Postscript to Chapter Five

The instrumental principle, as I interpret it, instructs us to take the means that we believe to be necessary relative to our ends. My treatment of this principle attempts to do justice to two dimensions of instrumental rationality. First, the instrumental principle applies even in cases in which we are acting in pursuit of ends that we do not ourselves fully endorse. Second, it applies in a way that involves a kind of rational pressure to revise our attitudes. We feel and respond to this pressure when we adopt necessary means to our ends, or give up our ends upon realizing that we are not willing to take the means that are necessary for their attainment. My conviction is that we can do justice to these twin features of instrumental rationality only by tracing the instrumental principle to requirements of theoretical rationality, concerning the coherence of our beliefs (including the beliefs that are bound up with the intention to pursue a given end). These are requirements whose force is familiar to us as believers, insofar as we feel and respond to a rational pressure to revise our beliefs when we recognize them to be inconsistent. Furthermore, the theoretical source of the rational pressure to revise inconsistent beliefs renders it applicable to situations in which we are acting in pursuit of ends that we do not fully endorse.

The defense of this approach in 'Normativity, Commitment, and Instrumental Reasons' leaves many questions unanswered. In this postscript I wish to address a few of the objections that have recently been brought against my approach; the aim is not to offer an exhaustive treatment of the issues raised by these objections, but rather to sketch in outline the resources available within the approach for dealing with some hard cases for the theory of instrumental rationality. I shall focus on three sets of issues.

1. Revision of Belief and of Intention.

The agent who fails to intend the means they believe necessary relative to an end they intend to achieve is subject to inconsistent beliefs—or so I maintain. But what is wrong with that? The set of beliefs we hold may well turn out to contain many latent inconsistencies, but this fact alone generates no particular rational pressure in the direction of making specific revisions. That our beliefs are inconsistent entails that not everything we believe can be true, but this in itself poses no special problem so long as we remain in the dark about where in our network of beliefs specific falsehoods may lie.

Of course, in the case of instrumental irrationality we can, it would seem, locate the falsehood in our beliefs. We believe the following three propositions, and know that they cannot together be true:

(1) It is possible that I do X.

(2) It is possible that I do X only if I also intend to do Y.

(3) I do not intend to do Y.

Joseph Raz has objected that the appeal to falsehood in this complex of beliefs fails to illuminate what is undesirable about the failure to take necessary means to one's chosen ends. In particular, he suggests, we need an explanation of the undesirability of this false belief 'that explains why the so-called principle of instrumental rationality is one of the standards that determine well-functioning deliberative processes.'[1] But the appeal to false beliefs does not illuminate the role of the instrumental principle as a norm of deliberative good functioning. Thus someone who intends to perform an action that it is in fact impossible for them to perform might end up wasting limited resources of time and energy in an essentially fruitless pursuit. But if such a person happens to be instrumentally

irrational, they will spare themselves this fate. They may be subject to a false belief, but in the present context this actually seems to be an advantage, something that is conducive to their overall good functioning as agents.

Raz himself explains instrumental rationality by appeal to standards of effective agency. 'If you are prone to instrumental irrationality, you are less likely to achieve your ends, whatever they are.'[2] A certain practiced knack for taking the necessary means to our ends is an important executive capacity, one that any successful agent will need to acquire, and our sense that we are irrational or defective when we fail to take the necessary means to our ends derives from the general role of this executive capacity in relation to deliberative good functioning. I think Raz is correct to emphasize the role of instrumental rationality as part of our good functioning as agents.[3] But this point does not suffice to account for the full normative significance of the instrumental principle. On Raz's account, that principle serves to specify a standard of excellence in agency, defining what we might think of as a kind of all-purpose executive virtue. But this turns the instrumental principle into an evaluative ideal, one that is not sufficiently anchored in the deliberative perspective of the agent. Someone who believes that a given means is necessary relative to an end that they intend to achieve will experience an immediate rational pressure to revise their attitudes. This pressure cannot derive from the fact that they will fail to exhibit an executive virtue if they do not revise their attitudes. For one thing, the agent may not be aware that the principle of instrumental rationality defines a standard of executive virtue. For another, even if they think in these terms, their doing so need not translate into any particular rational requirement to which they would feel pressure to respond. (The akratic agent in particular would probably not care very much about

the fact that they are failing to exhibit excellence in agency, given that they are not responsive to the acknowledged badness of their own immediate ends.)

