

## A PLANNING AGENT'S SELF-GOVERNANCE OVER TIME

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

7-10-17

What is it to govern your thought and action not only at a time but also over time? Let's focus the question a bit. We are planning agents. Almost all of our intentional activity is embedded in planned temporally extended activity that is structured by prior partial plans. These plans settle practical issues, pose problems of means and the like, filter solutions to those problems, and guide action. These structures of planning support important forms of cross-temporal and social organization in ways that are compatible with our cognitive and epistemic limits. Given their basic role in cross-temporal organization, we can expect these planning structures to play a central role in diachronic self-governance. So let's ask, more specifically, what is it for a planning agent to govern her thought and action over time? <sup>1</sup>

In the background are two ideas. The first is that self-governance over time involves, in part, responding to certain common threats to important forms of cross-temporal organization and stability. Our account of a planning agent's diachronic self-governance needs to say more about those threats and the response characteristic of diachronic self-governance. Second, we do not want to see such self-governance as, at bottom, the workings of a self who is separate from the psychic economy and stands back and pulls the strings. We want a model of a kind of cross-temporal psychic functioning that ensures a planning agent's relevant self-governance over time. So let's see.

### <1> 1. SELF-GOVERNANCE AT A TIME AND OVER TIME

An initial idea is that for a planning agent to govern her action over time she needs at the least to be synchronically self-governing at relevant times (or small temporal intervals) along the way. But what is such synchronic self-governance? Here we can learn from Harry Frankfurt's

appeal to “where (if anywhere) the person himself stands.”<sup>2</sup> Self-governance involves guidance of thought and action by the agent’s relevant practical standpoint. The standpoint needs to guide choice that is in accord with that standpoint; and the standpoint itself needs to be sufficiently coherent to constitute a clear place where the agent stands on relevant practical issues. In this way we understand synchronic self-governance without an irreducible appeal to a little person in the head who is pulling the strings.

A planning agent’s self-governance over time involves, then, such synchronic self-governance at relevant times or during small temporal intervals along the way.<sup>3</sup> On this model, not all cases of successful self-management over time will constitute such diachronic self-governance. Ulysses has a strategy of self-management over time by tying himself to the mast, thereby blocking his self-governance when the sirens call; but his successful diachronic self-management is not diachronic self-governance.<sup>4</sup>

There is a partial parallel here with rational planning agency quite generally. In rationally planning my actions over a stretch of time, I at least implicitly suppose that, given my plan, when the times of action arrive it will be rationally permissible for me to carry out my plan then. Rational planning for a temporally extended stretch of activity needs to accommodate the supposed rational execution of the plan at times along the way. This is a norm of rational planning. In the case of diachronic self-governance what we have is not a rationality norm but a metaphysical constraint: an individual’s self-governance over time needs to accommodate her self-governance at each relevant time (or relevant temporal interval) along the way. In each case we have a principle of over-time/at-a-time coordination.

I think we should retain such over-time/at-a-time coordination of diachronic and synchronic self-governance.<sup>5</sup> But I also take it that diachronic self-governance is not merely a

concatenation of self-governance of the same person at various times along the way.<sup>6</sup> It involves, as well, appropriate *inter-connections* across time. What inter-connections?

## <1> 2. INTERCONNECTIONS: THE PLAN-THEORETIC MODEL

As I see it, the basic context in which the issue of a planning agent's diachronic self-governance arises is one in which the agent is engaged in plan-shaped temporally extended activity.<sup>7</sup> This suggests that a planning agent's diachronic self-governance involves synchronically self-governed choices at the relevant times along the way of a relevant planned temporally extended activity, where these choices over time are tied together by the inter-connections characteristic of planned temporally extended activity.

What inter-connections are these? Well, the plans that guide such temporally extended activity will settle relevant practical questions and will normally at least implicitly 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 of means and preliminary steps in filling in one's so-far partial plans as time goes by and in ways that, taken together, mesh; 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 of intention and plan that are in the spirit of broadly Lockean models of personal identity over time.<sup>8</sup>

Given these roles of plans in structuring the agent's temporally extended practical thought and action, we can expect that her standpoints at times along the way will be plan-infused.<sup>9</sup> These standpoints will involve both plans for temporally extended activities and relevant general policies, including policies about weights for deliberation.<sup>10</sup> In planned temporally extended activity these plan-infused standpoints will be inter-connected in the cited

ways. And this leads to the proposal that the cross-temporal inter-connections that are characteristic of a planning agent's self-governance over time, given self-governance at times along the way, involve the inter-connections that are characteristic of relevant planned temporally extended activity. This is the *plan-theoretic model* of these inter-connections.

A complexity is that, given the hierarchical structure of plans, these cross-temporal inter-connections can be at different levels. In particular, there can be such interconnections at the level of an overarching plan despite a breakdown in interconnection at a more specific level of sub-plans. Perhaps I continue with my overall plan to earn a law degree but change my sub-plan from one focusing on criminal law to one focusing on tort law. There are relevant cross-temporal continuities of intention at the level of the overall plan; but there is a breakdown in continuity in my sub-plans concerning area of concentration. Such upper-level continuities would help support my diachronic self-governance. And it may well be that I now sensibly believe, perhaps in light of new information, that I have conclusive reason to make the change in my lower-level plans. So sticking with my prior plan in favor of criminal law would be stubbornness, not synchronic self-governance. However, if, in contrast, there were available a way of continuing with my lower-level plans so that there would be self-governance at times along the way and relevant cross-temporal inter-connections at *both* lower and higher levels of the hierarchy, then the plan-theoretic model will say that a breakdown in plan-continuity at the lower level would to some extent diminish the extent of diachronic self-governance.

This plan-theoretic model has an important implication. Consider Sartre's young man. He must choose between staying with his mother and fighting with the Free French.<sup>11</sup> Suppose at t1 he chooses to stay with his mother. At t2, however, he reconsiders and changes his mind and decides to fight with the Free French. Both at t1 and at t2 he sees the conflict as a conflict of

non-comparable values. So it seems that at each time in acting on his choice he is synchronically self-governing, since his choice and action at each time cohere with and are guided by a basic value of his, one that he supposes is not outweighed or overridden in the circumstance. However, in shuffling from his decision at t1 to his decision at t2<sup>12</sup> he is breaking the cross-temporal inter-connections characteristic of planned temporally extended activity. If, in contrast, he had stuck with his initial decision to stay with his mother, his intentions over time would have been connected in these ways. In shuffling from one decision to another, though he is self-governing at each time, he does not satisfy the plan-theoretic model of the interconnections involved in self-governance over time. In contrast, he would satisfy that model were he to stick with his prior decision.

Granted (and to return to the complexity noted earlier), even if he shuffles in this way he might satisfy the plan-theoretic model, of the interconnections involved in self-governance over time, at the level of some persisting, higher-level plan – for example, a plan to give due regard to duties of loyalty. Nevertheless, there was available to him a way of resisting shuffling and continuing with his lower-level plans so that there would be self-governance at times along the way *and* relevant inter-connections at both lower and higher levels of the hierarchy of his plans. So, on the model, his shuffling diminishes the extent to which he is diachronically self-governing.

But why accept this plan-theoretic model of relevant interconnections?

