system as a system of coordinated, effective control. And in other work I have also emphasized an even broader significance of this planning system in our lives as an element in both our self-governance and in our sociality.<sup>17</sup> By itself, however, this picture of our planning agency, and of the central roles in it of norms of agglomeration, consistency and means-end rationality, does not show that cognitivism is false. A cognitivist can agree with all this and then argue

<sup>15</sup> So the step from such a non-planning agent with settled objectives to a full-blown planning agent would be an important step in a Gricean “creature construction” model of agency. See Paul Grice, “Method in Philosophical Psychology (From the Banal to the Bizarre),” (Presidential Address), *Proceedings and Addresses of the American Philosophical Association* (1974–5): 23–53, and Bratman, “Valuing and the Will,” reprinted in *Structures of Agency: Essays* (New York: Oxford University Press, 2007).

<sup>16</sup> In this last sentence I have benefited from discussions with Luca Ferrero and Michael Nelson. And see *Intention, Plans, and Practical Reason*: 137–138 (though there I proceed in the language of “guiding desire” rather than McCann’s language of “settled objective”).

<sup>17</sup> “Intention, Belief, Practical, Theoretical.”

that these norms turn out, on examination, to derive from theoretical norms on associated beliefs. What these appeals to the proper functioning of the planning system do indicate, though, is that there is an initial plausibility to the idea that these norms are practical norms grounded in the practical roles of the planning system of coordinated control. If we are to be led to cognitivism we need some further arguments, ones that go beyond noting that we do, indeed, appeal to and depend on these norms insofar as we are planning agents. We need arguments for thinking that we should see these norms as, at bottom, theoretical; and we need a defense of the idea that theoretical norms really can do the requisite work. So let's see.

## 2 HARMAN'S BASIC IDEA

We can begin with what I will call *Harman's basic idea*, since it derives from his ground-breaking 1976 paper "Practical Reasoning" (though the way I will present it here goes a bit beyond that paper). Harman, like many cognitivists about IR,<sup>18</sup> supposes that intending E necessarily involves believing E. Further,

since intention involves belief, theoretical and practical reasoning overlap.

In theoretical reasoning, one seeks to increase the coherence of one's overall view of the world . . . Since intention involves belief, and theoretical and practical reasoning overlap, coherence must be relevant to any sort of reasoning about the future, theoretical or practical . . .

The thesis that intention involves belief associates practical reasoning about means and ends with theoretical reasoning. It brings these two sorts of reasoning under a single principle.<sup>19</sup>

<sup>18</sup> There are important qualifications. Wallace endorses only a weak belief condition: intending to *A* requires believing *A* is possible. (As I note in "Intention, Belief, Practical, Theoretical," this weak belief condition does not on its own explain the rational agglomerativity of intention.) Broome says only that if you believe you intend *A* then you believe *A*. I will return to this view of Broome below. In his 1976 paper Harman *identifies* intention with a kind of belief that one will *A*; whereas in a later paper ("Willing and Intending," in Richard Grandy and Richard Warner, eds., *Philosophical Grounds of Rationality*, Oxford: Oxford University Press, 1986: 363–380) Harman only says that intending *A* requires believing *A*. The argument for cognitivism I am now sketching is neutral as between these two ways of understanding the connection between intending *A* and believing *A*—though below (pp. 28–29), the purported identification will matter.

<sup>19</sup> *Reasoning, Meaning, and Mind*: 49–50. In the original essay, Harman had said that "[I]n theoretical reasoning, one seeks to increase the explanatory coherence of one's overall view of the world." In this more recent version of his essay he appeals broadly to coherence, where explanatory coherence is one kind of coherence (though the appeal specifically to *explanatory* coherence remains: 56 and 63).



Given the emphasis on the claim “that intention involves belief,” it seems that the “single principle” to which Harman alludes is a principle of coherence on one’s beliefs. It is because “intention involves belief” that this principle of coherence on belief extends to one’s intentions. And – though at this point I go beyond what Harman says – it seems plausible to see something like BC as an aspect of belief coherence that is importantly relevant in this way to IR. A violation of BC would normally be a form of belief incoherence,<sup>20</sup> and it is this form of incoherence that may seem to be at stake in violations of IR. After all, if I intend E (and so believe E) and believe M is a necessary means to E, BC requires that, if I do not change those two attitudes, I believe that M. If I also believe that M will obtain only if I intend M,<sup>21</sup> BC requires, if I do not change those two attitudes, that I believe I intend M. And it can seem that to believe I intend M I need actually to intend M (though this is a matter to which I will return). So the demand to satisfy BC – a demand that seems to be a central element in the requirement for belief coherence – seems to issue in a demand to satisfy IR. And that is cognitivism about IR.

