

## INTRODUCTION: THE PLANNING FRAMEWORK

Michael E. Bratman Forthcoming in *Planning, Time, and Self-Governance: Essays in Practical Rationality* (OUP) subject to adjustment Please quote from published version.

08-23-17

We – mature human agents in a broadly modern world – are planning agents. Our practical thinking is shaped in characteristic ways by our future-directed plans. This is a key to the striking forms of cross-temporal organization that are central to our human lives. It is also, I think, a key to important forms of our sociality and our self-governance. Somewhat stable prior partial plans normally provide a background framework within which much of our practical thinking – including the weighing of reasons – proceeds. This framework-providing role of these prior partial plans involves distinctive rationality norms – norms of consistency,<sup>1</sup> coherence, and stability over time. And a fruitful way to understand intention is to see intentions as plan-states in this plan-infused, normatively guided psychic economy. In this way, we can develop a naturalistically plausible and modest theory of the will.

Or so I have argued in a series of works, beginning in the mid-1980s.<sup>2</sup> As this work has progressed, however, it has become clear that there are unresolved and difficult issues about practical rationality at the foundations, issues that require further reflection. This collection of essays is my effort, so far, to respond further to these foundational issues.

<1> 1.

Let me set the stage. I developed some of these ideas about our planning agency in my 1987 book, *Intention, Plans, and Practical Reason*. I then explored two further ideas. The first was that this planning framework provides substantial resources for a rich model of our *shared* acting and thinking, a model in the space between the work of, among others, Margaret Gilbert and John Searle. This led to a series of papers beginning in the early 1990s, four of which were included in my 1999 collection, *Faces of Intention*; and this then led to my 2014 book *Shared*

*Agency: A Planning Theory of Acting Together*. The second idea was that this planning framework provides substantial resources for a rich model of our self-governance, one in the space between the work of, among others, Harry Frankfurt and Gary Watson. This led to a series of papers – many of which are in my 2007 collection, *Structures of Agency* -- in defense of a plan-based model of self-governance. These twin developments added further support for the idea that our capacity for planning agency is a core capacity, one that supports a wide range of important practical capacities and is in that sense fecund. (See essay 5.)

But unrest was brewing. Why are the planning norms to which I was appealing, norms of *rationality* with a distinctive normative significance or force? In my 1987 book I had proposed a two-tier pragmatic justification for these norms, one that appealed to the ways in which guidance by these norms in general supports, both instrumentally and constitutively, much of what we care about, especially given our cognitive and epistemic limits. And I also tried to block the worry, traceable to J.J.C. Smart's forceful concerns about rule-worship, that a direct inference from pragmatically supported general strategies of thought to a judgment about the particular case is fraught.<sup>3</sup> But as I thought more about this Smart-inspired concern, my confidence waned.<sup>4</sup> And in his criticism of my argument against what I had called the "simple view," Hugh McCann challenged the very idea that the relevant norm of plan consistency is a norm of rationality rather than just a rule of thumb for tracking what really mattered.<sup>5</sup> Related challenges later appeared in work of Joseph Raz and Niko Kolodny who both think that the idea that the cited planning norms are norms of rationality is a "myth".<sup>6</sup>

During the same period, others – including J. David Velleman, Kieran Setiya, and to some extent R. Jay Wallace -- developed the idea that these norms were the reflection of more basic norms of theoretical rationality in favor of consistency and coherence of belief. And some

such – as I called it -- cognitivism could already be discerned in Gilbert Harman’s ground-breaking essay, “Practical Reasoning”.<sup>7</sup>

I myself was persuaded neither by cognitivism (see essays 2 and 3) nor by the myth-theoretic challenge. (See essay 4.) But I realized that I needed to go beyond the (albeit, qualified) two-tier account in my 1987 book. So beginning in 2009 I published a series of papers in which I tried to work toward a more adequate view.

One idea that was not available to me was that these norms are constitutive of agency -- that it is a necessary feature of agency (or, anyway, intentional agency) that it involves guidance by these norms.<sup>8</sup> This was in part because I wanted to situate the planning theory within a theoretical strategy that Paul Grice had called “creature construction”.<sup>9</sup> The idea is to build up increasingly rich models of agency and to see planning agency as a specific form of agency, one that appears in later stages of this exercise in creature construction. Given this multiplicity of agency, it is not plausible that these distinctive norms of planning agency are constitutive of agency quite generally.