My claim is that we can make sense of a kind of rational pressure to which even akratic agents are susceptible by appealing to the apparent conflict in the beliefs to which agents are subject when they fail to adopt means that they believe to be necessary relative to their intended ends. But Raz apparently doubts this. He notes that 'nothing follows about what we ought to do or believe and when we should suspend belief from the mere knowledge that a set of beliefs contains a contradiction.'[4] The fact that some subset of one's beliefs is inconsistent is no reason in itself to change the beliefs in the inconsistent set. This may be true as far as it goes, but it underestimates the extent to which awareness of localized inconsistency in belief can generate rational pressures on the believer. Even if I do not immediately know which of my inconsistent beliefs to revise, the fact that they are inconsistent is at least a pro tanto reason to reassess the credentials of the items in the inconsistent set. Furthermore, in the particular case at issue there is a special feature of at least one of these problematic beliefs that simplifies the process of revision in response to acknowledged local inconsistency. Thus belief (3) in the above set is a belief about my own intentions; the truth to which it is answerable is a matter that is directly subject to my volitional control. In this unusual situation, I can restore coherence to the problematic set of beliefs by adjusting my intentions regarding the necessary means, and thereby giving myself grounds for revising belief (3). If I nevertheless refuse to adjust my attitudes in this way, and retain my conviction that the means is really necessary, then I have no option but to revise belief (1), which in turn precludes me from continuing to intend the original end.

It might be thought that forming an intention to take the means in response to this kind of inconsistency fails to do justice to the nature of belief, which is answerable to independent facts of the matter about the way things are. I should revise belief (3), that I do not intend to take the means, because I have now formed such an intention; but then it seems I must have had some independent reason for adjusting my intentions in this way before I can revise my beliefs in the matter.[5] I agree that if belief (3) is to be revised, the only way to achieve this compatibly with the nature of belief as answerable to the truth is to revise one's intention first, by deciding or resolving to take the means. It does not follow, however, that there needs to be an independent or prior ground for forming this intention (independent, that is, from considerations about the consistency of one's beliefs). In the case at issue, where the belief in question is about a matter that is directly under my own control, the rational pressure that leads me to form an intention might well be the fact that I will thereby bring about coherence in a set of attitudes that includes a belief about my own intentions.

2. Questions about Possibility.

The rational pressure toward revision of intentions in accordance with the instrumental principle operates not merely via beliefs, but via beliefs about the possibility of doing what one intends. Possibility is of course a vexed topic in its own right, and questions may be raised about how the notion is to be interpreted in the context of the debate about the instrumental principle.

Suppose that one intends to pass an examination, and believes that that in turn will happen only if one resolves to study for the examination. Suppose further that one does not intend to study, and that one knows this fact about one's intentions. In this situation, we have specific instances of belief-types (2) and (3), which

together are supposed to be incompatible with a belief of type (1), to the effect that it is possible that one pass the examination. But are (2) and (3) really incompatible with (1)? There is, it would seem, a perfectly respectable sense in which it remains possible that one should pass the examination, even if one does not intend the necessary means. So long as it is possible for the agent to change their intentions regarding the means, it is possible for them to achieve the end; their doing so remains an option for choice, and something they might reasonably be held to account for failing to do. (It would hardly be an excuse for someone to plead that they were unable to pass the examination, because they failed to intend to take the necessary means.) In the generic sense of possibility as rational capacity, it is possible for us to do something even if we intend neither to do it nor to take the means that doing it requires, so long as it remains in our power to alter the intentions in question.

This generic sense of rational capacity cannot be the one that is at issue in my argument for the instrumental principle, figuring in the content of beliefs (1) and (2).[6] This raises two questions: Can we give a principled explication of an alternative notion of possibility that might figure in the content of these beliefs? And does it remain plausible to suppose that intention requires a belief in the possibility of doing what one intends, when (1) is interpreted in this way?

The conception of possibility that intuitively seems relevant to the argument of my paper is the conception of what is possible, given facts about my intentions. Something is possible in this sense if its occurrence is compatible with the rest of what we know about the world, holding fixed certain natural assumptions about the course of events, including the assumption that our own intentions will be executed.[7] Thus in thinking about the examination I assume, for instance, that I will not be struck by a sudden brainstorm in the

examination room, and that the questions that are asked will not address exclusively the small segment of the syllabus that I have already mastered. For practical purposes these scenarios seem too remote and unlikely to make it sensible for me to waste any time taking them into account in my deliberations about the future. Similarly, I take it as provisionally fixed that I will execute the intentions I have formed when I am engaged in deliberative reflection. The basic function of intentions is to resolve practical questions about my own agency, taking certain questions off the immediate deliberative agenda, so that I can focus on other issues that need to be dealt with. It would defeat the purpose of deliberation, and indeed undermine our basic capacity for effective action, to form the intention to do X, and then proceed in our practical thinking as if it was an open question whether we were going to X.[8]