As I am understanding it, then, Harman’s basic idea is that we arrive at cognitivism about instrumental rationality in two steps: we begin by noting that intention involves corresponding belief; we then reflect on the nature of the theoretical pressures on those involved beliefs – where, as I have developed this idea here, these pressures include BC. And now I want to point to two problems for this route to cognitivism about IR.<sup>22</sup> The first concerns the idea that intention involves corresponding belief. Suppose, to take an example I have discussed elsewhere, I intend to stop at the bookstore on the way home.<sup>23</sup> Still, I know that I am forgetful; so I am not confident that I will stop – after all, once I get on my bicycle I do have a tendency just to pedal on home. About this case I am inclined to say: I intend to stop, but I do not believe I will stop (though I do not believe

<sup>20</sup> Putting to one side issues of “clutter avoidance.” See n. 21.

<sup>21</sup> Setiya argues that a cognitivist may appeal to something like this belief to block, for such cases, Harman’s worries about “clutter avoidance.” (“Cognitivism about Instrumental Reason.”) In order to give the cognitivist a sympathetic hearing, I am following Setiya here.

<sup>22</sup> I also believe a third problem looms, a problem concerning the distinction between intending X and merely expecting that X will be a result of something one intends. See “Intention, Belief, Practical, Theoretical”; and for a discussion of this problem as it arises for Velleman’s theory, see “Cognitivism about Practical Reason,” reprinted in *Faces of Intention*. I put this potential third problem to one side here. My own discussion of this distinction is in *Intention, Plans, and Practical Reason*: chapter 10.

<sup>23</sup> *Intention, Plans, and Practical Reason*: 37, and “Practical Reasoning and Acceptance in a Context,” in *Faces of Intention*: 31–32. My use of this example in the present context benefited from discussion with John Perry.

I will not stop).<sup>24</sup> Many think, though, that it is a misuse of the word “intend” to say that I intend to stop but do not believe I will. Concerning this issue about the word “intend” there seems not to be a consensus one way or the other.<sup>25</sup> What matters for our present discussion, though, is not primarily what we would say using this word, but whether my attitude toward stopping, however labeled, engages the basic demand at the bottom of IR. And it seems to me that it does.

My attitude toward stopping is not a mere preference to stop: I have, rather, settled on stopping. If there were two routes home, only one of which went by the bookstore, my commitment to stopping at the bookstore would require that I take the route that goes by the bookstore. It would also require that I not settle also on another alternative known by me to be inconsistent with my stopping.<sup>26</sup> So my attitude toward stopping engages demands of consistency and means-end rationality that are characteristic of intention. And it does this even in the absence of a belief that I will stop. This suggests that we should seek an account of, in particular, IR that applies to my commitment to stopping at the bookstore, whether or not I also believe I will stop, contrary to Harman’s basic idea.

My second reason for being skeptical about Harman’s basic idea can allow, for the sake of argument, that if you intend A you believe A. What this second reason for skepticism involves is the idea that you can misidentify what you intend: you can falsely believe you have a certain intention. The mind, after all, is not an open book, even to the person

<sup>24</sup> As I note in “Practical Reasoning and Acceptance in a Context”: 32, I may nevertheless accept, in the context of relevant deliberation, that I would stop if I were to decide to stop. As I explain in that essay, acceptance in a context is not the same as belief. Wallace also alludes to something like this idea in “Postscript” to “Normativity, Commitment, and Instrumental Reason,” in *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason* (Oxford: Oxford University Press, 2006): 116. But Wallace seems to suppose that such acceptance is itself a kind of belief, whereas I would balk at this. These observations about acceptance in a context do raise the question of whether there is available to us a kind of cognitivism that goes by way of acceptance in a context, rather than belief. I have my doubts that what will emerge is really a kind of cognitivism, since the connection between intending and accepting in a context seems itself to be grounded in practical rationality; but I cannot pursue these matters here. Both Facundo Alonso and Olivier Roy have been pursuing this and related issues in unpublished work.

<sup>25</sup> For a lively sense of the disagreement here, see Paul Grice, “Intention and Uncertainty,” *Proceedings of the British Academy*, 57 (1971): 263–279; Donald Davidson “Intending,” reprinted in *Essays on Actions and Events*, 2nd edn. (Oxford: Oxford University Press, 2001): 83–102; Gilbert Harman, “Willing and Intending.”

<sup>26</sup> So my commitment to stopping is not a mere “settled objective,” in the sense in which my commitment to the end of getting into Harvard Law, in the example described earlier, is. In contrast with such mere settled objectives, my commitment to stopping at the bookstore not only requires settling on means, it also needs to be consistent with my other intentions.

whose mind it is. However we understand our special first-person access to our minds in general and, more specifically, to our intentions, it does not ensure incorrigibility about our own intentions.

In particular, there can be cases in which one believes one intends a means but does not intend that means. Perhaps I get confused and believe I intend to go shopping next Tuesday, though in fact I do not intend this but intend, rather, to go shopping next Thursday.<sup>27</sup> So there can be cases in which one intends E, believes that E requires both M and one's intending M, believes both M and that one intends M, but still does not, in fact, intend M. Though one believes one intends M one is mistaken about this. One's beliefs satisfy BC; but this does not ensure satisfaction of IR since, in the case envisaged, one does not in fact intend M. So it seems that we should reject cognitivism about IR.

