

Sovereign Agency<sup>1</sup>  
Matthew Noah Smith  
University of Leeds

Forthcoming in:  
Brendan Balcerak-Jackson and Magdalena Balcerak-Jackson, eds. *Reasoning* (Oxford University Press, 2017)

I.

This essay makes a case for the practical authority of deliberations and the intentions they yield. I argue that sound deliberations yielding an intention to act are together a (i) content-independent reason not to re-open deliberations about how to act and (ii) a content-independent reason to act as intended on the basis of those sound deliberations.<sup>2</sup> Many philosophers have argued that this sort of ‘bootstrapping’ is impossible.<sup>3</sup> In this essay, I neither rehearse nor challenge those arguments.<sup>4</sup> Rather, my aim is to defend

---

<sup>1</sup> This essay benefited from comments from audiences at Oxford University, Leeds University, the Bled Philosophy Conference, and the Konstanz Reasoning Conference. This essay benefited enormously from written comments from Nishi Shah, Luca Ferrero, Kieran Setiya, Eric Wiland, Brendan Balcerak Jackson and Magdalena Balcerak Jackson. A lengthy discussion with Stephen Darwall was also extremely valuable. Special thanks again to Magdalena and Brendan for organizing the Konstanz conference.

<sup>2</sup> I explain the concept on content-independent reasons below.

<sup>3</sup> For the contemporary canonical statement of the bootstrapping objection, see Michael Bratman, *Intentions, Plans and Practical Reasons* (Cambridge, MA: Harvard University Press), pp. 24 – 7. In Bratman, “Intentions, Practical Rationality and Self-Governance,” *Ethics* 119 (April 2009): 411- 443, Bratman challenges the bootstrapping objection with respect to non-modifiable intentions that are expressions of one’s self-constituting policies or commitments. Other statements of the bootstrapping objection include the following. “Forming an intention to do something surely cannot give one a reason to do it that one would not otherwise have. If it did, we could give ourselves a reason to do something just by intending to do it; and that cannot be right.” (Richard Holton, “Rational

bootstrapping, i.e., to defend the claim that, given certain conditions, and when taken together, a deliberation whether to  $\phi$  and the intention to  $\phi$  that it yields are, for the intending agent, a reason to  $\phi$ .

---

Resolve,” *Philosophical Review* 113 (2004): 507 – 535, 513; “...it is not credible that, just by adopting some end, you make it the case that you have reason to pursue it” (John Broome, “Have We Reason to Do as Rationality Requires? A Comment on Raz,” *Journal of Ethics and Social Philosophy* Symposium I (2005): I – 8, I); and “... the blanket conclusion that having goals or intentions provides reasons [is false]...” (Joseph Raz, “Instrumental Rationality: A Reprise,” *Journal of Ethics and Social Philosophy* Symposium I (2005): I – 19, 19). See also Joseph Raz, “The Myth of Instrumental Rationality” I *Journal of Ethics and Social Philosophy* I (April 2005): I – 28; John Broome, “Are Intentions Reasons?” in *Practical Rationality and Preference: Essays for David Gauthier*, Christopher Morris and Arthur Ripstein, editors (Cambridge: Cambridge University Press, 2001), pp. 98–120; John Broome, “Reasons” in *Reason and Value*, ed. R. Jay Wallace, et. al. (Oxford University Press: 2004), 28 – 55; John Broome, “Does Rationality Give Us Reasons?” *Philosophical Issues* 15 (2005): 321 – 37; Kieran Setiya, “Cognitivism About Instrumental Reason” *Ethics* 117 (2007): 647 – 73 (Setiya describes bootstrapping as “illicit” but then goes on to defend a belief-based ‘cognitivist’ version of bootstrapping); and Garrett Cullity, “Decisions, Reasons and Rationality” *Ethics* 119 (2008): 57 – 95, 63 – 67. See also Luca Ferrero, “Decisions, Diachronic Autonomy & the Division of Deliberative Labor” 10 *Philosophers Imprint* 2 (2010): I – 23, especially pp. 3 – 6. An influential related discussion appears in Christine Korsgaard, “The Normativity of Instrumental Reason,” in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997), 215–54. In “Intentions, Practical Rationality and Self-Governance,” at footnote 20 (pp. 416 – 417), Bratman questions whether both Broome and Raz should be read as treating the bootstrapping objection as something other than a blanket rejection of intentions being reasons.