### <1> 3. THE ARGUMENT FROM SETTLING FUNCTION

I begin by considering an argument that is in the spirit of an earlier essay.<sup>13</sup> This argument is not, however, conclusive. So I will turn to a second argument.

This first argument draws from the settling function of intentions and plans, and can be articulated by way of a pair of premises. The first premise concerns a functional role that is characteristic of intention:

- (i) The functional roles of prior intentions and plans include settling relevant practical matters in a way that supports the cross-temporal intention-inter-connections that are characteristic of planned temporally extended activity.

The second premise draws on an insight from Sydney Shoemaker. Shoemaker noted that the psychological continuity that plays a basic role in Lockean theories of personal identity over time “is just the playing out over time of the functional natures of the mental states characteristic of persons.”<sup>14</sup> And an analogous idea seems apt here:

- (ii) The cross-temporal inter-connections that are characteristic of a planning agent’s self-governance over time involve the “playing out over time of the functional natures of the” relevant plan states.

When we put (i) and (ii) together we get

- (iii) The cross-temporal inter-connections that are characteristic of a planning agent’s self-governance over time involve those interconnections that are characteristic of planned temporally extended activity.

And that is the plan-theoretic model of the relevant interconnections.

On reflection, however, one might challenge premise (i).<sup>15</sup> Why not instead say only that the relevant functional role of plan states is to support the cross-temporal intention-interconnections that are characteristic of planned temporally extended activity so long as the agent continues to judge that considerations on balance strongly favor the plan? And Sartre’s

young man does not judge that considerations on balance strongly favor his plan to stay with his mother, though he does judge that these considerations weakly favor that plan.

So if we are going to defend the plan-theoretic model we need to see if there is a further argument in its defense. And I think that there is. The argument I think I see<sup>16</sup> appeals to a version of a parallel that has struck many philosophers – a parallel between the cross-temporal organization of an individual’s activity and inter-personal, social organization.

#### <1> 4. ACTING “TOGETHER” WITH ONESELF AT DIFFERENT TIMES: THE SHARED AGENCY MODEL OF DIACHRONIC SELF-GOVERNANCE

The idea is that in diachronic self-governance one and the same person’s agency at different times is drawn together in ways that to some extent parallel the ways in which different agents are inter-connected when they act together. I do not say that in diachronic self-governance there are, literally, multiple agents acting together. I retain the commonsense thought that there is one agent who acts at different times; and this contrasts with inter-personal shared agency involving multiple agents. Nevertheless, there is a revealing, if partial, parallel between an individual’s diachronic self-governance and inter-personal shared agency. And this parallel helps explain the significance to us of such diachronic self-governance. In diachronic self-governance I am not just governing my relevant actions at each relevant time along the way. In addition, my thought and action at the times along the way are, at least implicitly, interwoven and interconnected in ways that support the helpful metaphor that “we” – that is, me at these different times – are thinking and acting “together”.

Now, others have explored the idea of a bargain between the agent at t1 and the agent at t2.<sup>17</sup> Edward McClennen, for example, seeks “a theory of what constitutes a fair bargain between one’s different, time-defined selves.”<sup>18</sup> The agent at t1 strikes a bargain with his later t2-self, a

bargain that enjoins his t2-self to stick with the agent's decision at t1. The idea of a bargain, however, seems to involve a kind of reciprocal inter-relation that is not possible here. After all, the person-stage at t1 will not be around when the person-stage at t2 acts, and will not be in a position to respond to whether that later person-stage keeps up her end of the bargain.<sup>19</sup>

My proposal that a planning agent's diachronic self-governance involves an intrapersonal analogue of shared agency is not the proposal of a bargain between the person at t1 and the person at t2. Instead, the proposal is that certain structural features of socially shared agency have intra-personal, cross-temporal analogues,<sup>20</sup> and that this helps support the idea that these intra-personal analogues are an aspect of a planning agent's diachronic self-governance.

How then should we understand relevant forms of shared agency? In other work<sup>21</sup> I argue that a basic idea here is that of a shared intention: a shared intention of the participants to J is central to a shared intentional activity of J-ing together. And I argue that this shared intention is, in basic cases, constituted by a web of inter-connected intentions of each of the participants in favor of their J-ing, together with associated cognitive attitudes of those participants. This web of attitudes of the participants is constituted as follows:

(A) The participants each have intentions in favor of J, intentions that interlock in the sense that each intends that they J in part by way of the intentions of each that they J.<sup>22</sup>

(B) Each intends that they J in part by way of mutual responsiveness of each to each in the execution of these intentions of each, and so by way of relevant sub-plans of each that mesh with each other.

(C) Each believes correctly that there is persistence interdependence between the intentions of each that they J.

(D) Each believes that if these intentions do persist they will indeed J.<sup>23</sup>

(E) All this is out in the open among the participants.

When such a structure of inter-connected attitudes leads to the joint J-ing by way of the intended mutual responsiveness and mesh, there is shared intentional action.

And now my proposal is that the cross-temporal, intra-personal “glue” characteristic of a planning agent’s diachronic self-governance involves intra-personal analogues of the inter-personal connections characteristic of shared intention and shared intentional agency. These intra-personal analogues of conditions of shared intention and shared intentional agency support the metaphor that in diachronic self-governance the synchronically self-governing agent at each time is acting “together” with the synchronically self-governing agent – namely, herself -- at other times, in a sense of “together” that comes from our theory of shared agency.<sup>24</sup> And an idea in the background of this *shared agency model* of the cross-temporal interconnections involved in a planning agent’s diachronic self-governance is that it helps explain the significance to us of such diachronic self-governance.<sup>25</sup>

In thinking about a planning agent’s self-governance over time it is tempting to appeal to a kind of cross-temporal narrative unity: there is a “story” that ties together the different elements over time. Depending on what counts as such narrative unity, this idea may apply to many cases. But according to the shared agency model, the more basic cross-temporal unity at work in diachronic self-governance is the unity involved when the agent is acting “together” with herself over time, in a sense of “together” that comes from the theory of shared agency.

#### <1> 5. PLANNED TEMPORALLY EXTENDED AGENCY AND SHARED AGENCY

The next step is to argue that planned temporally extended agency does indeed involve intra-personal, cross-temporal analogues of conditions of socially shared intention. In this way

the shared agency model helps support the plan-theoretic model of the cross-temporal connections that are characteristic of a planning agent's diachronic self-governance.

There will, of course, be important differences between my planned temporally extended agency and our acting together. In particular, our acting together will normally involve two-way causal interactions: what I think and do affects what you think and do, which affects what I think and do. In contrast, there will be a causal asymmetry in the case of my planning agency over time: what I think and do now can causally influence what I think and do later, in a way in which what I think and do later cannot causally influence what I think and do now. (Though my present anticipation of what I will think and do later can of course affect what I think and do now.) So the analogies between planned temporally extended agency and shared intentional agency will at most only be partial. But we can still ask whether these partial analogues, if such there are, shed light on diachronic self-governance.

So let's consider conditions (A)-(E) of shared intention and see what their analogues would be in a case of an individual's planned temporally extended activity. To do this, let's work with an example. Suppose that, given expected time pressures at this evening's interview, I decide at t1 on strategy A for responding to certain anticipated questions at t2. At t1 I form an intention to engage in A at t2. I also fill in this plan a bit: I develop at t1 a partial sub-plan for A-ing at t2. Later, t2 arrives, conditions are as I had at t1 anticipated, and I proceed to fill in my sub-plans further and follow through with A then, at t2. Let's see if there are, in this temporally extended activity, analogues of conditions (A)-(E) of shared intention.