These, anyway, are two concerns about Harman's basic route to cognitivism. In each case there are complexities we need to examine; and I will proceed to some of these complexities below. This will put me in a position to reflect on several related ideas that others have offered in defense of a version of cognitivism about IR. But first I want to note, and try to defuse, a line of argument that can seem to make cognitivism attractive.

### 3 CONSTITUTIVE AIM: VELLEMAN

Belief aims at truth. Or so it seems plausible to say.<sup>28</sup> Let me note three aspects of this idea. The first is that beliefs are embedded in a psychic economy that tends, in belief-formation, to track the truth, though of course it can on occasion fail. An attitude embedded in a psychic economy that, in its formation of that attitude, tracked instead the pleasant-to-think-of, would be a candidate for fantasy, not for belief. A second and closely related aspect of this idea is that, as Bernard Williams puts it, "truth and falsehood are a dimension of an assessment of beliefs."<sup>29</sup> Beliefs are criticizable if they are false, if they fail to track what – given the kind of attitudes they are – they tend to track. And, third, part of the explanation

<sup>27</sup> For this example, see my "Intention, Belief, Practical, Theoretical." I first discussed this objection to Harman's cognitivism in "Intention and Means-End Reasoning," *The Philosophical Review*, 90 (1981): 252–265, at 255–256, n. 4.

<sup>28</sup> A classic discussion of this idea is Bernard Williams, "Deciding to Believe," in *Problems of the Self* (Cambridge: Cambridge University Press, 1973): 136–151.

<sup>29</sup> *Problems of the Self*: 137.

of how, for agents like us, beliefs track the truth will appeal to an internal norm that assesses beliefs in terms of their truth.<sup>30</sup>

That said, it is a hard question precisely how these three aspects of the truth-directedness of belief are related. So, for example, at one point J. David Velleman thought that the second, normative aspect derived from the first, descriptive aspect.<sup>31</sup> In later work with Nishi Shah, however, Velleman has come to a more complex view of the relation between these two aspects.<sup>32</sup> These are difficult issues, and I will not try to sort them out here. It suffices for my purposes here simply to include all three, inter-related aspects in the idea that belief aims at truth.<sup>33</sup> Following Velleman, we can express this idea by saying that a “constitutive aim” of belief is truth.<sup>34</sup> And it seems plausible to say that demands on belief of consistency and coherence are closely related to this purported truth-aim.

So far, all this is neutral with respect to cognitivism. But one might think that the availability of such a story about various normative demands on belief supports the idea that these very same demands account for the cited normative demands on intention. Or at least this may seem plausible given the assumption that to intend A is, at least in part, to believe A. After all, if belief aims at truth there is, on this assumption, a story about the demand for consistency of intention: consistency is needed for the associated beliefs all to achieve their constitutive aim. Granted, even given the assumed connection between intention and belief, cognitivism about IR is more delicate, since failure of one’s associated beliefs to conform to BC does not ensure that any of

<sup>30</sup> Something like this seems implicit in Williams’ remarks about the underlying problem with believing at will. See *Problems of the Self*: 148.

<sup>31</sup> In “Introduction” to *The Possibility of Practical Reason*, Velleman claimed that “belief aims at the truth in the normative sense only because it aims at the truth descriptively” (2000: 17). Velleman discusses these matters further in “On the Aim of Belief,” in *The Possibility of Practical Reason*. His most recent discussion is in Nishi Shah and J. David Velleman, “Doxastic Deliberation,” *The Philosophical Review*, 114 (2005): 497–534.

<sup>32</sup> “There is both a descriptive and a normative component to belief’s truth-directedness.” “Doxastic Deliberation”: 530, n. 10. See also Nishi Shah, “How Truth governs Belief”, *The Philosophical Review*, 112(4) (2003): 447–482. Shah and Velleman, in this joint work, are concerned with the issue of how “to explain the fact that the deliberative question *whether to believe that p* is transparent to the question *whether p*” (“Doxastic Deliberation”: 497). I do not try to address this issue here.

<sup>33</sup> As does Williams, who also includes the further idea that in *saying* one believes *p* one is *claiming* that *p* is true. “Deciding to Believe”: 137.

<sup>34</sup> For the language of “the constitutive aim of belief,” see “Introduction”: 16. Since Shah and Velleman say that “there is both a descriptive and a normative component to belief’s truth-directedness,” I am assuming that talk of constitutive aim carries over to include, in this new work, both the descriptive and the normative component, even though the latter is no longer seen as derivable from the former.

them are false (though it does ensure that if one's actual beliefs do achieve their constitutive aim then one's set of beliefs do not maximally achieve this aim – since the failure of closure would be a failure to add a true belief). Still, there may seem to be some close connection between the truth-aim of belief and BC.<sup>35</sup> So if we could see IR as grounded in BC, we would then have the beginnings of an account of IR that grounds it in the purported constitutive aim of belief. So there may seem to be philosophical pressure in favor of cognitivism.<sup>36</sup>

I think, however, that it is important to see that if we are attracted to this idea that norms on belief are tied to its constitutive aim of truth, then we can argue, in parallel fashion, that norms on intention are tied to the (or anyway, a) constitutive aim of intention. And this parallel argument is independent of cognitivism. So appeal to constitutive aims of relevant attitudes does not, on its own, provide support for cognitivism.

