Rational Planning Agency[1]

Michael E. Bratman 02-08-17

Forthcoming in Royal Institute of Philosophy volume on Philosophy of Action; subject to small changes; please quote from published version

ABSTRACT. Our planning agency contributes to our lives in fundamental ways. Prior partial plans settle practical questions about the future. They thereby pose problems of means, filter solutions to those problems, and guide action. This plan-infused background frames our practical thinking in ways that cohere with our resource limits and help organize our lives, both over time and socially. And these forms of practical thinking involve guidance by norms of plan rationality, including norms of plan consistency, means-end coherence, and stability over time.

But why are these norms of rationality? Would these norms be stable under a planning agent's reflection? I try to answer these questions in a way that responds to a skeptical challenge. While I highlight pragmatic reasons for being a planning agent, these need to be supplemented fully to explain the force of these norms in the particular case. I argue that the needed further rationale appeals to the idea that these norms track certain conditions of a planning agent's self-governance, both at a time and over time. With respect to diachronic plan rationality, this approach leads to a modest plan conservatism.

We are planning agents. In support of both the cross-temporal and the social organization of our agency, and in ways that are compatible with our cognitive and epistemic limits, we settle on partial and largely future-directed plans. These plans pose problems of means and preliminary steps, filter solutions to those problems, and guide action. As we might say, we are almost always already involved in temporally extended planning agency in which our practical thinking is framed by a background of somewhat settled prior plans.

In this plan-infused practical thinking we are guided by norms of plan rationality. These include norms of plan consistency, including plan-belief consistency and the possibility of agglomerating one's various plans without running into problems of consistency.[2] These consistency norms are in the background of the filtering roles of

our prior plans. These norms also include a norm of means-end coherence – a norm, roughly, that mandates intending believed necessary means to ends intended.[3] This norm is in the background of the problem-posing roles of our prior plans. And these norms include a norm of stability of plans over time, one that is in the background of the default stability of our plans in the support of cross-temporal and social organization. While these norms admit of qualifications, we do not simply treat them as rules of thumb: we normally see their violation as a mistake/a breakdown. And guidance by these norms helps frame practical thinking in which we weigh pros and cons with respect to specific decisions that are on the deliberative table.

In other work I have developed these ideas as part of what I have called the planning theory of our human agency.[4] In this essay I want to reflect on the way in which the status of these norms can seem puzzling. After all, it is common to desire non-co-possible things, or to desire an end without desiring means. Why are intentions and plans different? Further, there seem to be cases in which we can better pursue our basic ends by adopting plans that are not consistent with each other.[5] And we can wonder what the problem is in failing to intend believed necessary means to an intended end if that belief is false, or if we have no good reason for our intended end.

These reflections point to a fundamental challenge: in giving these norms their own independent significance are we endorsing an unjustified fetish for 'psychic tidiness'?[6] We can think of this as a challenge to the reflective stability of these norms.[7] According to this challenge, a planning agent who accurately reflects on these structures of her practical thinking will reject these as norms with independent normative significance, since she will reject a brute appeal to the significance of mere mental tidiness. She will come to see that appeal to these norms as basic norms of practical rationality is an indefensible 'myth'. So these norms will not be stable under her critical reflection.

Such a failure of stability under reflection would pose a challenge to the descriptive and explanatory ambitions of appeals to our planning agency. If the norms involved in such agency would not survive a planning agent's critical reflection then it would be less plausible that thinking shaped by these norms is a basic feature of human agency.[8] And my aim in this essay is respond to this challenge.[9]

In doing this I assume that if we can show that these norms are both central to the basic structure of a planning agent's practical thinking and would survive a planning agent's critical reflection, then we can justifiably conclude that these are indeed norms of practical rationality for a planning agent. This is not yet to determine whether we should be in a strong sense realists about these norms. But for our purpose of defending the explanatory ambitions of the planning theory we need not settle that metaphysical question.

So how might we establish the reflective stability of these norms? Focusing at first on synchronic norms of plan consistency and coherence, an initial idea might be to see these norms as riding piggy-back on norms of theoretical rationality that enjoin consistency and coherence of associated beliefs. This is cognitivism about these aspects of plan rationality. In other work, however, I have argued that this is not going to work.[10] This is primarily (but not solely) because one might believe one intends the necessary means to an end one intends and yet not in fact intend those means. In such a case one's beliefs might be theoretically coherent even though one's intentions do not conform to the norm of means-end coherence.

A second idea might be to say that our acceptance of these norms is inescapable for agents, so there is no real problem about their reflective stability. And, indeed, it is a central feature of planning agency, as understood within the planning theory, that one's intentions and plans are guided by one's (perhaps, implicit) acceptance of these norms. But this would only show that the acceptance of these norms is inescapable for agents if planning agency were inescapable for agents. But it isn't. One can be a goal-directed agent who acts purposively and for reasons but is nevertheless not a planning agent. This is an aspect of the multiplicity of agency.[11]

Granted, we may not have the capacity, just like that, to give up being a planning agent and become, instead, a non-planning agent. But even if there were this contingent incapacity, we would need to address the possibility that a planning agent would, on reflection, be alienated from these norms in a way that would threaten their longer-term stability and challenge their status as basic features of our agency.[12]

Donald Davidson's work on interpretation points to another kind of inescapability. Davidson treated norms of rationality as a single over-all package. With respect to that package he wrote:

> It is only by interpreting a creature as largely in accord with these principles that we can intelligibly attribute propositional attitudes to it …An agent cannot fail to comport most of the time with the basic norms of rationality.[13]

Broad conformity to certain basic norms of consistency and coherence is a fundamental feature of the attitudes we ascribe in interpreting an agent.

Something along these lines seems right. But it does not solve our problem about plan rationality. First, this would not explain why a violation of these norms in the particular case is a breakdown. At most what is claimed to be inescapable for a person with a mind is failing to 'comport most of the time' with relevant norms.[14] Second, Davidson sees the relevant norms of rationality as a single over-all package, one involved quite generally in interpreting minds. But once we note the multiplicity of agency we need to be alive to the possibility of a minded agent who is not a planning agent. So Davidson's idea about interpreting minds does not establish the inescapability, for a minded agent, of conformity with the norms of plan rationality.