<sup>4</sup> For an overview of the arguments and objections against them, see Matthew Noah Smith, “One Dogma of Philosophy of Action” 173 *Philosophical Studies* (2016): 2249 – 2260.

That such a defense is available invites revisiting objections to bootstrapping and reflection on whether such objections are as strong as they are typically taken to be.<sup>5</sup>

The essay proceeds in two steps. I first argue that deliberations and intentions have certain functional roles, namely, deliberations about whether to  $\phi$  that yield an intention to  $\phi$  function as a reason to intend to  $\phi$ , intentions to  $\phi$  function as reasons not to re-open deliberations about whether to  $\phi$ , and finally intentions to  $\phi$  function as reasons to  $\phi$  (these are all agent-relative reasons by the way). The second step is to argue that intentions ought to play these roles. I then consider an important objection.

2.

One must always take a stand on certain philosophical issues when defending a certain view. This makes whatever conclusion is reached conditional on the truth of those assumptions. The assumptions I am making, then, are the following. First, I assume that intentions are mental states. Second, I assume that intentions bear special action-establishing relationships to behavior such that when that relationship is instantiated (and, as they say, ‘the world cooperates’) there is action and, at least for the paradigmatic instances of action, when that relationship is not realized, there is only behavior (for example, what makes

---

<sup>5</sup> Nothing here is meant to apply to questions related to epistemic bootstrapping. But, there may be connections.

For more on epistemic bootstrapping, see, e.g., Stewart Cohen, “Basic Knowledge and the Problem of Easy Knowledge” *Philosophy and Phenomenological Research* 65 (2002): 309–29, Jonathan Vogel, “Epistemic Bootstrapping” 9 *Journal of Philosophy* 105(2008): 518 – 39; Stewart Cohen, “Bootstrapping, Defeasible Reasoning, and *A Priori* Justification” *Philosophical Perspectives* 24 (2010): 141–59; and quite generally Jonathan Weisberg, “The Bootstrapping Problem,” *Philosophy Compass* 7 (2012): 597 – 610.

my sneezing not an action is that even if I intend to sneeze and I sneeze, the relationship between the intention and the sneeze is not the right one, whatever that may be).

Furthermore, I am concerned primarily with prospective intentions and not intentions in action, at least insofar as the latter bear no significant rational dependence on the former.

Finally, this essay discusses only intentions that are products of deliberations. For now, call these *deliberative intentions*. Deliberative intentions are distinct from intentions adopted in the absence of any deliberations whatsoever. Such intentions are akin to sudden urges. Call these *arational intentions*. There are also intentions adopted contrary to one's stable best judgment about the balance of reasons. Call these *akratic intentions*.<sup>6</sup> The latter two classes of intentions are not paradigmatic products of healthy agency. They are instead features of lesser or degraded forms of agency. My methodological commitment is to investigate paradigmatic instances of some phenomenon in order to understand that phenomenon. Thus, I put aside arational and akratic intentions, as these intentions are not features of paradigmatic forms of human agency and therefore are not paradigmatic kinds of intention.

The deliberations that yield deliberative intentions need not be explicit, conscious, or executed immediately prior to that particular intention. They can instead be in the psychological background, or recoverable upon prompting, or recollected as older deliberations that drove one to adopt certain practical policies, or something else along these lines.<sup>7</sup> In these cases, the intention is easily or immediately

---

<sup>6</sup> There are also intentions to establish a Nash equilibrium. In these cases, one adopts an intention because one has conclusive reason to select *some* option from the available equilibria although this reason has nothing to do with the merits of one option over another. Call these *plumping intentions*.

<sup>7</sup> See, e.g., Michael Bratman, "Valuing and the Will." *Philosophical Perspectives* 14 (2000): 249-65; and Michael Bratman, "Intention, Planning, and Temporally Extended Agency," *The Philosophical Review* 109 (2000): 35-61.

recognizable to the agent as based on reasoning she has already performed, or values and commitments she has already adopted.