In particular, if we are attracted to the appeal to constitutive aims of attitudes we can interpret the planning model of intention as articulating a constitutive aim of intention, namely: coordinated, effective control of action. Each intention aims at its realization in coordination with one's overall system of intentions. Coordination involves consistency among one's intentions, given one's beliefs; effective control requires that one intend means intending which one knows to be needed to achieve intended ends. So it is plausible that we can see norms of consistency and means-end rationality – norms characteristic of intention – as related to the (or, an) aim of intention in a way that parallels the relation between analogous norms on belief and the truth-aim of belief.

We can put the idea this way: The planning theory of intention articulates characteristic roles of intention in coordination and effective agency: intentions are embedded in a planning system that tracks coordination and effective control and systematically adjusts, when need be, in their direction. The planning theory sees the achievement of coordinated effectiveness as, to return to Williams' remark, "a dimension of an assessment" of intentions.<sup>37</sup> And the planning theory supposes that the explanation of how plans support coordination and effectiveness will involve associated internal norms. So the planning theory provides

<sup>35</sup> It may be that we should here appeal instead to a concern with understanding; but I put this complexity to one side here.

<sup>36</sup> An argument broadly in this spirit is in J. David Velleman, "What Good is a Will?" I discuss this essay in "Intention, Belief, Practical, Theoretical," where I also develop further the ideas to follow about the aim of intention.

<sup>37</sup> Though, of course, not all failures of coordination or effectiveness will be cases of irrationality.

resources that parallel the trio of ideas I have included within talk of the aim of belief. So we seem to have as much reason to appeal to an aim of intention as we do to an aim of belief – though, of course, the aims are different.

This does not require that it is essential to *agency*, quite generally, that one be a planning agent, one who forms intentions that have – or so we are now supposing – the constitutive aim of coordinated, effective control of action. Planning agency is a distinctive form of agency, one that contributes substantially to the pursuit of complex, temporally extended aims, to structures of self-governance, and – though I have not emphasized this here – to forms of sociality.<sup>38</sup> There can be agents who are not planning agents,<sup>39</sup> and these agents can even act intentionally in an attenuated sense that doesn't bring with it planning structures. But if you are, as we are, a planning agent<sup>40</sup> your intentions and plans have – we are now assuming – characteristic aims, aims associated with norms of consistency and means-end rationality. And this does not require cognitivism about IR.

This possibility of appealing to an aim of intention, in contrast with the aim of belief, tends to be obscured from within Velleman's cognitivist theory because, when he turns from belief to intention, he turns not to a distinctive aim of intention (other than – since intention is, on Velleman's view, a kind of belief – the aim of belief) but to a purported aim of, most generally, agency.<sup>41</sup> His central idea is that agency itself has an intellectual constitutive aim – namely self-knowledge and self-understanding. And this leads to Velleman's overall cognitivism about practical reason. But we can seek a practical parallel to the appeal to the aim of belief, without appeal to a purported aim of agency. We can talk, rather, of the aim of intention. And the planning theory gives us a plausible way to do that, a way that avoids cognitivism.

<sup>38</sup> Concerning this connection to forms of sociality, see essays 5–8 in *Faces of Intention*.

<sup>39</sup> This is implicit in the general strategy of Gricean “creature construction” in the philosophy of action (see above, n. 15). In unpublished work, Jennifer Morton pursues further implications of, as she puts it, the “varieties of agency.”

<sup>40</sup> I discuss the question, why (continue to) be a planning agent?, in “Intention, Belief, Practical, Theoretical,” where I emphasize the roles of planning agency in cross-temporally effective agency, in self-governance, and in our sociality. In these ways there are distinctively practical pressures in the direction of a kind of planning agency within which intentions have (or so we are now supposing) the cited constitutive aims. Note that this does not entail that we actually have a choice about whether to be planning agents. Nor does it rule out the possibility of cases in which things go better if one's planning system does not on that occasion function properly.

<sup>41</sup> See *The Possibility of Practical Reason*.

Now, one might think that we need to appeal to cognitivism to explain the special nature of the demand expressed in IR. In particular, we need to explain why violations of IR are a kind of *incoherence*.<sup>42</sup> The cognitivist will say that this is because the demand expressed in IR just is the demand for a kind of belief coherence. And the cognitivist sees this demand of coherence on belief as tied to the very nature of belief – where this includes the way it must be embedded in a system that tracks truth. But we can say something similar about intention without being cognitivists. We can say that the demand of coherence on intention (taken together with belief) is tied to the very nature of intention – where this includes the way it must be embedded in a planning system that tracks coordinated and effective control of action. And we can say this while acknowledging that not all agents are planning agents.<sup>43</sup>

#### 4 MISTAKES ABOUT ONE'S OWN INTENTIONS: HARMAN AND SETIYA

Let me turn now to the complexities I promised concerning my two objections to Harman's basic idea. Begin with the second objection<sup>44</sup>: I might falsely believe I intend a certain means intending which is, I know, needed for my intended end. In such a case I might satisfy BC but not IR. So IR is not grounded in BC. Or so I have averred. What might a cognitivist say in reply?