That said, guidance by these norms is, on the theory, partly constitutive of, in particular, planning agency. But given that you can be an agent without being a planning agent, it is not clear exactly how this – as we might call it, planning-agency constitutivism – helps us defend the claim that these are norms of practical rationality for such an agent.<sup>10</sup>

So I sought a path between a myth theory and cognitivism, but a path that did not depend on agency-constitutivism. I continued to think that it was important that the general guidance by these norms helps shape our lives in profound ways that are central to much of what we care about. But I concluded that we needed to supplement these two-tier pragmatic ideas. Such a supplement should draw on the common view that rationality is at least in part a matter of

coherence.<sup>11</sup> But what I learned from McCann, Raz, and Kolodny was that a mere appeal to coherence was open to the challenge that we were grounding rationality simply in a kind of mental “tidiness”.<sup>12</sup> We need to explain why the relevant forms of coherence matter in a way that responds to the concerns of the myth theorist, and we need to understand the philosophical underpinnings and implications of that explanation.

While I was trying to figure this out I was also developing the approach to a planning agent’s self-governance articulated in the essays in my 2007 book.<sup>13</sup> This led me to the conjecture that we can forge a path of the right sort between a myth theory and cognitivism in part by studying the ways in which basic norms of plan rationality reflect conditions of a planning agent’s self-governance, both at a time and over time. We begin by explaining what is involved in central cases of a planning agent’s self-governance. We then see the demands imposed by planning norms as demands to satisfy certain of these conditions of self-governance. The forms of coherence that are required by these planning norms will be forms of coherence that are among these conditions of a planning agent’s self-governance.

The idea, then, is to treat central conditions of a planning agent’s self-governance as a starting point and then see relevant norms as demands in favor of satisfying those conditions. In this sense these norms track relevant conditions of self-governance. The (perhaps implicit) acceptance of these norms, once in place, will then support conformity with these conditions of self-governance. So these norms track conditions of self-governance, and their acceptance tends to lead to conformity to those conditions.<sup>14</sup>

These essays constitute an interim report of progress, so far, with this line of reflection. The essays reveal a progression of ideas, with adjustments along the way. While there remain many unanswered questions,<sup>15</sup> what emerges is, I think, a plausible strategy for steering a path in

our understanding of plan rationality between a myth theory, cognitivism, and agency-constitutivism. This path continues to acknowledge the significance of two-tier pragmatic considerations but supplements those considerations in an important way. My primary aim in this Introduction is to pull together, in broad strokes, some of the main ideas about this path to which I have been led in writing these essays.<sup>16</sup> In doing this I leave for those essays many of the details, complexities, qualifications, references, and credits -- though I will provide pointers to those essays along the way. And toward the end I will go somewhat beyond these essays, though in a way that is tied to ideas in these essays.

## <1> 2.

As I see it, our main concern is with the reflection of a planning agent on her basic forms of practical thinking, forms of practical thinking guided by distinctive norms of (as I claim) plan rationality. How might she make the most sense of these norms, and would they thereby be reflectively stable? (Essays 2, 9, 10) These queries are motivated both by our interest in the normative force of these norms and by our related interest in the explanatory significance of structures of planning agency. If these norms would not be stable under reflection then it would be to some extent less plausible to say – as the planning theory does say – that these forms of practical thinking are central to the explanation and understanding of our human action.

A planning agent's reflection would take due account of the pragmatic virtues, especially given her epistemic and cognitive limits, of these general forms of thinking. But it would also be sensitive to the Smart-inspired thought that a direct inference from the pragmatic benefits of these general forms of thinking to a conclusion about the particular case is fraught. (Essays 8, 9, 10) Our reflective planning agent will want to know how to supplement such a two-tier

pragmatic account (which is not to say that she will explicitly appeal to such a supplement in each case of her actual practical thinking).

It is here that (beginning with Essay 4) I try to develop a *strategy of self-governance*. This involves two ideas. The first is the *tracking thesis*: basic planning norms track certain of the conditions of a planning agent's self-governance, conditions that are articulated within our model of that self-governance. The second idea is the *reflective significance thesis*: the tracking thesis helps provide the supplemental support for these planning norms that is sought by the reflective planning agent in justifying her application of these norms to the particular case.

Concerning the tracking thesis, begin with a planning agent's self-governance at a time (or, anyway, during a small temporal interval). Here my proposal is a merger of Frankfurt-inspired ideas with elements from the planning theory.<sup>17</sup> Central to such self-governance is guidance by the agent's relevant, suitably coherent practical standpoint. Given the roles of planning in a planning agent's temporally extended agency, this standpoint will, for such a planning agent, be substantially plan-infused. So the plan consistency (including plan-belief consistency) and means-end coherence required by basic planning norms will be needed for the agent's standpoint to be sufficiently unified to play this Frankfurtian role in her synchronic self-governance. And this supports the tracking thesis concerning these synchronic norms.