A fourth idea would highlight the large benefits to us of our planning agency. Given general features of our minds and our environments, our pursuit of our most basic ends will normally be made more effective by our plan-shaped practical thinking, practical thinking that supports both cross-temporal and social organization and, as I will discuss, our self-governance. This fecundity of planning agency supports the idea that we have good reason to be planning agents. Since our planning agency involves the application of the cited norms to particular cases, we may then try to infer that we have good reason to conform to these norms in their application to particular cases. This would be a two-tier pragmatic justification of these norms.

But, as we have learned from J.J.C. Smart, there is a problem.[15] Even given the advantages of general patterns of thought guided by norms of plan consistency and coherence, there can be particular cases in which it is known that conformity to these norms would not be as effective, with respect to the very same benefits, as would divergence. Perhaps sometimes it is useful to have inconsistent or incoherent plans.

But what we are seeking is not just a defense of a general tendency to conform to these norms. We are also seeking a justification of the application of these norms to the particular case.

There is an insight built into the two-tier pragmatic approach: the general capacity for planning agency is good in myriad ways. But what we learn from Smart is that we need also to provide a further rationale that, given that one is (as there is reason to be) a planning agent, supports the application of these norms to the particular case. Otherwise, we cannot be fully confident that the acceptance of these as norms with independent significance in application to the particular case will be reflectively stable.

Here we can learn from Gilbert Harman's suggestion that in theorizing about such norms we follow

> a process of mutual adjustment of principles to practice and/or intuitions, a process of adjustment which can continue until we have reached what Rawls (1971) calls a reflective equilibrium. Furthermore, and this is important, we can also consider what rationale there might be for various principles we come up with and that can lead to further changes in principles, practices, and/or intuitions.[16]

Our concern is with the stability under reflection of planning norms. Following Harman's suggestion, we can understand such reflection on the part of a planning agent as 'a process of mutual adjustment' and search for a 'rationale' that underlies the norms that guide one's plan-infused practical thinking. We can suppose that this rationale will involve some sort of two-tier pragmatic support. But it will need to go beyond that. So we ask: is there some further consideration appeal to which could supplement the two-tier pragmatic approach and enable the reflective planning agent to make good normative sense of her application of these norms in the particular case? This would enable the reflective planning agent to defend her norms by way of a kind of inference to the best normative explanation.

In pursuit of such a best normative explanation I will frequently speak directly in my own voice. But in doing so I take myself to be speaking on behalf of a planning agent who is reflecting on her plan-infused practical thinking. It is the reflective stability for a planning agent of that practical thinking that is our main concern.

We can articulate three inter-related desiderata for a rationale that underlies these planning norms. First, and in partial response to the myth theorist's challenge, it should explain why the forms of coherence at stake in these norms are not merely a matter of mental tidiness. Second, it would be good if this rationale articulated a relevant commonality across these different norms, both synchronic and diachronic. And third, it should explain why there is a systematically present normative reason that favors conformity to these norms.[17]

How should we understand this talk about normative reasons? This is controversial territory. But I think that, given our concern with the stability of these planning norms in light of the agent's own reflection, it is reasonable for us to work with a model of reasons as anchored in ends of the agent where what those ends favor is desirable. Roughly: a consideration is a reason for S to A only if it helps explain why S's doing A is needed to promote relevant ends of S,[18] and only if what these ends favor is desirable. A planning agent reflecting on her own practical thinking will have a keen interest in what is needed to promote her ends and in whether these ends favor what is desirable. So it makes sense, for our present purposes, to work within this dual framework in exploring the reflective stability of the planning norms.

Let me briefly clarify my talk here of an agent's ends. Roughly: to have E as an end is to have a non-instrumental concern in favor of E. Not all such ends are intentions since, in contrast with intentions, not all ends tend to diminish when they are not co-realizable in light of one's beliefs. It is common in our complex lives to have ends that we know are not jointly realizable even while we believe that each is realizable. Further, one may intend X even if X is not in this sense one of one's ends, since one's intention may favor X solely instrumentally. Nevertheless, an intention in favor of X solely as a means can still induce rational pressure for an intention in favor of a known necessary means to X. So we need to be careful to understand the idea of an intended end, as it appears in the norm of means-end coherence, in a way that does not require that what is intended is, strictly speaking, an end of the agent's.

The second desideratum seeks a commonality across synchronic and diachronic norms. What diachronic norm? The idea is that our planning agency involves a norm of stability of plans over time. What norm?

Note two preliminary ideas. First, a prior intention at t1 to A at t2 will frequently lead to change in the circumstances between t1 and t2 in ways that reinforce that intention. Think about buying, at t1, a non-refundable ticket in support of your intention at t1 to fly to London at t2. This is the snowball effect.[19] Second, having formed the prior intention it may not be rational to reconsider. After all, reconsideration has its own costs and risks, especially for resource-limited agents like us. And normally, if one does not reconsider one's rationally formed prior intention, then one continues rationally so to intend.

In my 1987 book I focused on these two aspects of the rational stability of intention over time. But I have come to think that there is more to say about this rational stability. My reasons for this primarily involve two cases of potential intention stability. I will focus first on a case involving potential willpower.[20] Later I will turn to a second case. In the end, a virtue of the account I will propose will be that it treats both cases within the same overall framework, one that also supports a significant commonality in the rationale underlying synchronic and diachronic plan rationality.

Suppose that you know you will be tempted to drink heavily tonight at the party. You now think that, in light of what matters to you, this is a bad idea. However, you know that at the party your evaluation will shift in favor of drinking more. You also know that if you did drink heavily your evaluation would later shift back and you would regret that. So this morning you resolve to drink only one glass tonight. The problem is that, as you anticipate, if you were to stick with your resolve at the party, you would act against what would then be your present evaluation. And we normally suppose that action contrary to one's present evaluation is a rational breakdown. So how could you rationally follow through with your resolve?