The horizon of deliberative reflection is in this way shaped by a variety of assumptions about what will happen that have their rationale in our practical point of view as agents. We can abstract from that point of view, and think about ourselves in a way that is detached from the practical problems to which deliberation is a systematic response. When we do so, we might conclude that it is possible for us to do something we have already decided not to do, insofar as we retain the basic capacity to revise our intentions. We might equally concede, in this theoretical vein, that it is possible that we will be struck by a brainstorm in the examination room even if we do not study in advance. But these assumptions are not ones that it is reasonable to make within the deliberative context of agency. From this practical point of view we assume (defeasibly) that the intentions we have formed will be executed, accepting that it is possible for us to do something only if our doing it is compatible with the assumption that our intentions will in fact be translated

into action. The beliefs whose consistency is relevant to instrumental reasoning, on my account, include crucially the attitudes of acceptance within the context of practical deliberation that we have toward propositions such as (1) – (3).[9]

This explains, perhaps, why I am entitled to take it as fixed that it is not possible for me to pass the examination, when I have formed the definite intention not to study for it, and believe that I can pass only if I intend to study. But there is a different case to be considered: what if I merely fail to form an intention to study, while continuing to believe that doing so is necessary relative to my goal of passing the exam? Insofar as I have not yet formed any definite intention on the matter one way or another, I ought, even within the context of deliberative reflection, to consider it still an open matter whether I study for the examination or not. Cases of this kind bring out an indeterminacy in the content of belief (2) in the inconsistent set. This is the belief that it is possible that I do X only if I intend to do Y. But we need to ask when the intention to do Y needs to be formed. It might be, for instance, that the possibility of passing the examination requires only that I intend by tomorrow to devote my energy to studying. Or it might be possible to pass the examination only if I now form the intention to study for it. In the former case, the counterparts of (1) and (2) would be compatible with the belief that I do not currently have an intention to study for the exam; but they would not be compatible with that belief if the possibility claim in (2) is understood in the latter way, as requiring that I now intend to study. It follows that rational pressure to revise one's intentions will in some cases be generated by the trio of attitudes only when what is required to achieve the intended end is that one <u>now</u> intend the means. Prior to this point in time, the requirements of coherence of attitude may not yield any concrete pressure to make changes in one's intentions,

apart from ruling out a definite intention not to take the later means (so long as the other two attitudes remain in place). This is, it seems, as it should be—the core requirement expressed by the instrumental principle does not operate at a temporal distance, but only when we believe that steps must now be taken if we are going to realize our intended ends.

Finally, we must consider whether it remains plausible to suppose that intention requires the belief that it is possible to do what one intends, when the notion of possibility is interpreted as constrained by one's other intentions. Against this, it has been suggested that to understand intention as carrying this kind of belief in its train would rule out cases of irrationality, where we intend an end and fail to adopt the means that rationality requires of us.[10] But this objection misfires. Irrational intentions of the relevant kind are possible, on the account I favor, precisely to the extent that it is possible for the agent to hold inconsistent beliefs about the possibility (in the constrained sense) of achieving what they intend. Inconsistency of this kind is in fact rather difficult to tolerate, once it has been brought to our attention; the rational pressure we feel to revise beliefs we know to be inconsistent, in cases in which the beliefs are about matters directly under our immediate control, is ordinarily too great to resist. But this, again, accords with our ordinary understanding and practice. We are inclined to doubt whether someone really does intend the end if they don't choose the means they take to be necessary, under circumstances in which they are focused and aware and thinking clearly about their situation. It was for this reason that I suggested originally that the instrumental principle functions more like a constraint on interpretation than do other practical principles. Violations of the instrumental principle, when they occur, seem typically to involve the kind of cognitive errors in reasoning that make it possible for

us to overlook easily-rectifiable inconsistencies in our own beliefs, errors such as inattention and failure to put things together. This is what we should expect if the pressure to revise our intentions in accordance with the instrumental principle has its source, as I have maintained, in conflicts in our beliefs.