This brings us to the reflective significance thesis with respect to these synchronic norms. Here I envisage a four-part answer.<sup>18</sup> (1) We can now see why the coherence that is tracked by these norms is, plausibly, a coherence worth wanting: it is not merely mental tidiness but is, rather, partly constitutive of a planning agent's self-governance – where I take it that such self-governance is itself worth wanting. (2) We have thereby identified an important commonality across these norms. Indeed, as we will see below, we have laid the foundation for articulating a

common element that underlies not only these synchronic norms but also a plausible diachronic norm. (3) The tracking thesis supports the claim that if (a) one has a normative reason in favor of one's synchronic self-governance, and (b) one has in the particular case the capacity for such self-governance, then (c) this normative reason transmits to a distinctive reason of synchronic self-governance to conform to these synchronic norms in that particular case. (See essay 4, where I also respond to related concerns about inappropriate bootstrapping of normative reasons.<sup>19</sup>) And (4): our reflective planning agent will indeed have a normative reason in favor of her self-governance.<sup>20</sup>

Given the tracking thesis, and given (1)-(4), the reflective planning agent will be in a position to supplement her appeal to a two-tier pragmatic rationale for her synchronic planning norms in a way that suitably supports her application of these norms to the particular case. But there remain two issues. First, we are so far without a defense of (4), the claim about a normative reason for self-governance. And second, we have so far not explored whether – and if so, how -- this approach could be extended to the diachronic case.

I turn now to this second issue, about the diachronic case. What I say will then be of use when we return later to the first issue.

### <1> 3.

So what should we say about the tracking thesis and diachronic plan rationality? Here we will need to reflect on a planning agent's self-governance not only at a time, but also *over* time. But first let me note a methodological issue. When we were considering the tracking thesis concerning synchronic norms of plan consistency and coherence, we took as given the basic outlines of those norms and tried to see if, so understood, they did indeed track conditions of a planning agent's synchronic self-governance. In contrast, when we turn to diachronic plan

rationality it is less clear, at the start, exactly what the relevant diachronic norm(s) is/are. So in exploring the tracking thesis about diachronic plan rationality we will need to assess an overall package that includes both a model of a planning agent's diachronic self-governance and a view of the associated diachronic norm (or norms).

That said, how should we model a planning agent's self-governance over time? (Essays 6, 10, 11) An initial idea is that a planning agent's self-governance over time involves her self-governance at (during) relevant times along the way, together with relevant inter-connections across those forms of synchronic self-governance. What inter-connections? Here I have been led to two complementary ideas. The first is that these are the (perhaps implicit) plan-infused inter-connections that are characteristic of planned temporally extended activity. These include characteristic forms of continuity of intention, cross-references between intentions at different times, intended and actual mesh between sub-plans at different times, and inter-dependence between relevant intentions at times along the way. The second idea is that these plan-theoretic *intra*-personal inter-connections are analogues of the (perhaps implicit) *inter*-personal intention inter-connections characteristic of shared intention and shared intentional activity – where these inter-personal interconnections are those articulated in my planning theory of acting and thinking together. (Essays 5, 11.) These analogies between planned temporally extended agency and shared intentional agency support the useful metaphor that in diachronic self-governance one is, as it were, acting “together” with oneself over time. (Essays 10, 11)

So the initial proposal is that a planning agent's self-governance over time involves her self-governance at times along the way together with cross-temporal plan-infused *intra*-personal inter-connections that are analogues of the plan-infused *inter*-personal interconnections characteristic of shared intention and shared intentional activity.<sup>21</sup> The next step is to explore



implications of this proposal for a pair of cases that have received much attention in the literature. This will lead us to a further element of such diachronic self-governance.

First, there are cases highlighted by John Broome in which one makes a decision in the face of non-comparable considerations and then reflects later on whether to stick with that decision.<sup>22</sup> Since, on the model, a planning agent's diachronic self-governance normally involves relevant plan continuities and connections, it will normally involve sticking with one's earlier decision in such cases of non-comparability over time. (Essays 6, 10, 11)

Second, there are cases highlighted by Richard Holton of potential willpower in the face of temptation.<sup>23</sup> Here the implications are more complex, and this because of a complex interaction between synchronic and diachronic self-governance. (Essays 7, 10, 11.)

Suppose that you know you will be tempted to drink a lot at tonight's party. Since you now think that would be a bad idea, you decide now to stick with one drink tonight. You know, however, that at the party your judgment will temporarily shift and you will at least initially judge that it would be better to have many drinks – though you also know that if you did give into this temptation you would later regret that. If at the party you stick with your prior one-drink intention there will be a salient continuity in intention; in contrast, if you give into temptation there will be a break in this continuity. However, such continuity of prior intention helps constitute diachronic self-governance only given synchronic self-governance at times along the way. But if your judgment shifts at the time of the party it seems that sticking with your prior intention would not be a case of synchronic self-governance: given the judgment shift, where you stand at the time of the party seems to favor drinking more. So it is unclear whether such cases of willpower can satisfy our model of a planning agent's diachronic self-governance. It seems to be built into such cases that there is a breakdown in coordination between self-

governance at the time of temptation and the kind of cross-temporal continuity characteristic of self-governance over time. Yet it seems a plausible, commonsense thought that willpower can sometimes be a central case of governing one's life over time.