To keep track, let's use *italics* to indicate a relevant analogue condition. Begin with (A), the interlocking condition. And consider the relation between my intention at t1 to A at t2 and my intention at t2 to A at t2. An initial point is that in so intending at t1 I do not think my

intention will simply reach its ghostly hand over time and shape my action at t2.<sup>26</sup> Instead, I will normally suppose that my intention at t1 will persist between t1 and t2, shape associated thought and action along the way, and thereby issue at t2 in my updated intention to A then.<sup>27</sup> It is this updated intention to A that I expect will issue in my A-ing: I certainly do not expect some unrelated intention at t2 to do this work. And this is, at least implicitly, how I intend this process to work its way through. So, at least implicitly, my intention at t1 to A at t2 *interlocks* with my intention at t2 to A then: at t1 I intend, at least implicitly, that I A at t2 by way of my intention at t2 to A at t2; and at t2, at least implicitly, I intend to A then as a way of following through with my intention at t1 to A at t2.<sup>28</sup> So we have this intrapersonal analogue of condition (A) of shared intention.

Consider now the first part of condition (B), the intended mutual responsiveness condition. In the intrapersonal case, at t1 I intend, at least implicitly, that my intention at t2 to A at t2 work its way through to action in ways compatible with my present intentions (at t1) concerning how to A at t2. Further, at t2 I intend, at least implicitly, to A in a way that fits with my intentions at t1 concerning how to A. So there is a kind of *intended mutual responsiveness* (though not mutual causal interaction) between my intentions at t1 and at t2. So, there is this intrapersonal analogue of the first part of condition (B) of shared intention.

Consider now the second part of (B), the intended mesh condition. In the intrapersonal case, at t1 I intend, at least implicitly, that the sub-plans I have so far constructed for A-ing at t2 mesh with the sub-plans I will construct and act on at t2, and at t1 I intend, at least implicitly, that I A at t2 by way of those meshing sub-plans at t2. Further, at t2 I intend, at least implicitly, to A by way of sub-plans at t2 that mesh with the partial sub-plans I had formulated at t1. So at t1 and at t2 I intend, at least implicitly, that my A-ing at t2 go by way of sub-plans at t2 that *mesh*

with my sub-plans at t1. So, there is this intrapersonal analogue of the second part of condition (B) of shared intention.

Consider now condition (C), the condition of correct belief in interdependence. In the intrapersonal case, my intention at t1 to A at t2 depends on my expectation that, if conditions at t2 are as I at t1 expect them to be, I will continue so to intend at t2. Further, at least in some cases, at t2 I know that if I had not intended at t1 to A at t2 then I would not at t2 be intending to A then.<sup>29</sup> So my intention at t1 is dependent on my expectation of my later intention. And my later intention may be accompanied by the knowledge that it depends on my earlier intention. So, there is at least a partial analogue of condition (C).

Consider now condition (D), an expectation of success condition. In the intrapersonal case, at t1 I believe that, given that my intention does persist, I will A at t2. And at t2 I believe that, given that my intention to A has persisted, I will indeed A at t2.<sup>30</sup> So there is, throughout t1-t2, conditional *expectation of success* in A-ing.

Finally, all of these analogues in the intrapersonal case will normally be *out in the open* for me throughout t1-t2. At each stage along the way I will be in a position to know that I have so far satisfied the cited conditions and justifiably to expect that I will continue to do so. So there is an analogue of condition (E) of shared intention.

So in my planned temporally extended activity there are intrapersonal analogues of interlocking intentions, intended mutual responsiveness and mesh in sub-plans, and expectations of success. There is at least a partial analogy with the condition of true beliefs about interdependence. And throughout there can be cognitive accessibility to all these elements, and so an analogue to the out-in-the-open condition. Appeal to these shared-agency-analogue

conditions supports the metaphor that *I at t1* and *I at t2* are acting and thinking *together* in an intra-personal *analogue* of shared intentional activity.

Indeed, there are kinds of shared intentional activity that are even closer to such cases of planned individual activity over time. For example, in the multi-generational shared activity of the building of a cathedral, the inter-relations among the agents of different generations will involve causal asymmetries similar to those in the intra-personal temporally extended case.<sup>31</sup> So in this respect there will be an even closer analogy between such social cases and the cited kind of temporally extended individual planned activity.

Of course, the cited cross-temporal structure in the case of my temporally extended planned activity may not be present in cases in which the world develops between *t1* and *t2* in ways that are counter to my relevant expectations at *t1* about conditions at *t2*. But what matters for present purposes are cases in which these expectations are realized and one does indeed act over time in ways that involve carrying out one's prior plan. And in such cases, planned individual activity over time involves cross-temporal intra-personal connections that are substantially analogous to the inter-personal connections characteristic of shared intention.

Recall now that according to the shared agency model, the cross-temporal connections that are characteristic of a planning agent's diachronic self-governance involve intra-personal analogues of conditions of socially shared intention. What we have just seen is that such intra-personal analogues are provided by the intention inter-connections characteristic of planned temporally extended agency. Further, and as noted earlier, the issue of a planning agent's diachronic self-governance normally arises within the background of planned temporally extended agency. Taken together, these points support the conjecture that the inter-connections characteristic of planned temporally extended agency are central to a planning agent's diachronic

self-governance. And that is the plan-theoretic model of these cross-temporal inter-connections. A striking fact at the bottom of the plan-theoretic model of the cross-temporal, intra-personal glue characteristic of diachronic self-governance is that the cross-temporal structure of planned temporally extended activity involves an intra-personal analogue of acting together. We have sought a deeper rationale for that plan-theoretic model, and now we have it, one that draws on the shared agency model.

#### <1> 6. WILLPOWER AND AN END OF ONE'S DIACHRONIC SELF-GOVERNANCE

Let me turn now to a further question about a planning agent's self-governance over time. Suppose that you know you will be tempted to drink heavily tonight at the party.<sup>32</sup> 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 know, 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 breakdown in self-governance.

What if you abandon your prior resolve and drink heavily at the party, given your present evaluation in favor of that? Well, in that case your intentions and plans over time will not conform to the plan-theoretic model of the connections involved in diachronic self-governance. And that will tend to block your diachronic self-governance.

So in neither case will you be, in a clear way, diachronically self-governing. But it also seems an important commonsense idea that willpower can be a central case of diachronic self-governance. What to say?

We sometimes avoid such problems by, prior to the party, changing the world outside our minds in a way that induces new reasons to stick with our prior resolve. One might, say, make a side bet. This is the snowball effect.<sup>33</sup> Again, we sometimes manage to resist reconsideration of our prior resolve, and this can many times be a good strategy, especially given the normal costs and risks of such reconsideration.<sup>34</sup> Both snowball effects and sensible resistance to reconsideration provide important support to the stability of our temporally extended agency. But cases of temptation and potential willpower seem sometimes to involve something further.<sup>35</sup> After all, it seems that there will be many cases of temptation in which we do in fact reconsider our prior resolution and we are not protected by some prior side bet or other snowball effect.<sup>36</sup>

A thought here is that we have available a form of theoretical reasoning that can support sticking with our prior resolve. After all, in such a case you will at the time of temptation know that your then-present evaluation diverges from your earlier and expected-later evaluation. This might lead you to adjust your then-present evaluation to be more in line with your different evaluations at these different times. This might be an aspect of a form of theoretical reasoning that aims at smoothing out differences in judgment across time.<sup>37</sup> Could we appeal to such theoretical reasoning to explain how willpower can comport with self-governance?