Well, a cognitivist might argue that such a false belief about one's own intention ineluctably violates a further basic theoretical demand on one's beliefs. So the theoretical demands on one's beliefs – where these theoretical demands include those necessarily violated by one's false belief about one's intention – really do, taken together, fully account for the rational force of IR. So my appeal to the apparent possibility of false belief about one's own intentions does not work as an objection to cognitivism about IR, so long as that cognitivism is allowed to appeal not only to BC but also to broad theoretical constraints against false belief about one's own intentions.<sup>45</sup>

<sup>42</sup> Setiya raises this issue in his comments on my "Intention, Belief, Practical, Theoretical."

<sup>43</sup> In this last sentence I am in disagreement with Setiya.

<sup>44</sup> I discuss the first objection in the context of my discussion, in the next section, of the views of John Broome.

<sup>45</sup> Wallace offers a version of this reply. I discuss it in "Intention, Belief, Practical, Theoretical." Here I focus on versions of this reply due to Gilbert Harman and to Kieran Setiya.

For this to work there must be a form of theoretical irrationality – and not just fallibility – whenever one misidentifies what one intends. What could that be?

One idea here – once suggested by Harman<sup>46</sup> – is that in falsely believing I intend M I falsely believe I *believe* M. And a false belief about what one believes involves a set of beliefs that are incoherent. After all (though here I go beyond Harman's explicit remarks), if I believe I believe p I should be willing to use p as a premise in my ordinary reasoning; but if I do not believe p I should not.<sup>47</sup> And the claim is that this incoherence within one's beliefs is ineluctably triggered by a false belief about what one intends.

But why say that in falsely believing I intend M I falsely believe I *believe* M? Well, as noted, cognitivist theories see intention as at least involving corresponding belief; and some see intention as itself a special kind of belief. So perhaps it will seem that, on such assumptions about the connection between intention and belief, when I falsely believe I intend M I do falsely believe I believe M. But, on reflection, we can see that this need not be so.

Suppose we say only that to intend A is, in part, to believe A, though it also involves other elements as well – perhaps intention involves both such a belief and, as well, a preference for A.<sup>48</sup> Well, then, I might falsely believe I intend M and yet still in fact believe M; it is just that, as a matter of fact, I do not satisfy the further condition for intending M – in the example, a preference for M. I believe I intend M but I do not intend M – though I do believe M I do not, unbeknownst to me, satisfy the further condition for intending M. So it is not true in this case that in falsely believing I intend M I falsely believe I believe M.

Now suppose we identify intending with believing. Well, we cannot plausibly say that intending to M is simply believing one will M. If we identify intending with believing it must be with a special kind of

<sup>46</sup> In a footnote to his 1980 APA comments on "Intention and Means-End Reasoning," Harman wrote:

In "Practical Reasoning" I assumed that to intend to do *B* is to have a certain sort of self-referential belief. So in this case [that is, the case of falsely believing one intends the necessary means] one believes one believes something which in fact one does not believe, and this might count as a kind of incoherence in one's beliefs.

I also discuss this suggestion of Harman in "Intention, Belief, Practical, Theoretical."

<sup>47</sup> Here I put to one side complexities involved in the possibility of acceptance in a context that is not belief.

<sup>48</sup> See Davis, "A Causal Theory of Intending."



believing. To intend M is, in this special way, to believe M. Harman's 1976 theory has this form: to intend M is to believe you will M by way of this very belief, where this belief is a conclusion of practical reasoning. Intentions are reflexive beliefs<sup>49</sup> that are, as well, the output of practical reasoning. Well, then, I might believe I intend M and, indeed, reflexively believe M, and yet not actually intend M. This could happen if my reflexive belief that M is not, in fact, the conclusion of practical reasoning. So I might falsely believe I intend M even though I do in fact reflexively believe M.

Indeed, Harman himself provides an example (one he attributes to Derek Parfit) in which one's reflexive belief is not a conclusion of practical reasoning, and so is not an intention. An insomniac might believe that he will stay awake because of his very belief that he will; yet he does not intend to stay awake. On Harman's 1976 theory, the insomniac's reflexive belief is not an intention because it is not a conclusion of practical reasoning.

Suppose now that the insomniac somehow mistakenly thinks his reflexive belief is the conclusion of practical reasoning – reasoning that is concerned, perhaps, with his desire to stay awake in order to write his paper for a conference on practical reason. He thereby mistakenly thinks he intends to stay awake. His belief that he intends to stay awake is false, though he does (reflexively) believe that he will stay awake. So, again, it is not true in such a case that in falsely believing one intends x one falsely believes one believes x.