The solution I propose is to supplement our model of a planning agent's self-governance over time by appeal to an end of the agent's that favors relevant intention continuity and thereby can potentially help re-shift her standpoint at the time of the party back in favor of a single drink. What end? We don't want just to appeal to an end of simple diachronic continuity. That would threaten to be an end in favor of mere cross-temporal mental tidiness; and that would expose our associated account of diachronic rationality to a version of the myth theoretic charge that these norms are simply tracking mental tidiness. I also think that we should not follow J. David Velleman in appealing here to a theoretical end of self-understanding.<sup>24</sup> Given our pursuit of a uniform account of plan rationality that tracks self-governance, this would return us to cognitivism about plan rationality – a view I have offered reasons to reject. (Essays 2, 3)

My proposal is instead to appeal to the end of one's diachronic self-governance itself.<sup>25</sup> (Essays 10, 11) This would help us explain how willpower can sometimes be a form of diachronic self-governance; it would avoid the charge that we are appealing to mere mental tidiness; and it would put us in a position to provide an alternative to a cognitivist account of diachronic plan rationality. So – by way of a kind of inference to the best explanation – I infer that the end of one's diachronic self-governance is itself an element in basic cases of a planning agent's exercise of her capacity for diachronic self-governance. Further, given that diachronic self-governance essentially involves relevant synchronic self-governance at times along the way, the presence of this end of one's diachronic self-governance will normally bring with it an end in favor of relevant synchronic self-governance.

This supplemented model of a planning agent's diachronic self-governance in hand, we can return to our effort to support a norm of diachronic plan rationality by way of the strategy of self-governance. The first step is to consider an extended tracking thesis: as synchronic plan rationality tracks conditions of a planning agent's synchronic self-governance, diachronic plan rationality tracks conditions of a planning agent's diachronic self-governance. This extended tracking thesis would have a pair of implications. First, there would be rational pressure in favor of continuity of intention over time, so long as that continuity coheres with self-governance at times along the way. Second, there would be rational pressure in the direction of an end of one's diachronic self-governance. We could then put these ideas together with our earlier, partial account of synchronic plan rationality and synchronic self-governance.

What should we say about such a package of views about plan rationality and self-governance? Well, given this overall package, norms of both synchronic and diachronic plan rationality are tied together and unified, and thereby made more intelligible, by the ways in which they track conditions of a planning agent's self-governance. Further, given this overall package we can conclude that if one has a normative reason in favor of one's self-governance, both synchronic and diachronic, and if one has the capacity for relevant self-governance, then one will have a distinctive normative reason of self-governance to conform to these norms. (Essays 4 and 6 address potential worries about unacceptable bootstrapping of reasons.)

This returns us to the question postponed earlier: why think that a reflective planning agent has a normative reason in favor of her self-governance? Now, we are asking about the reflections of a planning agent who is considering the norms that are guiding her practical thinking and wondering whether to continue to think in these ways. And I think that in such reflections a planning agent will be primarily concerned with considerations that are anchored in

ends of hers that are desirable. So, for our purposes, it will be reasonable to think of normative reasons as anchored in ends of the agent that are desirable.<sup>26</sup>

So: what to say about a normative reason in favor of one's self-governance? We can, I think, suppose that self-governance is desirable *pro tanto*, though a specific instance can be on-balance bad. (An example from essay 4 of self-governance that may well be on-balance bad is Iago.) So our question turns on the question: what is the status of the end of one's self-governance?

And here we face a dilemma. If our theory says only that

(a) the end of one's self-governance is a contingent, rationally optional end it would not be in a position to account for the general normative force for us of basic planning norms. But an account that said that

(b) the end of one's self-governance is essential for all intentional agency would, I think, overburden our theory of intentional agency: there are too many cases of intentional agency that is not plausibly motivated by some such end.

Is there a path between (a) and (b), one that would suffice for our defense of the normative force of basic planning norms? We might try here to appeal to our earlier conclusion that the end of one's diachronic self-governance is involved in central cases of the exercise of the capacity for such self-governance. But it is not clear how this is to work. It seems that at most what we can infer is that a planning agent with the capacity for diachronic self-governance must have the *capacity* for the end of her diachronic self-governance. So we do not have an argument that a planning agent with the capacity for diachronic self-governance must in fact have that end.