As Sarah Paul has emphasized (in conversation), cases with this structure are ubiquitous in our lives.[21] We many times face temporarily shifted evaluations with respect to continuing with an ambitious project when, as it is said, the going gets tough. And even in the case of more modest temporally extended projects, we frequently face issues of procrastination. In following through with planned temporally extended activities one will frequently be tempted to procrastinate just a bit. It will frequently seem that one could get the benefits of the planned activity plus a small incremental

benefit of, say, reading just one more e-mail.[22] Problems of willpower and temptation pervade our planned temporally extended activities. If we are going to understand the deep ways in which our plans help support important forms of cross-temporal organization we will need to understand how those plan structures are responsive to such de-stabilizing pressures. So we will need to ask whether there is at work here a norm of diachronic plan stability that goes beyond snowball effects and issues of rational non-reconsideration.

I turn to this question below. But first we need to return to our general pursuit of a rationale that underlies the planning norms in a way that suitably supplements the two-tier pragmatic account.

Here I propose a *strategy of self-governance*: a basic rationale underlying these norms, one that supplements the two-tier pragmatic account, appeals to a planning agent's self-governance, both synchronic and diachronic.[23] Planning norms, both synchronic and diachronic, track forms of coherence that are essential to a planning agent's self-governance, and such coherence is not merely mental tidiness. Further, a systematically present reason in favor of conformity to these norms is grounded in one's reason to govern one's own life. A planning agent with the capacity for self-governance will be in a position to conclude, on reflection, that the best rationale for her planning norms, one that supplements the two-tier pragmatic account, appeals in these ways to the significance of self-governance. And given this rationale, her acceptance of these norms will be reflectively stable. Or so I will argue.

This is to focus on planning agents with the capacity for self-governance. At some point we would need to consider planning agents who do not have the capacity for self-governance – 3-year old humans, perhaps. But I put this issue aside here.

The idea is not to see self-governance as a constitutive aim of agency.[24] As I see it, an appeal to such a substantive constitutive aim would overburden[25] our descriptive and explanatory theory of action: there are just too many cases of agency that seems not to be guided by such an aim. R. Jay Wallace gives us a lively sense of this point when he highlights 'sheer willfulness, stubbornness, lethargy, habit, blind self-assertion, thoughtlessness, and various actions expressive of emotional states.'[26] Just as the legal

positivists distinguished between law as it is and law as it ought to be, we should distinguish between agency as it is and as it ought to be.

Again, the idea is not to appeal to self-governance to convince a purposive but non-planning agent to try to become a planning agent. To be sure, there are strong pragmatic reasons for making such a transition, if one can. But that is not the main focus of the strategy of self-governance. Its main concern is, rather, directly to address an agent who is already a planning agent, one whose reasoning accords the cited planning norms an independent normative significance. In addressing such an agent the self-governance strategy aims to articulate a rationale to which that agent can appeal to make good normative sense of her plan-infused practical thinking.

Such a rationale would need to be responsive to our three desiderata: articulate relevant forms of coherence that are not merely mental tidiness; articulate a relevant commonality across the different norms; and identify a systematically present reason in favor of conformity. While this last desideratum – as I will call it, the *reason desideratum* -- is fundamental, I will for now put it to one side and try to articulate a structure of self-governance-based norms that is responsive to concerns with coherence and commonality. In this way I will try to construct an initial, prima facie, though not yet conclusive self-governance-based case for these norms. I will then return to the reason desideratum.

A first step is to sketch a broadly naturalistic model of self-governance at a time (or anyway, during a small temporal interval). And here we can learn from Harry Frankfurt's idea of 'where (if anywhere) the person himself stands.'[27] Self-governance involves guidance of thought and action by where the agent stands, by the agent's relevant practical standpoint. Such a standpoint will need to be sufficiently coherent to constitute a clear place where the agent stands on relevant practical issues. It will need to guide choice. And choice will need to cohere with that coherent standpoint.

So coherence of relevant standpoint and coherence of choice with standpoint are elements in our model of self-governance at a time. And now we can propose, as part of our pursuit of inference to the best normative explanation, that there is a close connection between these forms of self-governance-related coherence and practical rationality. In doing this we will want a somewhat qualified connection. Incoherence of

standpoint with respect to trivial choices, or with respect to tragic conflicts,[28] may not be irrationality. And we will want to leave room for a pro tanto or local rational breakdown that, while it is not merely a potentially misleading, prima facie indicator of irrationality, nevertheless does not ensure all-in irrationality.[29] So consider:

> *Practical Rationality/Self-Governance (PRSG)*: If S is capable of self-governance it is, defeasibly, pro tanto irrational of S either to fail to have a coherent practical standpoint or to choose in a way that does not cohere with her standpoint.

PRSG says that if S is capable of self-governance and yet fails to satisfy the cited coherence conditions of synchronic self-governance then, defeasibly, S is pro tanto irrational. The connection it articulates between a breakdown in self-governance-related coherence and irrationality is doubly qualified: it is a defeasible connection to pro tanto irrationality. But such a breakdown in self-governance-related coherence is not merely a potentially misleading prima facie indicator concerning what really matters.

As noted, this is so far only an initial, prima facie case for PRSG. I will turn later to the reason desideratum; but first, let's see how this initial case can be extended to, more specifically, a planning agent.

A planning agent will have a web of plans that settle – frequently in the face of conflict -- on certain projects, as well as on certain considerations that are to matter in the pursuit of those projects. These plans will normally cross-refer to each other: one's plan for today will typically involve a reference to one's earlier and later plans; and vice versa. These issue-settling, cross-referring plans will frame much of one's practical thought and action over time. They will pose problems about how to fill in so-far partial plans with sub-plans about means and the like, sub-plans that mesh with each other. And they will filter options that are potential solutions to those problems. In playing these roles these plans will induce forms of psychological connectedness and continuity that are familiar from Lockean models of personal identity over time.

This leads to a proposal about self-governed planning agency.[30] Given the settling, cross-referring, framing, mesh-supporting, and Lockean-identity-supporting roles of her plans, a planning agent's practical standpoints will involve her web of plans concerning both projects and considerations that are to matter in her practical thinking: her practical standpoints will be *plan-infused*. A planning agent's plans help constitute

her practical standpoint at a time in part because of their roles in structuring her temporally extended practical thought and action over time. So the guidance of her thought and action by these planning structures will help constitute her relevant self-governance. In such self-governance, plan-infused standpoints will need to be both coherent and coherent with choice. And when we combine this point about a planning agent's self-governance with PRSG we arrive at

> *Practical Rationality/Self-Governance-Planning Agency (PRSG-P):* If S is a planning agent who is capable of self-governance it is, defeasibly, pro tanto irrational of S either to fail to have a coherent practical plan-infused standpoint or to choose in a way that does not cohere with her plan-infused standpoint.