Well, at the time of temptation you will think that your earlier and later evaluations are mistaken, and so wonder why you should change your present evaluation in order to smooth out these cross-temporal differences in evaluation. And it is not clear why you should have any intrinsic concern with such diachronic theoretical constancy --- why such a concern would not simply be an undefended concern with cross-temporal theoretical tidiness.<sup>38</sup> But in any case, what is central for present purposes is that such theoretical thinking is not directly concerned with or responsive to the presence of the prior resolution. Yet what we are trying to explore is

the potential role of that prior resolution in diachronic self-governance. A related point is that we seek an account of willpower that also applies to resistance to shuffling in cases like that of Sartre's young man. But in such cases of persisting non-comparability there is already constancy of evaluation over time: the problem is that this constant evaluation underdetermines what to do. So appeal to cross-temporal theoretical smoothing will not explain what blocks shuffling.

Perhaps we should instead challenge the assumption that in the conflict, at the time of temptation, between prior resolution and present evaluation it is the present evaluation that speaks for the agent and shapes her then-present standpoint. Perhaps we should instead say that in some cases it is the prior resolution that speaks for the agent. That is why acting in accord with that prior resolution at the time of temptation can cohere with synchronic self-governance.<sup>39</sup> Further, given that the evaluation at the time of action is made in light of what matters to the agent, and given that what matters to the agent is shaped by her standpoint, this impact of the prior resolution on the agent's standpoint can then ground a re-shift in the evaluation so that it favors sticking with one drink.<sup>40</sup>

But given just these two elements – prior resolution and present evaluation – and given the normal status of the agent's evaluation, it is going to be difficult to explain why the prior resolution trumps the evaluation in shaping where the agent stands.<sup>41</sup> A more straightforward strategy would be instead to appeal explicitly to a third element, one that can explain this significance of the prior resolution to present standpoint. And a natural idea here would be to appeal to an end in the agent's standpoint that favors follow through with the prior resolution and thereby potentially supports both a shift in standpoint and a re-shift in evaluation, both in the direction of willpower. In playing this dual practical role such an end could help explain how willpower can cohere with both synchronic and diachronic self-governance.

What is it to have an end? Well, to have X as an end is, roughly, to have a non-instrumental concern in favor of X. This is not the same as having an intention in favor of X. First, not all intentions are ends in this sense, since some intentions are solely instrumental. And, second, even for a planning agent not all ends are intentions, since – in contrast with intentions -- it is common to have ends one knows are not co-realizable.

The proposal, then, is that a relevant end, in this sense of end, can, together with the cited prior resolution, play a dual role of supporting both a re-shift in evaluation and a re-shift in standpoint. But, how precisely should we conceptualize this proposed dual role of such an end? Should we say that this end, together with the prior resolution, directly supports a re-shift of evaluation in favor of willpower, and that given this re-shifted evaluation there is a corresponding shift in what the agent's standpoint supports? Or should we say that this end, together with the prior resolution, directly supports a shift in standpoint, and that given this shift in standpoint there can be a corresponding shift in evaluation (since this evaluation is made in light of what matters to the agent, and what matters to the agent is shaped by her standpoint)?

It seems to me that the second proposal – one that prioritizes the impact of the end on the agent's standpoint -- fits best with our underlying model of self-governance. The idea is that what is fundamental is what is favored by the agent's standpoint. This shapes what matters to the agent. And the relevant evaluations are evaluations in light of what matters to the agent. What was wrong with the earlier, simple proposal that it is the prior resolution, rather than the present evaluation, that dominantly speaks for the agent was not that it focused primarily on the impact of these different elements on the agent's standpoint. The problem, rather, was that, in the absence of the sort of third element to which we are now appealing, it seemed difficult to defend this priority of the prior resolution in shaping the agent's standpoint. Once we introduce

this third element, however, we potentially have a straightforward explanation of the impact of the resolution on the agent's standpoint. And we thereby potentially have a straightforward explanation of its impact on the agent's relevant evaluation.

But what end is this? One idea, due to Jordan Howard Sobel, would simply be that the agent as a matter of fact "puts a premium on steadfastness".<sup>42</sup> In a related proposal, Wlodek Rabinowicz highlights the idea that one's preferences may well be "influenced by [one's] previously chosen plan of action".<sup>43</sup> In each case the idea would be that the agent has an end of "steadfastness", or of fitting with a "previously chosen plan", and this end, together with the prior resolve, explains how her standpoint at the time of temptation gives independent significance to relevant cross-temporal inter-connections in a way that potentially shifts her standpoint to favor willpower and thereby supports a corresponding re-shifting of evaluation in favor of willpower. So an end along the lines of Sobel's and Rabinowicz's discussions would potentially play the cited dual role of shaping standpoint and thereby shaping evaluation.

But here the worry is that in appealing to such an end we are appealing simply to a concern in favor of mere cross-temporal psychic tidiness. And we do not want a central element in diachronic self-governance to involve such a brute concern with tidiness.<sup>44</sup>

What we need, then, is an end that, in tandem with the prior resolve, can play these two coordinated roles by favoring something that involves but goes beyond cross-temporal coherence. Here we might consider a proposal by Thomas Kelly, in the context of a discussion of sunk costs, that it can be sensible to have an end of redeeming one's own earlier actions, of preventing them from having been in vain.<sup>45</sup> And this end is not simply a concern with cross-temporal mental tidiness. But even if we agree with Kelly that such a backward-looking end can sometimes be sensible, its application will not be sufficiently general to solve our problem about

willpower. Perhaps one's earlier resolve not to drink heavily at the party has already issued in actions that have themselves had significant costs, and so an end of preventing those costly actions from being in vain might support follow through at the time of the party. But this is a special case. One may simply have decided in the morning not to drink heavily in the evening, and now the time -- and the temptation -- has arrived. There has been no significant prior, costly action that might be the target of redemption.

What we need is a more generally applicable end, but one that is not simply a concern with cross-temporal mental tidiness. And a natural proposal, based on a kind of inference to the best explanation, would appeal to a (perhaps implicit) end of one's diachronic self-governance itself -- where that diachronic self-governance is understood along the lines we are currently developing.<sup>46</sup> This is not merely an end of diachronic coherence, though diachronic self-governance involves cross-temporal coherence. This end of one's diachronic self-governance 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, I have argued, an element in diachronic self-governance. In this way this end of one's diachronic self-governance would be poised to help stabilize the agent's temporally extended, planned activities in the face of anticipated temptation.