There are, however, two important truths in this neighborhood. (Essay 10) Consider a planning agent who *does* have this end in favor of her diachronic self-governance (and so, in

favor of related synchronic self-governance). This end grounds a normative reason for such self-governance; and, as we have seen, this reason helps support the application of norms of plan rationality to the particular case.<sup>27</sup> And here it is important to note that one of these norms is the cited norm of diachronic plan rationality, a norm that itself supports the presence of this very end (given the role of this end in the normal exercise of the capacity for such self-governance). So in this way this end of self-governance – where that includes both diachronic and, so, synchronic self-governance -- is, if present within the planning framework, *rationally self-sustaining*: it supports a norm of diachronic plan rationality that in turn supports it. Further, given this rationally self-sustaining end, the agent's package of plan-infused practical thinking, planning norms, ends, and reasons would be reflectively stable: this would be a *rationally stable reflective equilibrium*. This package would be reflectively stable even if it were in some sense possible (though pragmatically problematic, and perhaps not even psychologically possible) for the agent to give up being such a planning agent.

So this end of one's diachronic (and so, synchronic) self-governance is, if present within the planning framework, rationally self-sustaining and a keystone of a rationally stable reflective equilibrium. Let's call this observation about a rationally self-sustaining end of self-governance over time, and an associated rationally stable reflective equilibrium involving the norms of plan rationality, the *keystone claim*. And now we need to ask: How should we understand the significance of this keystone claim?<sup>28</sup> In particular, does this claim help us find the path for which we are looking, a path between the idea that (a) the end of self-governance is just one among many rationally optional ends, and the idea that (b) the end of self-governance is essential for all intentional agency? While essay 10 provides support for this keystone claim, it stops short

of a full answer to these further questions. And my hope is that we can make progress here by returning to, and further developing, an idea initially introduced in essay 2.

#### <1> 4.

In section 9 of essay 2 I began to explore a parallel between the framework of plan-infused practical thinking and the framework of reactive emotions, as understood by Peter Strawson.<sup>29</sup> And my conjecture here is that by further developing this parallel we can find an appropriate path between (a) and (b), a path supported by the keystone claim.<sup>30</sup>

Strawson famously highlighted the role of the reactive emotions in our practices of holding morally responsible, and so in our understanding of moral responsibility. As Strawson saw it, this framework of reactive attitudes is not strictly necessary for agency or mind; but it is a deeply entrenched framework for us, one that is integrated within much that is humanly significant, and one we may well not have the capacity to change at will. This framework has very many good consequences; but it is also important that built into this framework is an entrenched concern with quality of will, one that helps make sense of the normative structure of the reactive emotions including, in particular, the relevance of excuses of ignorance and absence of control.

Analogously, the planning framework is not strictly necessary for agency; but it is a deeply entrenched framework for us, one that is integrated within much that is humanly significant, and one we may well not have the capacity to change at will. This general framework has very many good consequences; but it is also important that built into this framework, at least for us, is an entrenched end of one's self-governance over time, an end that helps make sense of the normative structure of our plan-infused practical thinking including, in particular, the application of basic planning norms, both synchronic and diachronic, to the

particular case. This end of one's self-governance over time is (within the planning framework) rationally self-sustaining, favors a central, organizing commonality across the planning norms, and is thereby a keystone of a rationally stable reflective equilibrium involving the agent's plan-infused practical thinking. This is all true even if it is in some sense possible (though pragmatically problematic, and perhaps not even psychologically possible) for the agent to give up being such a planning agent.

Basic planning norms are, for us, embedded in such a rationally self-sustaining and stable reflective equilibrium. Whereas the keystone of the Strawsonian framework of reactive attitudes, and associated practices of holding responsible, is a concern with quality of will, the keystone of our framework of plan-infused practical thinking is a rationally self-sustaining end of one's self-governance over time. In each case the keystone helps make reflective sense of the normatively guided functioning of relevant attitudes – in one case, the system of reactive emotions; in another case, the system of plan-states. In each case the keystone – concern with quality of will/end of diachronic self-governance -- while not essential to mind or (intentional) agency *per se*, is also not merely one among many rationally optional, potential ends. After all, these keystones help make reflective sense of what are, for us, fundamental, pervasive, fecund, and entrenched webs of normatively guided thought and action. And, turning in particular to the planning framework, this puts us in a position to say that while basic planning norms are not (given the multiplicity of agency) essential to (intentional) agency *per se*, characterizing their purported special status as a “myth” fails to appreciate their keystone-supported status.