Again, what we have so far is only an initial, prima facie case in favor of PRSG-P. Keeping this limitation in mind, however, we can explore the implications of PRSG-P concerning plan consistency and coherence. And the basic idea here is that inconsistency or incoherence in plan, given one's beliefs, normally baffles the coherence of plan-infused standpoint that is needed for there to be a clear place where the agent stands with respect to relevant issues. If you intend A and intend B, while believing that A and B are not co-possible, there is no clear answer to the question of where you stand with respect to this practical question. If you intend A but believe not-A then you will normally be buffeted by conflicting dispositions to plan on the assumption that A and to plan on the assumption that not-A.[31] In this way there will be no clear answer to the question of where you stand with respect to A. And if you intend E but do not intend believed necessary means to E even though you believe it has come time to (as we say) fish or cut bait, there will be no clear answer to the question of where you stand with respect to E. In each case there is a contrast with ordinary desire: desires for non-co-possible things, or for things one believes will not happen, or for ends in the absence of desiring the means, are a common feature of our lives and need not block relevant coherence of standpoint.

A complication is that there can be intention analogues of the preface paradox.[32] Perhaps one has a wide range of plans for one's vacation, but sensibly believes that one will not accomplish everything one plans. So it is not possible to realize all one's

intentions in a world in which all of one's beliefs are true. Still, one may sensibly proceed to plan in the normal way with respect to each of one's intended ends.

This suggests that in certain preface-analogue cases plan-belief inconsistency may not induce a breakdown in coherence of plan-infused standpoint. So we have a double defeasibility. As noted earlier, coherence of plan-infused standpoint is, defeasibly, needed to avoid self-governance-grounded pro tanto irrationality. To this we add that plan-belief consistency is defeasibly needed for coherence of plan-infused standpoint. We thereby arrive at:

> *Plan consistency and coherence (PCC)*: If S is a planning agent who is capable of self-governance it is, doubly-defeasibly, pro tanto irrational of S to have plans that are inconsistent or means-end incoherent, given her beliefs.

What underlies this rational pressure against plan inconsistency or incoherence is not merely mental tidiness but the coherence of standpoint that is essential to a planning agent's synchronic self-governance.

Though this is so far only a prima facie case in support of PCC, we can go on to ask whether this approach to synchronic plan rationality could be extended to diachronic plan rationality. Does diachronic plan rationality track a kind of cross-temporal coherence that is central to a planning agent's diachronic self-governance?[33]

To explore this we need a model of a planning agent's self-governance not only at a time but also *over* time.[34] An initial idea is that a planning agent's self-governance over time involves her self-governance at times along the way while engaging in a planned temporally extended activity, where these forms of synchronic self-governance are appropriately interconnected. But what interconnections are these?

Here I propose that they involve the interconnections between plan-infused attitudes that are characteristic of planned temporally extended activity, all in the context of self-governance at times along the way. These interconnections will include forms of continuity of intention, cross-reference between intentions, intended mesh in sub-plans, and interdependence between intentions and/or expectations of intention.[35] Further, though I cannot defend this here, I think such cross-temporal intention inter-connections within planned temporally extended activity will be significantly analogous to the inter-personal intention inter-connections highlighted in the account of shared intentional

action I have developed elsewhere.[36] This is a version of an important parallel between the cross-temporal organization of an individual's activity and inter-personal, social organization. The idea that a planning agent's diachronic self-governance involves inter-connections characteristic of planned temporally extended activity, taken together with this parallel between the individual and the social, supports the metaphor that in her self-governance over time a planning agent is 'acting together' with herself over time.

Will willpower comport with a planning agent's diachronic self-governance, so understood? Well, in a temptation case involving evaluation shift, following through with one's prior resolution to resist the temptation, while it would involve the cited cross-temporal inter-connections, would conflict with one's then-present evaluation. It seems to follow that sticking with one's prior resolution would be incompatible with synchronic self-governance, and so with diachronic self-governance. But it also seems an important commonsense idea that willpower can be a central case of diachronic self-governance.

In responding, we do not merely seek some sort of causal mechanism in the psychology that can explain why one sometimes sticks with one's resolve. We want to explain why, at least sometimes, sticking with one's resolve coheres with one's present standpoint and, in part for that reason, coheres with self-governance. And here we will want to appeal to some general feature of the agent's standpoint that helps explain how the prior resolve sometimes helps re-shift the standpoint in favor of willpower. But what feature?

We do not want simply to appeal to an end of constancy of intention,[37] since such an appeal would face familiar concerns about a fetish for (in this case, cross-temporal) mental tidiness. In a discussion of related matters, J. David Velleman proposes that we appeal to our interest in understanding ourselves: 'my intellectual drives … favor fulfilling my past intentions.'[38] Given the commonality desideratum, however, this will lead to a general cognitivism about planning norms, with all its difficulties.

So let me propose instead that we appeal to the end of diachronic self-governance – where such self-governance is understood in terms of the model we are hereby developing. This is not an appeal simply to an end of constancy of intention; but it is also not an appeal that leads to cognitivism. This end would sometimes support willpower in the face of temptation, since such willpower would involve the cross-

temporal continuity and interconnection of plan structures that is an element in diachronic self-governance.[39] In this way this end would be poised to help stabilize the agent's temporally extended, planned activities. Further, the presence of this end would help explain why a planning agent's diachronic self-governance is, at least frequently, intentional under that description.

Granted, this end of diachronic self-governance, even if present, may be overridden by other ends in the agent's standpoint at the time of temptation. And if it is overridden then sticking with one's prior resolution will not comport with synchronic self-governance; and so it will not comport with diachronic self-governance. But sometimes this end of diachronic self-governance can indeed help re-shift the agent's standpoint at the time of temptation to favor willpower. So this end can sometimes support the coordination of synchronic self-governance and diachronic continuity in such cases. In this way, willpower in the face of temptation may comport both with synchronic and with diachronic self-governance, given the end of diachronic self-governance. This does not explain how willpower always comports with self-governance; but we do not need to explain that, since it is not true.