These values and commitments can in turn be products of prior reflections or a series of decisions made on the basis of deliberations and then enacted so many times that habits form. In this way, we can deliberatively settle on values and commitments that allow us to make the quick decisions that determine how much of our lives go. But, these habits and quick decisions are just expressions of the deliberations from which they initially sprang. In this way, habitual intentions (so understood) have the sort of provenance that makes them more similar to deliberative intentions than to arational intentions.

3.

The function of deliberation is to produce action by way of an intention. But, deliberation functions correctly not just by producing an intention that in turn produces an action. For, when one deliberates, one deliberates about what the thing to do is, or what the best thing to do is, or how one should live. That means that the function of deliberations is *to produce an action that one has decided to take on the basis of those deliberations*. So, deliberations do not merely cause actions. Rather, their function is to determine what it is reasonable for the agent to do and then to produce an action on the basis of those reflections about what it is reasonable to do. In this way, the function of deliberations is to make it such that the agent is *authorized* to take the action the deliberations recommend.

The intuition behind this is that insofar as we have practical authority over ourselves at all – insofar as we are the authors of our own actions – then we must be able to *authorize* our actions. But, authorization is a process. What process? Maybe it is something over and above practical deliberation. But, that would make authorization quite mysterious – a process we have never really noticed. Instead, we

might just treat deliberations yielding an intention to be the authorization process itself. Doing this allows us to explain how actions are not merely *caused* by the deliberations, as ripples on water are caused by a stone hitting that water. Furthermore, it helps us to see how diminutions in the capacity to deliberate limit or diminish one's authority over oneself.<sup>8</sup>

Deliberations alone, though, cannot authorize taking some action. For, deliberations that are never complete authorize nothing at all. It is only the completion of deliberation that allows for authorization. Completed deliberations produce intentions. It is this deliberation-intention combination that does the authorizing. But, this invites queries into the relationship between deliberations and intentions. Deliberations cannot merely cause any old intention. The deliberations must also authorize those intentions. In this way, there is a *chain of authorization* linking deliberation to intention to action. Thus, insofar as the function of deliberations is to authorize actions, they also function as authorizations of intentions, and these together function as authorizations of the intended action. In short, the function of deliberations and intentions is to give the agent both standing and reason to take the action.

4.

That is a very broad sketch of an argument for deliberations and intentions functioning as authorities. Let us take a closer look. Suppose I deliberate about whether to go to the store. On the basis of those deliberations, I form the intention to go to the store. But, before I leave, a mad scientist uses his new mind control technology to make me forget everything that has just happened, thereby destroying the deliberation-authorized intention to go to the store. The mad scientist also has a device that reads my

---

<sup>8</sup> For more on this and the moral significance of such agents, see Agnieszka Jaworska, "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value," *Philosophy and Public Affairs* 28 (1999): pp. 105 -138.

mind prior to his zapping me. On the basis of that, he implants in me, after destroying my memory and my authorized intention to go to the store, a new intention to go to the store. This intention causes me to go to the store. This sort of scenario is a case of compromised agency. For, although I am doing as I intended because of an intention, I am not going to the store on the basis of my deliberations. What is missing is the *authorization* of that intention. It is as if an army had swept into a country, deposed its government, destroyed the government agencies, and then rebuilt everything entirely in the image of the old government. Meet the new boss, it's not the same as the old boss.<sup>9</sup>

This thought experiment suggests that mere causal connections between deliberations, intentions, and action are not sufficient for the realization of agency to the fullest and richest extent. What is required are deliberations authorizing the formation of an intention, with that intention in turn authorizing the action. Causal connections of the right kind are merely necessary but partial *grounds* of the authorization, in the same way that utterances of the right kind are merely necessary but partial grounds of, e.g., promissory obligation. If authorization was not required, then all we'd have would be deliberations about how one should live, arational causal connections between those deliberations and an intention, and the intention and some behavior.