Further, such cases of temptation -- where these sometimes take the form of temptations to procrastinate -- pervade our lives.<sup>47</sup> If the end of diachronic self-governance is part of the structure of a planning agent's response, in diachronic self-governance, to such pervasive sources of instability, we can plausibly see that end as a central element in a planning agent's self-governance, one that stands guard against such instability. 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 relevant times along the way, and the coherence involved in relevant cross-temporal continuities and inter-connections of intentions over time. The proposal is that this coordination of these two forms of coherence is normally supported by standpoints that include the end of diachronic self-governance, and that the presence of this end stands guard to support this coordination in the face of pervasive potential threats to the needed stability over time.

So by modeling a planning agent's diachronic self-governance as involving this end of her diachronic self-governance we provide for the possibility of diachronically self-governing willpower. We help explain the potential robustness of diachronic self-governance in the face of certain characteristic threats of instability. And we do this by appeal to an end that – in contrast with mere mental tidiness – is, plausibly, worth wanting.

Is there an analogous argument that synchronic self-governance involves an end of synchronic self-governance? Well, what is needed for synchronic self-governance is guidance by the substantive concerns that constitute the agent's then-present standpoint. And it is not clear why these substantive concerns must include a concern specifically with self-governance. In contrast, when we turn to diachronic self-governance we face a distinctive, time-induced issue of coordination between present standpoint and cross-temporal continuity. The end of one's diachronic self-governance is a response to that issue. If this is right, there is an asymmetry here. However, insofar as synchronic self-governance is an element in diachronic self-governance, there will be an indirect argument for the need for an end of synchronic self-governance.

The next step is to see more precisely how the presence of this end of diachronic self-governance would help make possible a form of willpower that coheres with both synchronic and diachronic self-governance.

## <1> 7. DIACHRONICALIZED STANDPOINTS

The first step is to note that the relevant end will not simply be an end of maximizing the amount of one's diachronic self-governance – an end that would argue against taking a nap. It will, rather, be an end in favor of one's diachronic self-governance given that certain pre-conditions are met. In particular, this end will come to bear at a given time if the agent is engaged in a relevant planned temporally extended activity and her standpoint at that time – a standpoint that includes this very end -- would support a choice that also satisfies the intention-interconnection conditions of diachronic self-governance. Since this conditional end concerns the role of a standpoint that itself includes this end, this is a reflexive (and conditional) end in favor of diachronic self-governance.

We can now distinguish two cases. In a simple case the support of the standpoint at that time for a choice that coheres with one's diachronic self-governance does not depend on the presence in that standpoint of the end of that diachronic self-governance. This may happen in a case of potential shuffling given on-going non-comparability in which, even in the absence of the end of diachronic self-governance, the standpoint supports – albeit, weakly – refraining from such shuffling. In a second, more complex case, the support of the standpoint for a choice that coheres with one's diachronic self-governance depends on the way this end in favor of that diachronic self-governance – an end that is itself in that standpoint – favors that choice in part because that choice would satisfy the diachronic inter-connection conditions on diachronic self-governance. As we have begun to see, this may happen in a case of potential willpower in the face of temptation. In either case, and this is the central idea, the agent's standpoint at that time – one that includes this conditional, reflexive end in favor of one's diachronic self-governance --

would support the joint satisfaction of the conditions of both synchronic and diachronic self-governance.

Let's see in more detail how this works in a temptation case. In such a case the agent's conditional and reflexive end of her diachronic self-governance would support willpower if that end were brought to bear on that willpower; and, given her planned, on-going temporally extended activity, it will be brought to bear if her standpoint at the time of temptation, *supplemented by the support this end would provide*, would support that willpower. If the agent's conditional and reflexive end of diachronic self-governance is sufficiently important to her to make it true that her standpoint at the time of temptation, supplemented by the support this end would provide, would support that willpower, then the agent satisfies the relevant pre-conditions, and her conditional end in favor of her diachronic self-governance does indeed support sticking with her prior resolve.

A complication is that this end in favor of her diachronic self-governance may not always be sufficiently important to the agent to re-shift her standpoint in this way in a temptation case. After all, that standpoint involves a complex web of elements, only one of which is this end. Other ends can in a given case trump the potential impact of this conditional end in favor of diachronic self-governance.<sup>48</sup> When other ends do trump they block relevant diachronic self-governance (though they need not block synchronic self-governance), and so they block the satisfaction of the relevant pre-condition. This is what happens when the agent decisively rejects his earlier plan, despite her end in favor of her diachronic self-governance. But sometimes in a temptation case this end in favor of diachronic self-governance would have sufficient priority to support sticking with the prior resolve in a way that is sufficient to re-shift the agent's standpoint so that it now favors sticking with her prior resolve. In such a case the agent's sticking with her

resolve can cohere with both synchronic and diachronic self-governance. The agent will thereby be in a position to re-shift her evaluation concerning further drinking: she can now favor abstaining since that would comport with her self-governance, both at the time and over time, and so – given her end of her diachronic self-governance -- with where she now stands. We thereby explain how willpower can sometimes (though not always) cohere with self-governance.

So a way to insure that at least some cases of willpower involve synchronic self-governance at the time of resisting temptation is to appeal to the cited conditional end in favor of one's diachronic self-governance. As noted, this end will be reflexive in the sense that it supports a choice insofar as that choice would comport with both diachronic and synchronic self-governance *given this very end* (and its impact on the agent's standpoint, and so on synchronic self-governance, and so on diachronic self-governance).

The next point is that once this conditional, reflexive end is on board it will also favor constancy of decision in the face of non-comparability. Sartre's young man, having chosen to stay with his mother, comes to a time when he reconsiders. At this time he satisfies the relevant pre-condition: he has a standpoint that (weakly) supports the choice to continue helping his mother, where that choice would thereby comport with synchronic self-governance and, given the history, with diachronic self-governance. So, this choice to continue with his mother – in contrast with a shuffled choice instead to fight with the Free French -- would be supported by his conditional, reflexive end in favor of his diachronic self-governance.

Granted, there is no guarantee that this will result in his standpoint *strongly* favoring the option previously chosen. A basic feature of non-comparability is that if A and B are non-comparable, it is possible to add into the mix a further consideration, C, in favor of A, and yet it still be true that A together with C remains non-comparable with B.<sup>49</sup> Nevertheless, we have an

explanation of why, even if the non-comparability remains, shuffling in such a case is a breakdown in diachronic self-governance, whereas resistance to shuffling can comport with diachronic self-governance and so be supported by the end of diachronic self-governance.

In supposing that the agent's standpoint includes the cited conditional, reflexive end in favor of her diachronic self-governance I am supposing that this standpoint is, as I will say, *diachronicalized*: it includes an element that can sometimes accord significance to relevant plan-infused cross-temporal interconnections, interconnections that are not merely cross-temporal mental tidiness. And the idea is that a planning agent's diachronic self-governance in the case of willpower involves a standpoint that is diachronicalized in this way, and that such a standpoint will also support a planning agent's diachronic self-governance in resisting brute shuffling.