So let's model a planning agent's diachronic self-governance as involving this end of diachronic self-governance. A planning agent's self-governance over time involves coordination of two kinds of coherence within planned temporally extended activity: the synchronic coherence involved in self-governance at times along the way, and the coherence involved in relevant cross-temporal continuities and inter-connections of intentions over time. And this coordination of these two forms of coherence is to some extent supported by standpoints that include the very end of diachronic self-governance.

Return now to diachronic plan rationality. In discussing synchronic plan rationality I argued that a reflective planning agent with the capacity for self-governance would be led to the idea that there is, defeasibly, pro tanto rational pressure in favor of the coherence that is partly constitutive of synchronic self-governance. This would be an inference to the best normative explanation of her plan-infused practical thinking – though, as noted, the support for this is so far only prima facie, since we have so far not addressed the reason desideratum. So let us now, in the same spirit, ask whether a reflective planning agent with the capacity for diachronic self-governance would be led

to the idea that there is, defeasibly, pro tanto rational pressure in favor of the forms of coordinated coherence that are partly constitutive of a planning agent's diachronic self-governance.

A basic thought here is that there is a natural generalization available to the reflective planning agent. Just as there is rational pressure for the coherence central to her synchronic self-governance, so there is rational pressure for the coherence central to her diachronic self-governance. In each case the underlying idea, supported by an inference to the best normative explanation of her plan-infused practical thinking, is that there is rational pressure for coherence that is partly constitutive of her self-governance. So there is, in particular, rational pressure in favor of the coordination of synchronic and diachronic coherence that is characteristic of a planning agent's diachronic self-governance. So consider:

Diachronic Plan Rationality (DPR): If S is a planning agent who is capable of diachronic self-governance then the following is, defeasibly, pro tanto irrational of S:

(a) S is engaged in a planned temporally extended activity that has so far cohered with both synchronic and diachronic self-governance.

(b) Given her present standpoint, a choice to continue with her planned activity would cohere with that standpoint and so cohere with her continued synchronic self-governance and, in part for that reason, with her diachronic self-governance. And yet

(c) S makes a choice that blocks her continued diachronic self-governance.

Condition (a) is an historical condition: it matters whether the agent has been engaged in a relevant planned temporally extended activity. And condition (b) would not be satisfied if S's ends develop in a way such that a choice to continue with the planned activity would not cohere with then-synchronic self-governance.

Return to willpower. Suppose you resolve at t1 to have only one beer at the party at t2 while knowing you will at t2 at least initially think it better to have many beers, but also knowing that at t3 you would regret it if you did indeed have many beers at t2. How does DPR apply to this case?

Well, we do not yet know since we do not yet know whether at t2 condition (b) is satisfied. If the standpoint at t2 included the end of diachronic self-governance then perhaps it would favor willpower, and so (b) would be satisfied. Since abandoning the prior resolution would satisfy (c), DPR would then favor, instead, willpower. But we are not yet in a position to suppose that the standpoint at t2 does include this end of diachronic self-governance.

Granted, DPR focuses on a planning agent with the capacity for diachronic self-governance. This capacity includes the capacity for having the end of diachronic self-governance, since that end is involved in central cases of a planning agent's diachronic self-governance. But you could have the capacity for that end and yet not in fact have that end. So in order to understand how DPR would apply to cases of potential willpower we need to reflect further on the status of this end of diachronic self-governance.

But first we need to address a different issue about DPR. Let's assume that the end of diachronic self-governance is present, and so that at least some cases of willpower would cohere with both synchronic and diachronic self-governance and so be favored by DPR. The idea is that in such cases the end of diachronic self-governance re-shifts the agent's standpoint at the time of temptation so that it now favors following though with her prior resolution. So a failure of willpower would now be a failure of synchronic plan rationality. But then we can ask whether we really need a distinctive norm of diachronic plan rationality, a norm along the lines of DPR. Why not simply work with a norm of synchronic plan rationality along the lines of PRSG-P? DPR does have the implication I have emphasized concerning temptation cases: given the end of diachronic self-governance, it can explain why a breakdown in willpower in such cases can sometimes be irrational. But our question now is whether we need DPR for this. Why not just work with synchronic plan rationality, given the way in which the end of diachronic self-governance may shift the agent's standpoint at the time of temptation?

To respond to this challenge we need to consider, as anticipated earlier, a second kind of case that poses a problem of plan stability. These are cases in which one makes a decision in the face of non-comparable temporally extended options and then, in the process of follow through, is later faced with continued non-comparability.[40]

In Sartre's famous example, the young man needs to decide between staying with his mother and fighting with the Free French, where he (plausibly) sees this as a decision between non-comparable values.[41] Suppose he decides in favor of staying with his mother. Later he (sensibly) reconsiders and notes that the non-comparability remains. Is there any rational pressure for him to stick with his earlier decision?

One virtue of DPR is that it articulates a rational pressure in favor of constancy in such cases. Whether or not the young man has the end of diachronic self-governance, each option -- the option of staying with his mother, as well the option of instead fighting with the Free French -- is supported by his now-present standpoint. But what is crucial is that, if the young man does stick with his prior decision to stay with his mother, his intentions over the relevant time will have the inter-connections characteristic of a planning agent's diachronic self-governance. In contrast, if he changes his mind in favor of the Free French then his intentions over the relevant time will not have these inter-connections. So DPR will favor his sticking with his decision.

In this way DPR can help us understand the rational pressure for constancy in such cases of decision in the face of on-going non-comparability.[42] And once we are led in this way to DPR we can note that it promises to contribute to an overall treatment of the rational stability of plans in both such non-comparability cases and, to return to our earlier discussion, temptation cases. But, as noted earlier, the relevant implications of DPR concerning willpower depend on the presence of the end of diachronic self-governance. So we need to return to our question: what is the status of this end of diachronic self-governance?