---

<sup>9</sup> There is no sense in which this is about free will. It is about whether an action is authorized by deliberations. So long as one recognizes an important difference between the intention produced by deliberations and the intention produced by a mad scientist, then one recognizes the significance of authorization. More needs to be said about the conditions on which authorization supervenes. But, it seems that there is a referential component to authorization: X authorizes *this* intention, and any other intention implanted, however similar, is not authorized. Notice also that if one knew about the mad scientists and one's deliberations authorized any intention with a certain content regardless of its provenance then the mad scientist-implanted intention is authorized.

At this stage, someone might resist this by arguing that deliberation's role is entirely instrumental. The best possible deliberations do no more than reliably cause the correct intention, i.e., the one that the balance of reasons supports. On this picture, then, the character of the psychological process that leads up to behavior is irrelevant. Whether the intention meets independent standards is all that matters. There would be no difference between good deliberations about whether to live life in some way and mere causal influences that force one to live life in that way. But, there is pressure for our deliberations to bear a proper rational relationship to our intentions.

For example, imagine the following deliberative train: I love cooking, I ought to do what I love (so long as it's feasible), I've got a lot of experience working on the line in a restaurant, I've been told I am pretty good at cooking, restaurant life is better than lives associated with other career paths open to me (despite the fact that I will make more money in some of those career paths), and so on along those lines. These deliberations cause me to decide to pursue a career as a chef.

Now imagine this deliberative train: my cat's breath smells like cat food, a name is just a sound somebody makes when they need you, the NHS should be fully funded, these fingers have veins that run straight to the soul of man – the right hand, friends, the hand of love. This train of thoughts causes me to decide to pursue a career as a chef.

Let us assume that pursuing a career as a chef is the course of action I ought to take. Are there any differences between these two cases? One case is a case of full-blooded agency and the other is something far less than that. But, if the sole function of deliberations is merely to cause certain intentions to perform 'correct' actions, then there is no functional difference between these deliberation-intention pairs. But, there clearly is a difference. The first is a case of deliberation authorizing an intention. The second is just a mess and it cannot authorize anything at all. To see this, suppose the second deliberation was publicly



revealed. The recommendation that I revisit my decision to become a chef would be appropriate, even given the fact that becoming a chef was the correct course of action. Such a recommendation wouldn't be warranted in the case of the first deliberation, though. That's because the first deliberation authorized the intention to become a chef.

One might respond that I am confused. The function of deliberations is to reliably yield correct decisions, goes this objection. The formal character of the second wild train of thought is such that if repeated it is not likely to yield a correct decision. So, that train of thought having functioned as deliberations is a problem, even if it luckily yielded the correct decision. The recommendation to reconsider is therefore warranted. For, that would presumably steer the agent towards more reliable deliberative practices. Thus, insofar as deliberations are significant with respect to current (correct) decision, then, they are at best attractive window dressing – they make the decision 'look better' and so have at best aesthetic significance. They do not make a practical difference to the authority of the intention.

This is odd, though. First, the decision itself does not seem more or less aesthetically attractive given the quality of the deliberations behind it. Only the *deliberations* are apt bearers of that value. For, it is false that because some aesthetically attractive X produced Y, Y is therefore made more attractive. Having beautiful parents does make a child more beautiful.

Second, suppose that we are *certain* that this instance of bad thinking is a one-off case. Even if we know the decision it yielded is the correct one, it is still appropriate to criticize the deliberations and recommend reconsideration (assuming that the agent has the capacity to re-deliberate correctly). For example, suppose that wise King Solomon has up until now always made the wise decision on the basis of reliable deliberations. This morning, though, he tried Biblical marijuana for the first time. Now high, he is

faced with the question of which of the two women was the mother of a baby. Stoned, Solomon thinks about the cats prancing about his palace, then about the attractiveness of their swishing claws, and then he reflects on the ridiculous weight of his crown. As a result of this train of thought, he decides to threaten to cut the baby in two in order to determine who gets the baby. Because Solomon does not enjoy being stoned, this will be a one-off case. And because the Biblical marijuana is weak, he will return to brilliant reasoning very quickly. Let us now suppose that the threat to cut the baby in two is the *correct* decision. Is there something defective about Solomon threatening to cut the baby in two on the basis of his stoned reasoning? Yes. For, the decision-generating train of thought is utterly corrupt. It must be tossed out along with the decision it generated. So, even if Solomon's decision is correct, and even if everyone knows that Solomon will never again reason this poorly, that he has made the decision on the basis of such corrupt reasoning is grounds for him to throw out that decision and try again. That he is reliably comes to the right decision is just not germane to this issue.