This allows that we can sometimes put to one side the end of our diachronic self-governance.<sup>31</sup> As Carlos Núñez has emphasized (in conversation), Strawson noted that we each have the power sometimes to bracket our concern with quality of will, perhaps as “a relief from

the strains of involvement".<sup>32</sup> But it remains true, according to Strawson, that our normatively guided practice of holding morally responsible is to be understood in terms of the expression of reactive emotions that involve this concern with quality of will. Analogously, we each perhaps have the power sometimes to bracket our concern with our own diachronic self-governance. But it remains true that our normatively guided plan-infused practical thinking is to be understood within an overall framework that involves that basic concern.<sup>33</sup>

This does not rule out the possibility of a thin kind of planning agency in the absence of a concern with self-governance. Perhaps this is how we should think about the planning of young human children, or the great apes, or certain artificial intelligence systems.<sup>34</sup> About such thin forms of planning agency the myth theorist might be right to say that it involves no distinctive form of practical rationality that goes beyond a two-tier pragmatic rationale, with its limits in supporting the application of general modes of thinking to the particular case. But such a thin planning agent would not be in a position to make direct sense to herself, on reflection, of her direct application of her planning norms to the particular case. In contrast, a planning agent who does have the cited concern with her own self-governance would be in a position to make reflective sense to herself of this basic aspect of her plan-infused practical thinking.

And my conjecture is that we ourselves are such reflectively successful and reflectively stable planning agents. While the myth theory may be right about, say, very young children, it is not right about us – mature human agents for whom structures of planning play a reflectively supportable and stable role in our cross-temporal organization, our sociality, and our self-governance. These pervasive and fecund forms of plan-infused practical thinking are not only in general extremely useful, though they are that. They are for us also in part supported by our



entrenched, reflectively stable end of our own self-governance over time, and in that way by who we in fact, and on reflection, are.<sup>35</sup>

---

<sup>1</sup> I include here both a norm that precludes planning p while believing not-p, and an agglomerativity norm that precludes planning X and planning Y while believing X and Y are not co-possible.

<sup>2</sup> This work took place against a background of field-defining work by Donald Davidson and under the significant influence of work of Gilbert Harman -- and, to some extent, Herbert Simon and Héctor-Neri Castañeda. See Donald Davidson, *Essays on Actions and Events*, 2nd edn. (Oxford: Oxford University Press, 2001). Gilbert Harman, "Practical Reasoning," in *Reasoning, Meaning, and Mind*. (Oxford: Oxford University Press, 1999). Gilbert Harman, *Change in View: Principles of Reasoning*. (Cambridge, MA: MIT Press, 1986). Herbert Simon, *Reason in Human Affairs*. (Stanford, CA: Stanford University Press, 1983). Héctor-Neri Castañeda, *Thinking and Doing*. (Dordrecht, Holland: Reidel. 1975). (For discussion of Castañeda's views see Michael E. Bratman, "Castañeda's Theory of Thought and Action," in *Faces of Intention*. New York: Cambridge University Press, 1999.) The main work of mine, prior to this present volume, in which I develop this plan-theoretic view of our agency is: Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987; Reissued CSLI Publications, 1999); Michael E. Bratman, *Faces of Intention* (New York: Cambridge University Press, 1999); Michael E. Bratman, *Structures of Agency* (New York: Oxford University Press, 2007); Michael E Bratman, *Shared Agency: A Planning Theory of Acting Together* (New York: Oxford University Press, 2014).

<sup>3</sup> Smart, J.J.C. "Extreme and Restricted Utilitarianism." In Philippa Foot, ed., *Theories of Ethics* (Oxford: Oxford University Press, 1967), 171-83.

---

<sup>4</sup> See my “Planning and the Stability of Intention,” *Minds and Machines* 2, no. 1 (1992): 1–16, and my “Following Through with One’s Plans: Reply to David Gauthier,” in Peter A. Danielson, ed. *Modeling Rationality, Morality, and Evolution* (New York: Oxford University Press, 1998), 55–66. And see essays 8 and 9 in this volume.

<sup>5</sup> Hugh McCann, “Settled Objectives and Rational Constraints.” *American Philosophical Quarterly* 28, no. 1 (1991): 25–36. The simple view says that intentionally A-ing always involves an intention to A. My argument against the simple view first appeared in my “Two Faces of Intention,” *The Philosophical Review* 93 (1984): 375–405. Large parts of “Two Faces of Intention” were then incorporated into Chapter 8 of my *Intention, Plans, and Practical Reason*.

<sup>6</sup> Raz, Joseph. “The Myth of Instrumental Rationality.” *Journal of Ethics and Social Philosophy* 1, no. 1 (2005): 1–28. Kolodny, Niko. “The Myth of Practical Consistency.” *European Journal of Philosophy* 16, no. 3 (2008): 366–402.