3. False and Incomplete Beliefs about One's Intentions.

The cognitivist approach to the instrumental principle faces a different set of potential problems in the possibility that our intentions might diverge from our beliefs about our intentions. Any of the following discrepancies between intentions and beliefs about intentions seem, in principle, conceivable: I might intend to X, but not believe that I intend to X; I might intend to X, but believe that I intend to do Z rather than X; I might have no intention to do X, but believe that I have such an intention; I might have no intention to X, but lack the belief that I have no intention to X. To the extent discrepancies of these kinds are possible, it would seem possible that one might exhibit failures in instrumental reasoning, without being subject to the kind of inconsistency in beliefs to which I have traced the instrumental principle.

In 'Normativity, Commitment, and Instrumental Reason' I argued that possibilities of these kinds could safely be ignored, since there are good grounds for thinking that it would independently be irrational to have mistaken beliefs about one's (consciously accessible) intentions, in the deliberative contexts to which the instrumental principle applies. This assumption might be challenged, however. Thus John Brunero writes: 'While it is plausible to claim that an agent who intends not to Y will believe that he does not intend to Y, it is not plausible to claim that someone who fails to intend to Y has the belief that he does not intend to Y. Alice may fail to intend to study and fail to notice this; it just slips her

mind. But if she does not hold the belief that she does not intend to study, then there is no way on Wallace's account to convict her of instrumental irrationality.'[11]

It is of course correct that one can fail to intend to Y without believing that one does not have this intention, and that this is a fact about one's intentions that is easier to overlook than (say) the fact that one intends not to Y. Furthermore, there is in general nothing necessarily irrational about lacking the belief that one does not have an arbitrary intention that it would at the time have been possible for one to adopt. But the deliberative contexts to which the instrumental principle is relevant have special features, which generate rational pressure to form accurate beliefs of this kind about one's own lack of certain intentions. These are contexts in which one is reflecting about how to proceed in executing one's intention to do X, and one has arrived at the conclusion that it is possible that one will X only if one forms the intention to do Y. I contend that a rational agent who is minimally self-aware (as I put it) will, in this special context, realize that they have not formed the very intention whose necessity relative to their larger end they have themselves explicitly affirmed. Having arrived at the conclusion that it will be possible to realize one's end only if one intends to Y, it would be a gross lapse in deliberation to overlook the fact that one has not formed that intention.

What explains the plausible suggestion that a lapse of this kind, in the specific deliberative context at issue, would amount to a form of irrationality? We might want to say here that a tendency to form accurate beliefs about whether one has formed an intention whose necessity to one's other purposes one has explicitly affirmed is a kind of executive virtue, to be included among the traits and capacities that make us, in general, effective in the pursuit of our goals. Construed in these terms, the appeal to rationality functions

to explain why it is safe to assume that minimally competent agents will be aware of their own lack of the relevant instrumental intentions in the context of the kind of deliberation that is at issue.

Alternatively, we might say that there is reason for us to scrutinize our intentions to this degree, insofar as doing so is a strategy that enhances our ability to realize the broader aims that are given with our nature as deliberating agents.[12] This answer might appear circular in the context of an account of the instrumental principle, presupposing that we have reason to take effective means to our ends, where the point was to explain why this is the case. But there is in fact no circularity here. In 'Normativity, Commitment, and Instrumental Reason' I distinguished between the core instrumental principle that it was my aim to offer an account of, and more wide-ranging norms of means-end coherence. The idea that we have reason to take steps that facilitate the pursuit of our broader aims is an argument that appeals to norms of the latter kind. Reliance on such norms might be an embarrassment for a theory that attempted to account for all of instrumental rationality in cognitive terms; but that was not my aim, nor do I believe it plausible to suppose that the whole field of instrumental rationality can be explained in terms of the requirements of coherence in belief that operate where the core requirement to take necessary means to one's ends gets a grip.

We have, then, two ways in which the rationality of believing that one does not intend to Y might be accounted for, in contexts in which one has formed the intention to X and affirmed that it is possible for one to X only if one intends to Y. At this point, however, a different and final worry might be raised. We want to say, it seems, that a failure to take the necessary means to our chosen ends is, in and of itself, a ground for rational criticism. The argument canvassed above, however, fails to do justice to this

important idea. That argument establishes that agents who form beliefs (1) and (2), and who also fail to intend the necessary means Y, are irrational if they do not also form belief (3), to the effect that they do not intend to Y. But the argument does nothing to show that, under these circumstances, the failure either to intend the means Y, or to revise the original intention to do X, is itself a form of irrationality. Rational pressure to revise one's intentions in a case of this sort emerges only after one has exhibited deliberative rationality of the prior kind, by registering the fact that one has failed to form the intention to take the necessary means. To the extent this is the case, it appears we still haven't accounted for the independent force of the instrumental principle.