I conclude that if misidentification of one's own intentions is always a form of theoretical irrationality, it will need to be for a reason different from that alluded to by Harman. And, indeed, Kieran Setiya has sketched a different argument for thinking there will always be a form of theoretical irrationality.<sup>50</sup>

Setiya begins with a theory of intention that is close to Harman's 1976 view: to intend to A is to have a self-referential belief that one will A by way of this very belief, where this belief is itself motivating.<sup>51</sup> Setiya thinks that if this is what intention is, then false belief about what one intends is

<sup>49</sup> That is, intentions are beliefs whose contents have the form: I will M in part because of this very belief.

<sup>50</sup> An initial formulation was in his comments on "Intention, Belief, Practical, Theoretical"; a more detailed version is in "Cognitivism about Instrumental Reason."

<sup>51</sup> In saying that the belief itself is motivating, Setiya may be going beyond what Harman claims in his 1976 paper, though Harman does sketch a similar view in his 1986 essay "Willing and Intending."

always a form of theoretical incoherence in which one violates “the epistemic ‘should’.”

In defense of this last claim, Setiya begins with an idea he derives from work of Richard Moran: “In the epistemic sense of ‘should’, one should never make an inference” – where Setiya is focusing on non-deductive inference “on the basis of evidence” – that is of a kind that “could never be both sound and ampliative.”<sup>52</sup> But, given Setiya’s theory of intention,

An inference to the conclusion that I intend to *x*, from *any* premise, will instantiate a pattern that cannot be both sound and ampliative. For suppose that I infer that I intend to *x* on the ground that *p*. If the conclusion is false, the inference is unsound. But if the conclusion is true, the self-reference of intention ensures that the inference is redundant. If I intend to *x*, I already believe that I am going to *x* because I so intend . . .

It follows . . . that there is something incoherent about the belief that I intend to *x*, unless it is constituted by the intention to *x*. It is an inherently defective belief.<sup>53</sup>

Is it really true, on Setiya’s theory of intention, that “If I intend to *x*, I already believe that I am going to *x* because I so intend”? If this were true it would be surprising, since it seems at least *possible* to have intentions one does not believe one has (which is not yet the possibility of mistakenly believing one has an intention one does not have). So if Setiya is accurately representing an implication of his theory we should be wary of that theory. The theory combines the idea that (a) intention is a kind of belief, with the idea that (b) intention involves reflexivity. Each of these ideas has a certain plausibility, though I myself would want to resist at least (a). And – it is important to note – these ideas are independent: one could accept (a) without (b), and vice versa.<sup>54</sup> But, if Setiya is accurately representing an implication of his theory, then what has happened is that when we put these two ideas together we get a surprising conclusion that one *never* intends something without believing that one so intends. By my lights, this should lead us to go back and re-examine the proposed merger of (a) and (b).

But perhaps Setiya is not accurately representing an implication of his theory. To be sure, on the theory, if I intend *x* then I believe I will *x* because of this very *belief*. But not all such reflexive beliefs – even true

<sup>52</sup> “Cognitivism About Instrumental Reason”: 670. Setiya is here extending talk of soundness to non-deductive inference.

<sup>53</sup> “Cognitivism about Instrumental Reason”: 671 (I have changed the action variable).

<sup>54</sup> John Searle, for example, accepts (b) but not (a). See *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983): chapter 3.

ones – are intentions; that is the lesson we learn from Parfit’s insomniac case. So perhaps, even on Setiya’s theory, I can in fact intend  $x$  without believing I *intend*  $x$ , though in intending  $x$  I reflexively believe I believe  $x$ , and this reflexive belief is, in fact, my intention. If so, I might newly come to believe, on the basis of evidence, that I do indeed intend  $x$  and do not merely reflexively believe I will  $x$ . If I do in fact intend  $x$  then the reasoning that leads me to this new belief about my intention could be both sound and ampliative. So it would not violate Setiya’s epistemic prohibition of inference that is of a kind that “could never be both *sound* and *ampliative*.” If I do not in fact intend  $x$  then I will have a false belief about what I intend even though the reasoning leading to that false belief is of a kind that *could* be both sound and ampliative (though, of course, it is not in this case).

My conclusion, then, is that Setiya has not convincingly shown that a false belief about one’s own intentions ineluctably violates an epistemic “should.” If we interpret his theory of intention in a way that does seem to show this, the theory is problematic; if we interpret the theory in a less problematic way, it does not show this.

#### 5 BROOME ON PRACTICAL REASONING

Both of my objections to Harman’s basic idea interact with recent work by John Broome on practical reasoning, work that leads Broome to a version of cognitivism about IR. Broome’s views are complex; but for present purposes we can understand the relevant aspects of his view in terms of the following claims<sup>55</sup>:

- (1) Intention is not belief.
- (2) It is not in general true that if you intend  $E$  you believe  $E$ .