I conclude, then, that the function of deliberations is to authorize actions via a certain pattern of reasoning. The deliberation function as an action-authorizing process via it the authorization of intentions, and it is through these intentions that the actions are authorized.<sup>10</sup>

---

<sup>10</sup> One might object at this stage that my account of deliberations appears to commit me to an anti-naturalist metaethical view. No good account of deliberations should do that! This objection fails, though. First, I am working within the realm of reasons, in which people happily talk of the balance of reasons making some action the right thing to do. That is a purely normative relationship, and not in any way a causal one (despite the unintentionally deceptive physics-y talk of the weight of reasons, the balance of reasons, and so on). Furthermore, everything I am saying can be interpreted within a suitably powerful expressivist framework. If we can use

5.

So far, I've argued that deliberations have the function of authorizing actions. Furthermore, if deliberations are to authorize actions, then deliberations must also authorize the intentions that are the next proximate source of behavior. Finally, if the authorization is to flow without disruption from deliberation to action, then the intentions also must function as authorities with respect to the action in question. In short, the function of an agent's deliberations and intentions is to authorize the agent's actions.

In the next sections, I offer further support for this claim. The argumentative strategy is to show that a parsimonious partial understanding of three interrelated phenomena, namely, responsibility for actions, compliance with reasons, and self-knowledge via one's actions, involves intentions functioning as action-authorizing attitudes.

6.

The paradigmatic case of responsibility for action is the case of someone acting on the basis of an intention that was produced by sound deliberation.<sup>11</sup> By 'paradigm' here, I mean the pattern of phenomena

---

expressivism to analyze normative language, then we can apply that to accounts of deliberations that authorize intentions and actions. So nothing I am saying here commits me to an objectionable metaethical view.

<sup>11</sup> Thus the old legal saw, *actus reus non facit reum nisi mens sit rea*. For philosophical discussion, see, e.g., Jonathan Glover, *Responsibility* (London: Routledge, 1970); Thomas Scanlon, "The Significance of Choice," in Sterling M. McMurrin (ed.), *The Tanner Lectures on Human Values, Vol. 8* (Salt Lake City, Utah: The University of Utah Press, 1988), pp. 149-216; Susan Wolf, *Freedom Within Reason* (Oxford: Oxford University Press, 1990); the essays collected in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press,

that realizes the archetypical instance of responsibility. There are certainly other patterns of phenomena that realize responsibility, but these patterns do so conditional on their being properly similar to the paradigm. Any case of responsibility for action that departs radically from this paradigm is *ipso facto* a radically unusual case – one that would either require an *ad hoc* amendment to our received conception of responsibility, or a wholesale reconsideration of whether that conception of responsibility is correct. Thus, when I say that the paradigmatic case of responsibility for action is the case of action produced by intentions that were in turn produced by healthy deliberation, I am saying that this is our starting point for thinking about responsibility, and quite far from a complete theory of responsibility. Nonetheless, it is a starting point with a bite: every step away from it must be made on the basis of good argument.

It would be quite odd if in this paradigmatic case of responsibility there was no *normative* connection between the agent’s psychology and her behavior. For, we generally think that *mere* causal links between deliberations and intentions, and intentions and behavior, are not sufficient for responsibility. First, were intentions merely causes, then their effects would be akin to the effects of an outside force. The