To this I would respond that it is not at all clear to me that the instrumental principle really has the kind of independent rational force attributed to it in the objection. Recall that we are attempting to explain the rational pressure to revise one's attitudes in accordance with the instrumental principle that is generated by the mere fact that one has adopted the intention to achieve a given end, independently of whether the end really would be good to pursue in the circumstances, or is believed by the agent to be good. When we abstract from normative considerations about the justification for pursuing a given end, and consider merely the difference it makes that the agent has formed the intention so to act, it is no longer obvious at all that the failure to adopt the means that are believed necessary relative to the end is a form of irrationality. We may grant that a reason to pursue an end is eo ipso a reason to take other steps that would facilitate such a pursuit, so that the reason for the end transfers its rational force across the means/end relation.[13] But in the case at issue we are not assuming that there really is a good reason for the agent's pursuit of the end in the first place. Perhaps, under these circumstances, there is no rational

pressure to revise one's intentions that operates independently of one's beliefs about (inter alia) one's own intentions.

This response could be supplemented by borrowing on Raz's executive virtue account of instrumental rationality. I have already suggested that a tendency to form accurate beliefs about the state of one's own intentions may be a rational excellence, something that contributes to one's effectiveness as an agent engaged in practical deliberation. Perhaps it is a further and distinct excellence to be disposed to revise one's intentions in accordance with the instrumental principle, as Raz maintains. This concession to Raz would not supplant the cognitivist account I have offered of the core requirement; as I argued above, the appeal to deliberative virtue fails to explain the internal rational pressure that deliberating agents register and respond to when they revise their intentions in accordance with the instrumental principle. This pressure, it seems, is present only to the extent agents are competent enough to have taken notice of the fact that they have not yet formed the intentions that they acknowledge to be necessary for the realization of their ends. But the idea that a tendency to revise our intentions in accordance with the instrumental principle is part of what it is to be an effective agent might explain the residual thought that there is something independently defective about an agent who does not make such revisions—independently, that is, of whether or not they have formed accurate beliefs about the state of their own intentions as they deliberate.[14]

[1] See Joseph Raz, 'The Myth of Instrumental Rationality', Journal of Ethics and Social Philosophy, www.jesp.org, (April 2005), vol. 1, no. 1, p. 17.

[2] Raz, 'The Myth of Instrumental Rationality', p. 17.

[3] A similar view is defended by Michael Bratman, in Intention, Plans, and Practical Reason (Cambridge, Mass.: Harvard University Press, 1987), at e.g. p. 51—a source upon which Raz draws.

[4] Raz, 'The Myth of Instrumental Rationality', p. 20.

[5] See John Brunero, 'Two Approaches to Instrumental Rationality and Belief Consistency', Journal of Ethics and Social Philosophy, www.jesp.org, (April 2005), vol. 1, no. 1, p. 4.

[6] Compare Brunero, 'Two Approaches to Instrumental Rationality and Belief Consistency', pp. 4-6.

[7] I say 'executed' here, to leave open the possibility that one will not succeed in achieving the end that one intends to realize. We hold it fixed that we will at least make a concerted effort to achieve what we are intending, and in that sense we assume that our intentions will be executed.

[8] See Bratman, Intention, Plans, and Practical Reason, for an extended defense of this conception of the role of intention.

[9] On the notion of 'acceptance in a context', and its relevance to the practical perspective of agency, see Michael Bratman, 'Practical Reasoning and Acceptance in a Context', as reprinted in Bratman, Faces of Intention (Cambridge, England: Cambridge University Press, 1999), pp. 15-34.

[10] Brunero, 'Two Approaches to Instrumental Rationality and Belief Consistency', pp. 5-6.

[11] Brunero, 'Two Approaches to Instrumental Rationality and Belief Consistency', p. 6.

[12] Michael Bratman makes this suggestion, in 'Intention, Belief, Practical, Theoretical', unpublished manuscript.

[13] Compare Raz's discussion of what he calls facilitative reasons, in 'The Myth of Instrumental Rationality', sec. 1.

[14] It might be doubted, however, whether we can really make sense of this tendency as one that is genuinely independent of the tendency to form accurate beliefs about one's own intentions. In practice, someone will probably exhibit a knack for revising their intentions in accordance with the instrumental principle only to the extent they have a knack for registering when they lack the intentions they believe necessary relative to their ends.