This last is because (and here Broome and I agree)

- (3) You can sometimes intend  $E$  but fail to believe you intend  $E$ ; and in such cases you may well not believe  $E$ .
- (4) But, if you do believe you intend  $E$  then you will believe  $E$ .
- (5) And for your intention to  $E$  to enter into practical reasoning about means to  $E$ , you need to believe you intend  $E$ .
- (6) So when your intention to  $E$  enters into your practical reasoning about means to  $E$ , you believe  $E$ .

<sup>55</sup> In this discussion I focus on Broome’s “The Unity of Reasoning?” I quote from the manuscript of August 2008.

- (7) And it is this belief that E that provides the premise for your reasoning, namely: E.
- (8) If you also believe that E only if M, and if these beliefs do not change, BC requires that you believe M; and that is where your reasoning can lead.
- (9) But if in the “background” you believe that M will obtain only if you intend M, then if you do arrive at the belief that M this will normally be by way of intending M. In satisfying BC in this way you will satisfy IR.

This, then, is Broome’s broadly cognitivist picture of reasoning from intended ends to intended means. Broome wants to acknowledge, though, that it remains possible to intend E, believe that this requires both M and that you intend M, but falsely believe that you intend M. Broome grants that in such a case you fail to satisfy IR, though you may well satisfy BC. But, says Broome,

- (10) In such a case “your false belief blocks any reasoning that can bring you to satisfy” IR.<sup>56</sup> So,
- (11) Insofar as IR is a rational demand *that can be satisfied by reasoning* it is demand that derives from BC. Insofar as IR seems to impose demands that go beyond what is imposed by BC, these are not demands that can be satisfied by reasoning.

In this sense, it is BC that is fundamental for a theory of practical reasoning from ends to means.

Let me focus here on two ideas. The first is that my intention to E enters into my means-end reasoning by providing the believed premise that I will E. I will believe this premise since, for my intention to enter into my reasoning I need to believe I so intend; and if I believe I intend E then I believe E. The second idea is that a false belief that I intend M blocks the possibility of arriving at an actual intention to M by way of practical reasoning that begins with my intention to E. The first idea is that the relevant practical reasoning that can lead me to satisfy IR is theoretical reasoning concerning the contents of my associated beliefs. The second idea is that insofar as IR may seem – in cases of false belief about what one intends – to impose a rational demand that goes beyond what such theoretical reasoning can satisfy, it imposes a rational demand that no reasoning can satisfy. Taken together, these ideas amount to a kind of cognitivism about IR. And I think that both of my reasons for

<sup>56</sup> “The Unity of Reasoning?": msp. 17.

objecting to Harman's basic idea also suggest challenges to this form of cognitivism.

Broome supposes, in claim (5), that my intention to E can enter into my practical reasoning only if I believe I so intend. This seems delicate. On the one hand, it seems that we do not suppose that the belief that p can enter into theoretical reasoning only if one has the second-order belief that one believes p. It seems, for example, that a child might reason theoretically without having the concept of belief, and so without having a belief that she believes.<sup>57</sup> (Though perhaps, if her reasoning is *conscious*, she needs some higher-order thought that is in fact about her belief.<sup>58</sup>) So why should we think that intention is different from belief in this respect? On the other hand, a reasoning system needs to keep track of whether an attitude involved in reasoning is a belief or an intention. And one way to do this is to have second-order beliefs – or perhaps some other sort of higher-order thought – about which attitude is in fact involved.

Since I do not want to try to sort out this matter here, I will proceed by bracketing this complication and simply granting claim (5) for present purposes. Note though that Broome also needs claim (4), the claim that if you believe you intend E then you believe E. Claim (4) assumes that the only breakdowns between intending E and believing E occur when you fail to believe you intend E. But there is reason to be skeptical about this assumption: that is the point of the example of my commitment to stopping at the bookstore while being aware of my absent-mindedness, an example I offered as part of my first objection to Harman's basic idea.

Broome notes the possibility of such examples, and he acknowledges that we sometimes call the agent's attitude towards his action in such examples "intention." Broome says that in the case of such a non-confident intention one would normally express one's intention not by saying (or thinking) that one will so act (as one would, according to Broome, in the normal case of intention) but, rather, by saying (or thinking) that one intends so to act. And Broome acknowledges that such a non-confident intention in favor of E will not provide, as a premise for one's practical reasoning, the believed proposition E. Since Broome's story of the role of an intention to E in providing a premise for such practical reasoning is that it provides the believed proposition that E, he must grant – as he does – that his story of

<sup>57</sup> Example courtesy of John Broome, in correspondence.

<sup>58</sup> See David Rosenthal, "A Theory of Consciousness," in Ned Block, Owen Flanagan, and Guven Guzeldere, eds., *The Nature of Consciousness: Philosophical Debates* (Cambridge, MA: MIT Press, 1997): 729–753.

practical reasoning from ends to means does not apply to the case of a non-confident intention in favor of E: such non-confident intentions are “beyond the scope” of his account.<sup>59</sup>

But even if my intention to stop at the bookstore is non-confident it involves a distinctive kind of practical commitment that goes beyond mere preference: I have, in some practical sense, settled on stopping there. This returns us to the basic question of whether it is this practical step that newly engages a requirement like IR, or whether the relevant requirement of instrumental rationality is only engaged once one actually believes one will do as one intends. If we say the former – that even non-confident intention engages a requirement along the lines of IR – then we should worry that Broome’s account of practical reasoning from ends to means is inappropriately limited in scope. Broome identifies such practical reasoning with a form of theoretical reasoning that is commonly – though, it seems, not necessarily – associated with practical reasoning from ends to means. And that identification seems problematic.