#### <1> 8. LATER REFLECTION AND PLAN-INDUCED TEMPORAL FOOTPRINT

We need one further idea. Suppose that in Gregory Kavka's toxin case you decide at  $t_1$  to drink a disgusting toxin at  $t_3$ .<sup>50</sup> You make this decision at  $t_1$  because you know that given that decision you will get a great reward at  $t_2$ . You also know throughout that your reward at  $t_2$  depends on your decision at  $t_1$  but does not depend on your following through with that decision at  $t_3$ . You decide at  $t_1$  to drink at  $t_3$ , you get the reward at  $t_2$ , and  $t_3$  arrives. At  $t_3$  you are faced with a choice of whether to stick with your intention to drink the toxin at  $t_3$ . If you stick with it, your intentions during  $t_1$ - $t_3$  will satisfy the shared-agency-analogue conditions with respect to this time interval. But won't you nevertheless be acting contrary to your standpoint at  $t_3$ ? Well, on our way to a model of a planning agent's diachronic self-governance we have supposed that you have the conditional, reflexive end of your diachronic self-governance. So we need to ask whether, given this end, your standpoint at  $t_3$  may turn out to support drinking the toxin. If it

does then your drinking the toxin at t3 may fully comport with both your synchronic and your diachronic self-governance. But that seems wrong.

In response we might try simply to say that in this toxin case at t3 your end of your diachronic self-governance will not outweigh your end of not getting very sick. But in discussing the temptation case we supposed that your end of your diachronic self-governance might sometimes in effect outweigh your end of drinking more wine. And it is not clear how we could defend in a principled way this purported, systematic difference between the cases.

I think there is a structural issue here, one distinct from the question of the relative weights an agent assigns to avoiding the discomfort of toxin-drinking in comparison with diachronic self-governance. We can articulate this structural issue by considering your relevant reflections at a later time, t4.<sup>51</sup> In particular, we consider your reflections at t4, on the basis of what you at t4 know were your relevant non-evaluative beliefs at t3, on your decision at t3. In the temptation case, if you did stick with your prior resolve to resist the temptation, such reflections at t4 would, we may assume, not issue in regret about your decision at t3. In the toxin case, in contrast, if you did at t3 stick with your prior resolve to drink toxin, such reflections *would*, we may assume, issue in a kind of regret at t4 about your decision at t3. Drinking toxin is, after all, awful. In the toxin case – in contrast with the temptation case – sticking at t3 with your prior intention would be something you would regret at t4 when reflecting in the cited way on that decision.

This suggests that sticking with your intention at t3 to drink the toxin does not fit into the kind of cross-temporal structure of plan-infused attitudes that is an element of diachronic self-governance. And this is a contrast with the kind of temptation case at issue here. The end of one's diachronic self-governance conditionally supports sticking with one's resolve in a temptation case in part because sticking with one's resolve fits together, in shared-agency-

analogue ways, not just with one's earlier resolve but also with one's relevant, anticipated later attitude concerning one's exercise of willpower. In sticking with one's resolve one is acting "together" with oneself at both earlier and later times. It is this relevant anticipated later regret at abandoning one's resolve that distinguishes this temptation case from the toxin case in which what is anticipated is, rather, relevant later regret at having followed through with one's intention to drink the toxin. In drinking the toxin one would not be acting "together" with oneself at the later time,  $t_4$ . This is what allows us to resist the idea that drinking the toxin would be an element in diachronic self-governance and so potentially be supported by the conditional, reflexive end of one's diachronic self-governance.

But why does this yet later time matter? The answer seems to be that appeal to such later times is at least implicitly built into the plans that frame the relevant temporally extended activity. The planned temporal shape of the activity will frequently extend to relevant later times – to, as I once said, plan's end.<sup>52</sup> In the temptation case we are thinking of the planned project of resisting temptation as extending beyond the moment of expected temptation and including one's relevant later reflections on earlier follow through. And in the toxin case we are thinking of the plan for getting the money as extending beyond the time of drinking the toxin and including one's relevant later reflections on earlier follow through. In each case, diachronic self-governance with respect to the planned temporally extended activity needs to involve relevant shared-agency-analogue plan-theoretic continuities that extend through to plan's end. And that is why the conditional, reflexive end of one's diachronic self-governance, while it will sometimes support willpower, will not (on current assumptions) support drinking the toxin. It will not support drinking the toxin because following through with the intention to drink will not fit

appropriately with one's relevant, anticipated regret, at plan's end, concerning such follow through.

#### <1> 9. A PLANNING AGENT'S DIACHRONIC SELF-GOVERNANCE

We now have in place the building blocks for a model of a robust form of a planning agent's self-governance over time:

First, the agent is engaged in *planned temporally extended activity*. Given the hierarchical structure of plans, this planned temporally extended activity can be shaped by a higher-level plan that remains in place even as lower level sub-plans are adjusted. And these plans specify a *temporal footprint*, one that will commonly include one's relevant later responses to one's then-earlier efforts.

Second, there is synchronic *self-governance at (during) relevant times (small temporal intervals) along the way* in the execution of the plan that frames the relevant temporally extended activity. This synchronic self-governance involves coherence of standpoint and coherence of choice with coherent standpoint.

Third, there are appropriate *cross-temporal interconnections* between relevant plan-infused attitudes. These inter-connections involve cross-temporal inter-connections of intention that are characteristic of planned temporally extended activity, all in the context of self-governance at times along the way. These plan-theoretic inter-connections induce associated shared-agency-analogue inter-connections. And these shared-agency-analogues support the metaphor that in the temporally extended activity involved in diachronic self-governance the agent at different times along the way is "acting together" with herself at the other relevant times along the way.

As noted, a complexity is that there can be these inter-connections at a higher level of the hierarchy of plans despite breakdowns in inter-connection at a lower level of sub-plans. Such higher-level interconnections can, together with other relevant elements, sometimes suffice for diachronic self-governance. However, if the breakdown at the lower level could have been avoided in a way that cohered with diachronic (and so, synchronic) self-governance at that level, then that lower-level breakdown diminishes the extent of diachronic self-governance even given the higher-level interconnections.

Fourth, the standpoints at times along the way are *diachronicalized*: they include a conditional, reflexive end in favor of one's diachronic self-governance. A pre-condition of this end is that the agent is engaged in a relevant, planned temporally extended activity, and his diachronicalized standpoint at the time at issue would support the joint satisfaction of conditions of both synchronic and diachronic self-governance. This conditional, reflexive end to some extent helps coordinate the coherent standpoints at each relevant time along the way with the cross-temporal inter-connections involved in diachronic self-governance. It does this by sometimes inducing within the standpoint at a given time relevant significance of the cited connections across time. And the time frame that is relevant to this end of diachronic self-governance is specified by the plan that frames the underlying temporally extended activity.

In short, this model of a planning agent's self governance over time highlights: (1) planned temporally extended activity, and an associated time frame; (2) self-governance at (during) relevant times along the way of that activity; (3) plan-infused cross-temporal inter-connections that have a structure that is to some extent analogous with the inter-personal structure of inter-locking intentions of individuals in shared agency; and (4) diachronicalized

standpoints at (during) relevant times along the way. These diachronicalised standpoints help support the coordination between (2) and (3) in response to characteristic threats to stability.