<sup>7</sup> Gilbert Harman, “Practical Reasoning”; J. David Velleman, *Practical Reflection* (Princeton, NJ: Princeton University Press, 1989; Reissued Stanford, CA: CSLI, 2007); R. Jay Wallace, “Normativity, Commitment, and Instrumental Reason,” with postscript in his *Normativity and the Will* (Oxford: Oxford University Press, 2006); Kieran Setiya, “Cognitivism about Instrumental Reason.” *Ethics* 117, no. 4 (2007): 649–673. And Velleman in a 2007 essay [“What Good is a Will?” as re-printed in *Rational and Social Agency: The Philosophy of Michael Bratman*, Manuel Vargas and Gideon Yaffe. eds, (Oxford: Oxford University Press, 2014)], and Setiya in a 2014 essay [“Intention, Plans, and Ethical Rationalism,” in *Rational and Social Agency: The Philosophy of Michael Bratman*, Manuel Vargas and Gideon Yaffe. eds, (Oxford: Oxford University Press, 2014)], also went on to provide insightful critiques of my approach to these matters. My initial response to Velleman’s critique is in essay 2. My initial response to

---

Setiya is in my “Rational and Social Agency: Reflections and Replies,” in Manuel Vargas and Gideon Yaffe, eds., *Rational and Social Agency: The Philosophy of Michael Bratman* (Oxford: Oxford University Press, 2014), at 313-315. This present volume of essays constitutes my overall response to both of these philosophers.

<sup>8</sup> Ideas in the spirit of such agency-constitutivism are in Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press, 2009) and J. David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).

<sup>9</sup> Paul Grice, “Method in Philosophical Psychology (From the Banal to the Bizarre),” Presidential Address, *Proceedings and Addresses of the American Philosophical Association* 48 (1974-5): 23–53; Michael E. Bratman, “Valuing and the Will” in *Structures of Agency*.

<sup>10</sup> Setiya in his “Intention, Plans, and Ethical Rationalism” insightfully poses this issue in terms of what he calls “pluralistic rationalism”. My initial response is in my “Rational and Social Agency: Reflections and Replies,” at 313-315. In essay 10 and this Introduction I try to deepen this response. Though my appeal to a two-tier pragmatic rationale contrasts with such pluralistic rationalism, the version of the strategy of self-governance developed here is to some extent in the spirit of such pluralistic rationalism.

<sup>11</sup> For example, John Broome writes: “Rationality is concerned with coherence among your attitudes such as your beliefs and intentions...” See John Broome, *Rationality Through Reasoning* (Wiley-Blackwell, 2013), at 193.

<sup>12</sup> Niko Kolodny, “How Does Coherence Matter?” *Proceedings of the Aristotelian Society* 107, no. 1 (2007): 229–263, at 241.

<sup>13</sup> I was also helped by Kenneth Stalzer’s work on his 2004 Stanford University Ph.D. thesis, *On the Normativity of the Instrumental Principle*.

---

<sup>14</sup> This paragraph was aided by correspondence with Sarah Paul.

<sup>15</sup> Since each essay was written to stand on its own, there are also redundancies, for which I request the reader's patience.

<sup>16</sup> In writing this Introduction I have drawn also on my "Plan Rationality," in Ruth Chang and Kurt Sylvan, eds., *The Routledge Handbook of Practical Reason* (forthcoming), my Pufendorf Lectures at Lund University in June 2016, and my Franz Brentano Lectures on Practical Philosophy at the University of Vienna in April, 2017.

<sup>17</sup> See essays 4 and 5; though see also an important challenge in Elijah Millgram, "Segmented Agency," in Manuel Vargas and Gideon Yaffe, eds., *Rational and Social Agency: The Philosophy of Michael Bratman* (Oxford: Oxford University Press, 2014): 152-89.

<sup>18</sup> See Essay 10, though I there in effect collapse the third and fourth element to be distinguished below.

<sup>19</sup> For a challenge to my response to concerns about inappropriate bootstrapping, see John Brunero "Self-Governance, Means-Ends Coherence, and Unalterable Ends" *Ethics* 120 (2010), 579-591.

<sup>20</sup> I discuss below the relation between the desirability of self-governance and such a reason for self-governance.

<sup>21</sup> A complexity in the background concerns the hierarchical structure of plans: there can be relevant intra-personal inter-connections at the level of an overarching plan despite a breakdown in such inter-connections at the level of relevant sub-plans. I discuss this complexity in essay 11.

<sup>22</sup> John Broome, "Are Intentions Reasons? And How Should We Cope with Incommensurable Values?" in Christopher W. Morris and Arthur Ripstein, eds., *Practical Rationality and Preference: Essays for David Gauthier* (Cambridge: Cambridge University Press, 2001): 98-120.

---

<sup>23</sup> Richard Holton, *Willing, Wanting, Waiting* (Oxford: Clarendon Press, 2009). And see Michael E. Bratman, “Toxin, Temptation, and the Stability of Intention,” in *Faces of Intention*.