In discussing synchronic plan rationality I argued that a reflective planning agent would be led to the thought that the best normative explanation of her plan-infused practical thinking draws on the significance of the coherence involved in her synchronic self-governance. I then generalized: the best normative explanation of her plan-infused practical thinking will draw on the significance of the coherence involved in her self-governance, both synchronic and diachronic. This led us to DPR. And this suggests yet a further generalization: we appeal to constitutive conditions of a planning agent's self-governance, where these include, but may not be limited to, coherence conditions. To this we then add our conjecture that diachronic self-governance, at least in (ubiquitous)

cases of temptation, involves the end of diachronic self-governance. We thereby have an argument for a norm that supports an end of diachronic self-governance:

> *Rational End of Diachronic Self-Governance (REDSG)*: If S is a planning agent who is capable of diachronic self-governance then it is pro tanto irrational of S to fail to have an end of diachronic self-governance.[43]

The argument for REDSG involves three ideas. First, there is the general idea, in the spirit of inference to the best normative explanation, that there is rational pressure in favor of satisfying constitutive conditions of self-governance. Second, there is the idea that a planning agent's diachronic self-governance, at least in cases of temptation, involves her end of diachronic self-governance. And third, there is the idea of the ubiquity of forms of temptation as potential destabilizers of planned temporally extended activities.[44]

So, the self-governance-based rationale for norms of plan coherence, both synchronic and diachronic, leads, on further reflection, to a rationale for a rationally supported end. In reflecting on our planning agency we are led to the self-governance strategy in order to support norms of plan coherence, both synchronic and diachronic, and in response to the challenge that these norms express a fetish for mere mental tidiness. This promises a significant commonality of rationale underlying synchronic and diachronic norms. And it leads us to an argument for a norm that supports the end of diachronic self-governance.

REDSG is a weak principle in at least two ways. First, it does not require, even pro tanto, that the end of diachronic self-governance be pre-eminent within the agent's standpoint. Different agents might satisfy REDSG by way of ends of diachronic self-governance that have different relative weights within their standpoints. Second, REDSG does not address the issue of how to respond in cases in which there is a tension between what is called for by diachronic self-governance over different temporal intervals.[45]

Nevertheless, it remains true that REDSG, together with DPR, can sometimes induce rational pressure in favor of willpower; and DPR on its own induces rational pressure for constancy in non-comparability cases. So when we combine DPR with REDSG we arrive, prima facie, at a modest plan conservatism, one that includes but

goes beyond the support of plan stability that is traceable to snowball effects and the rationality of non-reconsideration.

So we have an initial prima facie, self-governance-based case in favor of seeing PRSG-P, PCC, DPR, and REDSG as norms of plan rationality. These norms are, plausibly, central to the basic structure of a planning agent's practical thinking. They track forms of coherence central to self-governance, both synchronic and diachronic, and –- in the case of REDSG – a basic form of support for the coordination of such synchronic and diachronic coherence. So these norms do not merely track mental tidiness. And this self-governance-based case promises to provide a common justificatory framework for this package of norms of synchronic and diachronic plan rationality.

We can now return, as promised, to the question whether for a planning agent with the capacity for self-governance there is a systematically present normative reason that favors conformity to these norms. I take it that if we could defend an affirmative answer to this question we would then be justified in going beyond the cited initial case for these norms and concluding that they are indeed norms of practical rationality for a planning agent. But how can we defend such an affirmative answer?

An initial observation is that a planning agent will have such a reason if she has a reason for her self-governance and that self-governance is attainable. After all, these norms track necessary constitutive features of a planning agent's self-governance. And a reason for self-governance will transmit to a reason of self-governance for those necessary constitutive features if the self-governance is attainable.[46]

But why think that a planning agent with the capacity for self-governance has a normative reason for her self-governance? Given our approach to normative reasons, and given the plausible assumption that self-governance is a human good, she will have this reason for self-governance if, but only if, she has the end of her self-governance. But what is the status of this end?[47]

The key is to proceed in two stages. We note first that, as I have argued,

(1) there is an initial self-governance-based, prima facie case for REDSG.

We then note that

(2) if a planning agent who is capable of diachronic self-governance conforms to REDSG by having the end it supports, she will thereby have a normative reason (a reason of self-governance) to conform to this norm.

And my conjecture is that, given (1), (2) constitutes a sufficiently systematic connection to a supporting normative reason for REDSG to satisfy the reason desideratum, appropriately understood.[48] Granted, this does not show that there is normative reason to conform to REDSG whether or not one does conform to REDSG. But it does show that if one does conform to REDSG by having the end it supports then there is normative reason in support of this conformity. And my conjecture is that, given (1), this suffices for REDSG to satisfy the reason desideratum, appropriately understood. So we can conclude that REDSG is indeed a norm of practical rationality for such a planning agent. So the end of diachronic self-governance is in this way rationally self-sustaining.

A planning agent with the capacity for diachronic self-governance who conforms to REDSG by having the end it supports will have a reason for her diachronic self-governance, and so for the synchronic self-governance that is partly constitutive of that diachronic self-governance. We now note that this reason also supports conformity to PRSG-P, PCC, and DPR, since each of these norms tracks a constitutive element of relevant self-governance. So we can conclude that these norms also satisfy the reason desideratum, appropriately understood. So given the initial self-governance-based case for these norms we can conclude that they too are norms of practical rationality for a planning agent.

Return now to a planning agent who has the capacity for self-governance and is reflecting on basic norms involved in her planning agency. She will see that the best rationale for these norms treats self-governance, both synchronic and diachronic, as the basic consideration that supplements a two-tier pragmatic rationale. She will see that given that she is, as there is reason to be, a planning agent, the application to the particular case of the norms that are central to her planning agency is supported by appeals to the significance of her self-governance. She will thereby be in a position to see the rational dynamics of her planning agency as having a justifying rationale that involves both two-tier-pragmatic and self-governance-based support. In this way her

plan dynamics will make sense to her and be reflectively stable. And that is what we needed to show to defend the planning theory from the challenge posed by the myth theorists.[49]

Stanford University

bratman@stanford.edu

---

[1] This is a substantially revised version of my talk at the Royal Institute of Philosophy in October 2015. A version of this essay was presented at the April 2016 Conference on Practical Reason and Meta-Ethics at the University of Nebraska. The ideas in this essay are developed in more detail in my 2016 Pufendorf Lectures, delivered at Lund University in June 2016. (See http://www.pufendorf.se/sectione195f.html?id=2864)

[2] For this way of formulating a norm of agglomerativity see Gideon Yaffe, 'Trying, Intending and Attempted Crimes', *Philosophical Topics* **32** (2004), 505-32, 510-12.