Such diachronic self-governance on the part of a planning agent is responsive to common threats of instability within her temporally extended agency, namely: threats of potential shuffling and of temptation. These responses to these threats involve diachronicalised standpoints and go beyond snowball effects and sensible non-reconsideration. But these responses need not involve an implausible rigidity. We highlight the twin ideas that a planning agent's diachronic self-governance normally involves the end of her diachronic self-governance and can involve the stability of prior intentions in cases of resisting shuffling and in some cases of willpower. But we also acknowledge the Sartre-inspired thought that in some cases "the prior project collapses into the past in the light of a new project which rises on its ruins."<sup>53</sup>

---

<sup>1</sup> This essay develops ideas about a planning agent's diachronic self-governance that are sketched in "Rational Planning Agency," *Royal Institute of Philosophy Supplement* (2017) 80 [this volume, essay 10]. I offer a brief overview of the ideas in both of these essays in my "Plan Rationality," in Ruth Chang and Kurt Sylvan, eds., *The Routledge Handbook of Practical Reason* (forthcoming). Many of these ideas are drawn from my Pufendorf Lectures, delivered at Lund University in June 2016. I presented earlier versions of the present essay at the August 2016 Workshop on "Time in Action" at Oslo University; as part of my Franz Brentano Lectures on Practical Philosophy at the University of Vienna in April, 2017; and at the Copenhagen workshop on Shared and Temporally Extended Agency in April, 2017. Many thanks to the audiences on these occasions and to Philip Pettit, Sergio Tenenbaum, Johanna Thoma, and Gideon Yaffe,

---

<sup>2</sup> 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

<sup>3</sup> In saying this I am supposing that synchronic self-governance can involve self-governance during a relevant, small temporal interval. (This is in the spirit of Sergio Tenenbaum’s thought that “a ‘synchronic’ norm is not necessarily a time-slice norm.” See Sergio Tenenbaum “Reconsidering Intentions,” *Nous* (2016): 14.) Can we say more about such intervals? My tentative proposal is that we look for intervals within which relevant basic concerns and plans are constant, there is not reconsideration of relevant plans, and there are not relevant action plans that specify different stages to be carried out at different times within that interval. Given that we are normally involved in planned temporally extended activities within which our plans specify different stages over time, we can expect these intervals normally to be small. And in any case, once there is reconsideration of a relevant plan we can no longer continue to understand the process as simply one of synchronic self-governance. Nevertheless, it remains possible that one fails to act in accord with one’s standpoint during a relevant interval even if we cannot specify a specific moment in that interval at which this failure occurred. (These comments were aided by correspondence with Sergio Tenenbaum, though they are in the service of an approach to diachronic self-governance that I do not think he would endorse. And see Tenenbaum “Reconsidering Intentions,” at 21-22.)

<sup>4</sup> This paragraph responds to a query from Gideon Yaffe.

<sup>5</sup> Agnes Callard (in conversation) and Jennifer Morton (in correspondence) have pointed to cases in which a breakdown in synchronic self-governance at  $t$  results in downstream changes that then support later, associated diachronic self-governance. As I understand it, such later diachronic

---

self-governance begins at a time later than  $t$ . But I agree that we might – as it were, by courtesy – speak of the earlier breakdown in synchronic self-governance at  $t$  as the time at which, in an extended sense, the diachronic self-governance begins.

<sup>6</sup> See my “Time, Rationality, and Self-Governance,” *Philosophical Issues* 22 (2012): 73-88, at 81 [this volume, essay 6, at 00].

<sup>7</sup> Granted, even for a planning agent there may be cases of intentional agency that are not embedded within such planned temporally extended activity: spontaneously scratching an itch, perhaps. But in such cases of spontaneous activity the question of whether the agent is diachronically self-governing does not arise in a clear way.

<sup>8</sup> For this last idea see my “Reflection, Planning and Temporally Extended Agency” as reprinted in my *Structures of Agency* (2007) New York: Oxford University Press, at 28-33. And see Gideon Yaffe, *Liberty Worth the Name: Locke on Free Agency* (Princeton University Press, 2000), chap. 3. This paragraph draws from my “Rational Planning Agency”.

<sup>9</sup> See my ‘Three Theories of Self-Governance’ as reprinted in my *Structures of Agency* (Oxford University Press, 2007), 222-253, and my “Rational Planning Agency”.

<sup>10</sup> See my “Three Theories of Self-Governance” at 239, where I call these “self-governing policies”. For an overview of my approach to a planning agent’s self-governance at a time see my “A Planning Theory of Self-Governance: Reply to Franklin,” *Philosophical Explorations* 20 (2017): 15-20.

<sup>11</sup> 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. The problem I note in this paragraph owes to 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. .

<sup>12</sup> I owe the terminology of “shuffling” to Richard Kraut (in conversation). In my “Time, Rationality, and Self-Governance” I call this brute shuffling.

<sup>13</sup> “Time, Rationality, and Self-Governance”.

<sup>14</sup> Sydney Shoemaker, “Personal Identity,” in Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Basil Blackwell, 1984) at 95.

<sup>15</sup> See Luca Ferrero, “Diachronic Constraints of Practical Rationality,” *Philosophical Issues* 22 (2012): 144-164, at 160.

<sup>16</sup> An argument that goes beyond my earlier discussion of a planning agent’s self-governance over time in my “Time, Rationality, and Self-Governance”.

<sup>17</sup> Edward McClennen, *Rationality and Dynamic Choice: Foundational Explorations* (Cambridge: Cambridge University Press, 1990); Edward McClennen, “Rationality and Rules” in Peter A. Danielson, ed., *Modeling Rationality, Morality, and Evolution* (Oxford University Press, 1998), 13-40; George Ainslie, *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person* (Cambridge University Press, 1992).

<sup>18</sup> “Rationality and Rules,” 25.

<sup>19</sup> I make a related point in response to a similar idea from George Ainslie in my “Planning and Temptation,” as reprinted in my *Faces of Intention* (Cambridge University Press, 1999), 48. As I note there, Frank Döring makes a related point in response to McClennen.

<sup>20</sup> Carol Rovane explores a related parallel between “long-term activities” and “joint activities” in *The Bounds of Agency: An Essay in Revisionary Metaphysics* (Princeton University Press, 1998), pp. 144 ff.

---

<sup>21</sup> Michael E. Bratman, *Shared Agency: A Planning Theory of Acting Together* (Oxford University Press, 2014)

<sup>22</sup> Since these intentions, as it were, interlock with themselves they involve a kind of reflexivity.

<sup>23</sup> As I indicate in *Shared Agency* at pp. 76-77, such a belief condition is, strictly speaking, too strong if what we are seeking are necessary conditions for shared intention. But since what we are seeking is, rather, sufficient conditions for robust forms of shared intention, it is reasonable to appeal here to such a belief condition. (And see below, note 30.) More generally, in *Shared Agency* I do not claim that the conditions highlighted there (and cited here) are strictly necessary for shared intention or shared intentional action. I leave open the possibility of other, perhaps weaker, forms of shared agency. However, my conjecture here is that these strong conditions of shared agency help give us an apt model of a planning agent's diachronic self-governance

<sup>24</sup> In my *Shared Agency* I discuss -- and note important complexities concerning -- the mirror image of this idea, namely the idea that shared agency involves "quasi-Lockean" interconnections between relevant participants. See pp. 97-97 and 128.

<sup>25</sup> A further, interesting question that I will not try to pursue here is whether, and to what extent, this shared agency model can help us understand the legal doctrine of *stare decisis*.

<sup>26</sup> *Intention, Plans, and Practical Reason*, p. 5.