<sup>24</sup> See e.g., J. David Velleman, “Centered Self,” in his *Self to Self* (Cambridge University Press, 2006), 253-283; J. David Velleman, *How We Get Along* (Cambridge University Press, 2009), chap. 1.

<sup>25</sup> In essay 7 I say: “a planning agent is committed to shaping present thought and action with an eye to how her planned activities will stably look as they develop over time.” (This volume p. 00.) Here (and in essays 10 and 11) I replace this appeal to a concern with how things will look over time with an appeal to a concern with one’s self-governance over time; and I interpret this concern with one’s self-governance over time as an end of the agent’s. I then try, toward the end of this Introduction, to articulate a more nuanced account of the relation between planning agency and this concern.

<sup>26</sup> As I discuss in essay 10, not all ends are intentions and not all intentions are ends.

<sup>27</sup> Though, again, see the responses in essays 4 and 6 to concerns about inappropriate bootstrapping.

<sup>28</sup> Part of the answer concerns the relation between this keystone claim and ideas about “self-reinforcement” in essays 8 and 9. Essay 8 considers the idea, within the framework of work of David Gauthier, that a deliberative standard can be “self-reinforcing in the sense that its very acceptance provides, by way of considerations of self-governance, reasons that ...[can] tip the scales in favor of employing that deliberative standard in the particular case.” (Essay 8 at p. 00.) Essay 9 claims that if (a) we take it as given that a planning agent has a normative reason in favor of her self-governance, then we can argue that if we are (as there are strong pragmatic reasons to be) planning agents, and if we have the capacity for self-governance in the particular

---

case, then (b) we have a reason of self-governance in support of the application to that particular case of the norms of plan consistency and coherence whose acceptance is partly constitutive of our planning agency. In this sense our planning agency is, given (a), self-reinforcing.

In each case, the cited self-reinforcement -- of a deliberative standard; of planning agency itself -- works against a background of a reason or end in favor of self-governance. The keystone claim goes on to focus on the end of one's self-governance. It claims that this end is, within a planning framework, a rationally self-sustaining element of a reflectively stable psychic economy. And our question now is what exactly this tells us about the status of that end, and thereby also about the kinds of self-reinforcement noted in essays 8 and 9.

For a different development of the metaphor of a keystone see Keith Lehrer's discussion of "the keystone loop of reason," and of a related idea of a self-supporting "power preference," in his *Self-Trust: A Study of Reason, Knowledge, and Autonomy* (Oxford: Oxford University Press, 1997), esp. chaps. 1 and 4. (Quotations are from pp. 22 and 100.)

<sup>29</sup> Peter Strawson, "Freedom and Resentment," in Gary Watson, ed., *Free Will* second edition (Oxford: Oxford University Press 2003): 72-93.

<sup>30</sup> For a related discussion see Luca Ferrero, "Inescapability Revisited," (unpublished manuscript, April 2016), section 2.4. In this present section I have benefitted greatly from detailed comments from Ferrero on an earlier draft.

<sup>31</sup> Cp. Sarah Paul, "Diachronic Incontinence is a Problem in Moral Philosophy," *Inquiry* (2014): 337-55, at 345.

<sup>32</sup> Strawson, "Freedom and Resentment," at 82.

<sup>33</sup> As Facundo Alonso has noted (in correspondence), a cognitivist about these planning norms might appeal here to its own version of an analogue of Strawson's framework, one within which

---

a concern with, say, one's self-understanding plays an essential role. However, this would not address the substantive problems for such cognitivism that I discuss in essays 2 and 3. That said, there remains the further question of whether the approach to plan rationality I am sketching here can be extended to norms of theoretical rationality of belief.

<sup>34</sup> Concerning this last idea see Michael E. Bratman, David Israel, and Martha Pollack. "Plans and Resource-Bounded Practical Reasoning." *Computational Intelligence* 4, no. 3 (1988): 349–355.

<sup>35</sup> Thanks to Facundo Alonso, Olle Blomberg, Jorah Dannenberg, Luca Ferrero, Carlos Núñez, Sarah Paul, and Steven Woodworth.

---

## PLANNING, TIME, AND SELF-GOVERNANCE: ESSAYS IN PRACTICAL RATIONALITY

### Preface

1. Introduction: The Planning Framework
2. Intention, Belief, Practical, Theoretical (2009)
3. Intention, Belief and Instrumental Rationality (2009)
4. Intention, Practical Rationality, and Self-Governance (2009)
5. Agency, Time, and Sociality (2010)
6. Time, Rationality, and Self-Governance (2012)
7. Temptation and the Agent's Standpoint (2014)
8. The Interplay of Intention and Reason (2013)
9. Consistency and Coherence in Plan (2014)
10. Rational Planning Agency (2017)
11. A Planning Agent's Self-Governance Over Time

### Bibliography