[3] For a more precise formulation see my *Intention, Plans, and Practical Reason* (Harvard University Press, 1987; reissued CSLI Publications, 1999), 31.

[4] Bratman, *Intention, Plans, and Practical Reason*.

[5] This is the structure of the video games example I discuss in Chapter 8 of *Intention, Plans, and Practical Reason*.

[6] An early version of this challenge is in Hugh McCann, 'Settled Objectives and Rational Constraints', *American Philosophical Quarterly* **28** (1991), 25-36. It is developed further in Joseph Raz, 'The Myth of Instrumental Rationality', *Journal of Ethics and Social Philosophy* **1:1** (2005) and in Niko Kolodny, 'The Myth of Practical Consistency', *European Journal of Philosophy* **16** (2008), 366-402. Talk of 'psychic tidiness' is from Niko Kolodny, 'How Does Coherence Matter?' *Proceedings of the Aristotelian Society* **107** (2007), 229-63, 241. The worry about being 'fetishistic' is from Kolodny, 'Why Be Rational?' *Mind* **114** (2005), 509-63, 547.

[7] Cp. Christine Korsgaard: 'If the problem is that morality might not survive reflection, then the solution is that it might' -- though my concern here is not with morality but with basic plan-theoretic norms. See Christine Korsgaard, *The Sources of Normativity* (Cambridge University Press, 1996), 49.

[8] This is to some extent in the spirit of Niko Kolodny, 'Reply to Bridges', *Mind* **118** (2009), 369–376.

[9] My hope is thereby also to respond further to trenchant challenges to my earlier treatments of these issues in J. David Velleman, 'What Good Is a Will?' in Manuel Vargas and Gideon Yaffe, eds., *Rational and Social Agency: The Philosophy of Michael Bratman* (Oxford University Press, 2014), 83-105, and in Kieran Setiya, 'Intention, Plans, and Ethical Rationalism' in Vargas and Yaffe, eds., *Rational and Social Agency,* 56-82.

[10] See Michael E. Bratman, 'Intention, Belief, Practical, Theoretical', in Simon Robertson, ed., *Spheres of Reason: New Essays on the Philosophy of Normativity* (Oxford University Press, 2009), 29-61; and 'Intention, Belief and Instrumental Rationality', in David Sobel and Steven Wall, eds., *Reasons for Action* (Cambridge: Cambridge University Press, 2009), 13-36.

[11] An idea built into the strategy of creature construction in H.P. Grice, 'Method in Philosophical Psychology (From the Banal to the Bizarre)', *Proceedings and Addresses of the American Philosophical Association* **48** (1974), 23–53. And see my 'Valuing and the Will,' as reprinted in Michael E. Bratman, *Structures of Agency: Essays* (Oxford University Press, 2007), 47-67; and Jennifer Morton, "Toward an Ecological Theory of the Norms of Practical Deliberation," *European Journal of Philosophy* **19** (2011), 561-584.

[12] For helpful discussion of related issues see Luca Ferrero, "Inescapability Revisited', unpublished manuscript, April 2016, section 6.

[13] 'Incoherence and Irrationality,' as reprinted in Donald Davidson, *Problems of Rationality* (Oxford University Press, 2004), 189-98, 196-7.

[14] Niko Kolodny makes this point in his 'The Myth of Practical Consistency', 386.

[15] J.J.C. Smart, 'Extreme and Restricted Utilitarianism', in Philippa Foot, ed.,*Theories of Ethics* (Oxford University Press, 1967), 171-83.

[16] Gilbert Harman, *Change in View: Principles of Reasoning* (Cambridge, MA: MIT Press, 1986), 9. Harman is here focusing on what he calls principles of reasoning, whereas our focus is on principles of plan rationality. We can nevertheless apply the spirit of Harman's comments to our concerns about plan rationality. This is also to some extent in the spirit of Nadeem Hussain's emphasis on a strategy of reflective equilibrium in his 'The Requirements of Rationality', vers 2.4. (unpublished manuscript, Stanford University), though Hussain would not be sympathetic to what I later call the reason desideratum.

[17] For this broad issue see John Broome, *Rationality Through Reasoning* (Wiley Blackwell, 2013), chap. 11. Broome, however, does not work with the model of normative reasons to which I turn in the next paragraph. Further, talk of a "systematically present" reason is mine, not Broome's; and I will have more to say about this idea below.

[18] A classic source of this idea is Bernard Williams, 'Internal and External Reasons', in his *Moral Luck* (Cambridge University Press, 1981), 101-113. My formulation follows, with important adjustment, Mark Schroeder, *Slaves of the Passions* (Oxford University Press, 2007), 59.

[19] *Intention, Plans, and Practical Reason*, 82.

[20] See Richard Holton, *Willing, Wanting, Waiting* (Oxford: Clarendon, 2009). And see my 'Toxin, Temptation, and the Stability of Intention', as reprinted in my *Faces of Intention* (Cambridge University Press, 1999), and my 'Temptation and the Agent's Standpoint', *Inquiry* **57** (2014), 293-310.

[21] For Paul's approach to these matters see her "Doxastic Self-Control," *American Philosophical Quarterly* **52** (2015), 145-158, and her "Diachronic Incontinence is a Problem in Moral Philosophy," *Inquiry* **57** (2014), 337-355. For a discussion of related phenomena see Jennifer Morton and Sarah Paul, "Grit," (unpublished).

[22] See Sergio Tenenbaum and Diana Raffman, 'Vague Projects and the Puzzle of the Self-Torturer', *Ethics* **123** (2012), 86-112, esp. section III.

[23] A related idea is in David Copp, 'The Normativity of Self-Grounded Reason', in his *Morality in a Natural World* (Cambridge: Cambridge University Press, 2007), 309–53, 351. A somewhat related idea is Kenneth Stalzer's thought that a violation of these norms is a breakdown in "self-fidelity". See his *On the Normativity of the Instrumental*

Principle (Ph.D. Thesis, Stanford University, 2004), chap. 5.