What about Broome’s claim (10), the claim that in a case in which you knowingly intend E and know that this requires your intending M, but you falsely believe you intend M, “your false belief blocks any reasoning that can bring you to satisfy” IR? Broome’s view here is that

[t]here is simply no way you can reason your way to an actual intention, past your belief that you have an intention.<sup>60</sup>

This leads to Broome’s idea that insofar as IR goes beyond what is required by BC – since IR requires that you actually intend M in cases in which BC is satisfied by your false belief that you so intend – it does not require something you can achieve by reasoning. This is Broome’s strategy for responding to the issues raised by my second objection to Harman’s basic idea. What to say?

Well, consider reasoning in which one aims to *reconfirm* what one in fact already believes. I believe I locked the door when I left home earlier today. But I find myself engaged in reasoning that aims at reconfirmation:

Susan was there. She would have seen whether I locked it. If she had seen that I had failed to lock it she would have said something. She didn’t say something. So, I locked it.

Granted, in many such cases I suspend my belief that I locked it, once I embark on the reasoning. But I don’t see that this is necessary. A concern

<sup>59</sup> “The Unity of Reasoning?": msp. 13.

<sup>60</sup> “The Unity of Reasoning?": msp. 16.

with reconfirmation need not begin with doubt about what is to be reconfirmed; it might be focused, rather, on articulating the precise grounds for one's belief.

Suppose now that I didn't really believe I had locked the door, though I somehow believed I believed that. It seems that I could engage in the cited reasoning – reasoning that I mistakenly thought of as merely reconfirming my belief – but which, in fact, leads to my newly believing I locked the door. As a matter of fact, this reasoning finally fully convinces me that I locked the door. My earlier false belief that I already believed I locked the door need not block this.

Return now to the case of a false belief about what one intends. Suppose I intend E (and I know this), I know that E requires M by way of intending M, I do not in fact intend M, but I falsely believe I intend M. Though I satisfy BC (in the relevant respect) I am in violation of IR, though I do not know that I am. Suppose I aim to go through the practical reasoning in favor of intending M as a way of articulating the rational support for the intention in favor of M that I falsely believe I already have. (Perhaps, for example, M is my engaging in unpleasant physical therapy, and I seek to reconfirm the grounds for doing this unpleasant thing as a means to my intended end of recovery from my accident.) It seems I can go through the relevant means-end reasoning and thereby in fact be led newly to intend M, though by my own (false) lights my intention to M is not new and this reasoning merely reconfirms an intention I already have. My earlier false belief that I already intend M need not block this way of newly coming into conformity with IR by way of reasoning. So – though I grant that such cases are unusual – I do not think we should accept Broome's claims that “[t]here is simply no way you can reason your way to an actual intention, past your belief that you have an intention” and that “your false belief blocks any reasoning that can bring you to satisfy” IR. So we need a version of IR that goes beyond BC in order to understand how reasoning your way to an intention in favor of necessary means despite a prior false belief that you already so intend can newly bring you into conformity with a requirement of instrumental rationality. And that means that we should not be cognitivists about IR.

## 6 CONCLUSION

What conclusion should we draw from these reflections? Well, cognitivism seems beset by difficulties associated with the possibility of mistaken

beliefs about what one intends; and many versions of cognitivism do not sufficiently come to terms with the way in which non-confident intentions seem to engage a basic demand of instrumental rationality even in the absence of a belief in success. Further, it seems plausible that we do not need cognitivism to see the relevant norms as tied to a constitutive aim of intention, or thereby to explain their special nature. So, though there remain important unresolved problems that arise for both cognitivism and for the practical commitment view, the weight of these reflections seems so far to argue in favor of the practical commitment view as a better model of this fundamental aspect of reason.<sup>61</sup>

<sup>61</sup> As noted, I am including in these reflections the arguments in both this present essay and in “Intention, Belief, Practical, Theoretical.” Let me note, however, an issue I have not tried to address in either of these essays. One might conjecture that cognitivism about IR is needed in order to avoid the kind of unacceptable bootstrapping described in “Intention and Means-End Reasoning” and in *Intention, Plans, and Practical Reason*: 24–27. The idea, in a nutshell, is that it is only by seeing IR as, at bottom, a theoretical requirement on belief rather than a requirement of practical rationality on intention, that we can avoid seeing intending *E* as always providing a kind of practical reason for *M* that constitutes unacceptable bootstrapping. Setiya develops an idea along these lines in “Cognitivism about Instrumental Reason.” I hope to address this issue on another occasion.