<sup>27</sup> While my intention persists from t1 to t2 it will also normally be filled in with further sub-plans. The persistence of intention that is characteristic of planned temporally extended agency is compatible with filling in the partial, hierarchical structure that is normally involved in prior intention. (I am responding here to an inquiry from Thomas Smith.)

<sup>28</sup> The idea that relevant contents of intentions can be tacit or implicit is also an aspect of the background view of shared intention. See *Shared Agency*, 104-5. I also note there the possibility

---

that “certain less demanding social psychological phenomena might in certain cases to some extent functionally substitute for these more demanding attitudes of each” (105). And we can allow for an analogous possibility in the intra-personal case. So in the shared case and in the intra-personal case there are analogous possibilities of non-explicit but implicit content and less demanding functional substitutes.

<sup>29</sup> See my remarks below about the snowball effect.

<sup>30</sup> Here I am mirroring the belief conditions cited in the above model of shared intention, where that model aimed at sufficient conditions for robust shared intention. As noted above (note 23), we could weaken those belief conditions somewhat, and then those weaker belief conditions would be reflected in our account of relevant shared agency analogues. But to keep our discussion manageable I put this complexity aside here.

<sup>31</sup> The example owes to Seamus Miller. I discuss it in *Shared Agency*, 100.

<sup>32</sup> 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), my “Temptation and the Agent’s Standpoint,” *Inquiry* **57** (2014), 293-310 [this volume, essay 7], and my “Rational Planning Agency”. In this paragraph I draw from that last essay.

<sup>33</sup> See my *Intention, Plans, and Practical Reason*, 82, and my “Agency, Time, and Sociality.” *Proceedings and Addresses of the American Philosophical Association* 84, no. 2 (2010): 7–26, 10 [this volume, essay 5, at 00.] In a standard case of a snowball effect, the new reason to stick with one’s prior intention is induced by an initial stage in the execution of that intention (e.g., buying a nonrefundable airplane ticket), a stage whose aim is not that of inducing such a reason.

---

Here I extend the idea of a snowball effect to include cases in which the new reason is induced by activity (e.g., making a side bet) whose aim is to induce that reason.

<sup>34</sup> This is a general theme in my *Intention, Plans, and Practical Reason*.

<sup>35</sup> Here I am disagreeing with Holton's approach in his *Willing, Wanting, Waiting*.

<sup>36</sup> Sarah Paul, "Review of Richard Holton's *Willing, Wanting, Waiting*," *Mind* 120 (2011): 889-92.

<sup>37</sup> There is a parallel here with Bas van Fraassen's "Reflection" principle in his "Belief and the Will," *Journal of Philosophy* 81 (1984): 235-56. Related ideas are in Sarah Paul, "Doxastic Self-Control," *American Philosophical Quarterly* 52 (2015): 145-58, and Arthur Lau, "Temptation, Resolution, and Epistemic Self-Trust" (unpublished manuscript, Stanford University 2013). I discuss these ideas further in "Temptation and the Agent's Standpoint" section 4.

<sup>38</sup> This parallels concerns that have been raised about certain apparently basic norms of practical rationality. See below note 44.

<sup>39</sup> A related idea is in my "Temptation Revisited" as reprinted in Michael E. Bratman, *Structures of Agency* (Oxford University Press, 2007), 264-74.

<sup>40</sup> A related idea is in my "Temptation and the Agent's Standpoint".

<sup>41</sup> I made some efforts in this direction in "Temptation Revisited" at 271-4 and 278; but I now think that these efforts will not apply in a sufficiently general way to cases of potential willpower.

<sup>42</sup> Jordon Howard Sobel, "Useful Intentions," in his *Taking Chances: Essays on Rational Choice* (Cambridge University Press, 1994): 237-254, at 249.

<sup>43</sup> In exploring the idea of preferences that are "influenced by the previously chosen plan of action" Rabinowicz notes the parallel with Sobel's paper. Such preferences are at the heart of

---

Rabinowicz's model of "wise choice". But, as I go on to note in the text, we can ask whether giving such significance to continuity with the past is an undefended concern with mere cross-temporal mental tidiness. To answer we need to embed such cross-temporal continuity within a larger framework. And that is what the appeal to diachronic self-governance – an appeal to which I turn below -- tries to do.

That said, both Sobel and Rabinowicz highlight an idea that is also part of the model I will be discussing, namely: that however we understand this concern for cross-temporal continuity, it is one concern among others and can be overridden in a particular case. Thus Rabinowicz: "The influence of the previously accepted plan on my preferences at node  $n$  may, but need not, be decisive." See Wlodek Rabinowicz, "To Have One's Cake and Eat It Too: Sequential Choice and Expected-Utility Violations," *Journal of Philosophy* 92 (1995): 586-620, at 606.

<sup>44</sup> An analogous worry is at the heart of a challenge that has been posed by Joseph Raz and Nico Kolodny to norms of plan rationality. I agree with these philosophers that if these norms were simply responsive to mere psychic tidiness then it would be unclear why they should have normative significance. But in my "Rational Planning Agency" [this volume, essay 10] and my "Introduction: The Planning Framework" [this volume, essay 1] I argue that these norms are not simply responsive to mere psychic tidiness: they are responsive to conditions of self-governance. In turning later in this discussion to an end of diachronic self-governance I am pursuing a related strategy with respect to present issues about diachronic self-governance. See Joseph Raz, 'The Myth of Instrumental Rationality', *Journal of Ethics and Social Philosophy* 1:1 (2005) and in Nico 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, at 241.

<sup>45</sup> Thomas Kelly, “Sunk Costs, Rationality, and Acting for the Sake of the Past,” *NOUS* 38 (2004): 60-85, esp. 73-75.

<sup>46</sup> One consideration in favor of appeal to this specific end is that this appeal fits well with a plausible theory of plan rationality. I explain this in “Rational Planning Agency” and “Introduction”. This connection with issues about plan rationality is also relevant to my assessment of an alternative, intellectualistic strategy developed by J. David Velleman. In his “Centered Self,” in his *Self to Self* (Cambridge University Press, 2006), 253-283, at 272, Velleman says: “my intellectual drives ... favor fulfilling my past intentions.” But, as I note in “Rational Planning Agency,” this will lead to a cognitivist treatment of basic norms of plan rationality, and there are good reasons to be wary of such a cognitivism.

<sup>47</sup> Sarah Paul, “Diachronic Incontinence is a Problem in Moral Philosophy,” *Inquiry* **57** (2014), 337-355; Sergio Tenenbaum and Diana Raffman, ‘Vague Projects and the Puzzle of the Self-Torturer’, *Ethics* **123** (2012), 86-112, esp. section III.

<sup>48</sup> This is the parallel with views of Sobel and Rabinowicz anticipated in note 43. It is also a feature of Velleman’s view cited in note 46.

<sup>49</sup> Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon, 1986), 325-6.

<sup>50</sup> Gregory S. Kavka, “The Toxin Puzzle,” *Analysis* 43 (1983): 33-36.

<sup>51</sup> This is a basic idea in my “Toxin, Temptation, and the Stability of Intention” and my “Temptation and the Agent’s Standpoint”.

<sup>52</sup> “Toxin, Temptation, and the Stability of Intention”.

---

<sup>53</sup> *Being and Nothingness* (Hazel Barnes translation) (New York: Washington Square Press, 1984), 612.