[24] J. David Velleman sees the constitutive aim of action as self-intelligibility. However, I take it that on his account this constitutive aim is, in effect, an aim of autonomy. See J. David Velleman, *How We Get Along* (Cambridge University Press, 2009) chapter 5, esp. 131-5. See also 26-27. (In his 'The Possibility of Practical Reason', in his *The Possibility of Practical Reason* (Oxford University Press, 2000), 170-199, 193, Velleman appeals explicitly to a constitutive aim of 'autonomy' and notes the continuity of that appeal with his account in his *Practical Reflection* (Princeton, N.J.: Princeton University Press, 1989.))

[25] Thanks to Jon Barwise and John Perry (in conversation) for this apt term.

[26] 'Normativity, Commitment, and Instrumental Reason', as re-printed in R. Jay Wallace, *Normativity and the Will* (Oxford University Press, 2006), 82-120, 91.

[27] Harry Frankfurt, 'Identification and Wholeheartedness', as reprinted in Harry Frankfurt, *The Importance of What We Care About* (Cambridge University Press, 1988), 159-76, 166. And see also Gary Watson, 'Free Agency', *The Journal of Philosophy* **72** (1975), 205-220, 216

[28] E.g., William Styron, *Sophie's Choice* (Random House, 1979).

[29] For a distinction between local and global rationality see Michael Smith, 'The Structure of Orthonomy', in John Hyman and Helen Steward, eds., *Agency and Action* (Cambridge: Cambridge University Press, 2004), 165-93, 190.

[30] See my 'Three Theories of Self-Governance' as reprinted in my *Structures of Agency* (Oxford University Press, 2007), 222-253, and my "A Planning Theory of Self-Governance: Reply to Franklin," *Philosophical Explorations* (forthcoming). For a deep challenge see Elijah Millgram, 'Segmented Agency', in Vargas and Yaffe, eds., *Rational and Social Agency* 152-89.

[31] See *Intention, Plans, and Practical Reason*, 38-9. Carlos Núñez develops a forceful challenge to a prohibition on intention-belief inconsistency. See Carlos Núñez, *The Will and Normative Judgment* (Stanford University PhD Thesis, 2016).

[32] I note this complexity in my 'Intention, Practical Rationality, and Self-Governance', *Ethics* **119** (2009), 411-443 at note 7. It is the target of an extended discussion in Sam Shpall, 'The Calendar Paradox', *Philosophical Studies* **173** (2016), 801-825.

[33] In asking this question here I continue with my strategy of postponing the question whether there is, systematically, a reason that favors conformity with these norms.

[34] This is the focus of my 'A Planning Agent's Self-Governance Over Time' (unpublished).

[35] Given the hierarchical structure of plans, there can be such interconnections at a higher level despite a breakdown in interconnection at a lower level. If the lower level breakdown in interconnections is grounded in a sensible reassessment of lower-level plans, perhaps in light of new information, the higher-level interconnections will, in the context, normally support a judgment of diachronic self-governance. But in some cases a more complex judgment about the extent of diachronic self-governance will be apt.

[36] Michael E. Bratman, *Shared Agency: A Planning Theory of Acting Together* (Oxford University Press, 2014). I defend this analogy in my 'A Planning Agent's Self-Governance Over Time'.

[37] For ideas broadly in this spirit see Jordon Howard Sobel, 'Useful Intentions', in his *Taking Chances: Essays on Rational Choice* (Cambridge University Press, 1994), 237-254, and Wlodek Rabinowicz, 'To Have One's Cake and Eat It Too: Sequential Choice and Expected-Utility Violations', *Journal of Philosophy* **92** (1995), 586-620.
[38] J. David Velleman, 'The Centered Self', in his *Self to Self* (Cambridge University Press, 2006), 253-283, 272.
[39] A full story would also appeal to the agent's expected regret at giving into temptation, but I put that aside here. See my 'Toxin, Temptation, and the Stability of Intention'.
[40] For a seminal discussion of the case of Abraham and Isaac, see 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 University Press, 2001), 98-120, esp. 114-119. My earlier discussion of such cases is in Michael E. Bratman, 'Time, Rationality, an Self-Governance', *Philosophical Issues* **22** (2012), 73-88.

[41] Jean-Paul Sartre, 'Existentialism Is a Humanism', in *Existentialism from Dostoevsky to Sartre*, edited by Walter Kaufmann. rev. and expanded. (New York: Meridian/Penguin, 1975), 345-69.

[42] This is to disagree with the Sartrean theme of 'the total inefficacy of the past resolution'. *Being and Nothingness* (Hazel Barnes translation) (New York: Washington Square Press, 1984), 70.
[43] I take it that concerns with trivial cases, tragic cases, and preface-analogue cases do not apply here; so we can express REDSG without the appeal to defeasibility that appears in our earlier principles.
[44] This is an argument for rational pressure in favor of a certain end, given capacities for planning agency and diachronic self-governance; it is not an argument for rational pressure in favor of the introduction of a new basic capacity.
[45] A point made by Gideon Yaffe
[46] See Michael E. Bratman, 'Intention, Practical Rationality, and Self-Governance', *Ethics* **119** (2009). I here put aside complexities about this inference. For an insightful overview of related issues, see Benjamin Kiesewetter, 'Instrumental Normativity: In Defense of the Transmission Principle', *Ethics* **125** (2015), 921-46 (though Kiesewetter focuses on the transmission of what he calls the deliberative 'ought', whereas the issue here is the transmission of normative reasons).
[47] I take it that the value of X does not by itself induce even *pro tanto* rational pressure to have the end of X: there are too many goods and, in our finite lives, not enough time.
[48] This is where my talk, in my formulation of the reason desideratum, of a "systematically present" reason is doing important work.
[49] Many thanks to audiences at the Royal Institute of Philosophy, the University of Nebraska, and Lund University, and to participants in my winter 2016 seminar on plan rationality at Stanford University. Special thanks to: Ron Aboodi, Facundo Alonso, Gunnar Björnsson, Olle Blomberg, John Broome, David Copp, Jorah Dannenberg, Luca Ferrero, Amanda Greene, Carlos Núñez, Herlinde Pauer-Studer, Sarah Paul, Björn Petersson, David Plunkett, Johanna Thoma, Han van Wietmarschen, Gideon Yaffe, and an anonymous reviewer.