---

1993), pp. 119–148), some of which reject this view but nonetheless treat it as paradigmatic; R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994); the essays collected in Gary Watson, *Agency and Answerability: Selected Essays* (Oxford: Clarendon Press, 2004), especially “Two Faces of Responsibility”; Angela Smith “Control, Responsibility, and Moral Assessment,” *Philosophical Studies* 138 (2008): 367 – 392, 373; R. Jay Wallace, “Caring, Reflexivity, and the Structure of Volition” in Wallace, *Normativity and the Will* (Oxford: Clarendon Press, 2006), pp. 190 – 211, p. 200). For an opposing view, see Robert Adams, “Involuntary Sins,” *Philosophical Review* 94 (1985): 3–31. An influential outlier is Harry Frankfurt’s work, which doesn’t treat deliberation and intention as philosophically salient. See Harry Frankfurt, “The Importance of What We Care About,” reprinted in Harry Frankfurt, *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), and Harry Frankfurt, *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999).

more one is affected by non-rational, normatively inert forces, the less responsible one is for one's mental life and one's behavior. Second, if the deliberations merely caused the intentions, then it is entirely possible that one could deliberate about whether to  $\phi$ , never decide to  $\phi$  (and so it could still be open that one might decide not to  $\phi$ ), and yet as a causal upshot of that deliberation, one intends to  $\phi$ . This is like the cause of a deviant causal chains producing behavior. Deviant causal chains are deviant instances of responsibility (it says so right there on the tin). So, in the paradigmatic case of responsibility, one's deliberations and intentions cannot bear a *purely* causal relation to one's action.

For example, suppose I am fantasizing about stealing a sandwich. What is a fantasy? A fantasy it is, for our purposes, a first-person train of thought that is *deliberatively offline*. It is just idle imagining in which I am the protagonist.<sup>12</sup> Now, suppose that I am fantasizing in the following sort of way: "Mmmm I'm hungry... and look at that amazing sandwich... it's got everything I could ever want in a sandwich... I

---

<sup>12</sup> This is a crucial point. We must distinguish our off-line, first-personal imaginings from our deliberations, even if when represented propositionally they are indistinguishable. The reason why is that the psychological processes in which the propositions are tokened – fantasy and deliberation – have different functions. For further clarity, consider the following example:

The next day Rastignac dressed himself very elegantly, and at about three o'clock in the afternoon went to call on Mme de Restaud, indulging on the way in those dizzily foolish dreams which fill the lives of young men with so much excitement: they then take no account of obstacles, nor of dangers, they see success in everything, poeticize their existence simply by the play of their imagination, and render themselves unhappy or sad by the collapses of projects that had as yet no existence save in their heated fancy... (Honoré de Balzac, *Le Père Goriot*, translation from Dorrit Cohn, *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction* (Princeton, NJ: Princeton University Press, 1983), p. 24.)

In this instance, Rastignac is fantasizing taking a course of action and then further imagining how it will play out..

could easily steal the sandwich. . . I should steal and then eat that sandwich. . .” Unbeknownst to me, prior to this fantasy, I’d been slipped a drug that affects my mind such that I perform my fantasies. As a result I steal the sandwich I’ve been fantasizing about. In this case, my fantasy is an off-line script that I was press-ganged (by the drug) into enacting. Am I responsible for stealing the sandwich? Perhaps, I am. But, this would clearly be an attenuated sort of responsibility and not a paradigmatic instance of responsibility. For, although I am the one who supplied the script by fantasizing, it is not the case that I was engaged in the sort of mental exercise whose function is to authorize my own actions. Thus, in a responsibility-diminishing way, I am estranged from the action, even if it enacts a script I wholeheartedly produced.

The lesson from this case is that authorization by deliberation and intention is a necessary feature of the paradigmatic instance of responsibility. (Recall that this is a *paradigm* of responsibility, which is absolutely not the same thing as the *only* kind of responsibility; it is just the pattern we use when constructing other conceptions of responsibility). But, since authorization partially establishes full responsibility for an action, and since this occurs by way of deliberations authorizing intentions that in turn authorize actions, the authority of deliberations and intentions is a condition of full responsibility. Simply on grounds of parsimony, we ought not posit a *different* normative relationship between intentions and actions in addition to the authority relationship in order to explain why deliberation/intention-authorized actions are the paradigmatic instances of responsibility for action.

7.

It is widely held that an important feature of human agency is that we can comply with reasons in addition to merely conforming to them. Complying with certain reasons as opposed to merely conforming

with those reasons requires not accidentally acting in response to those reasons.<sup>13</sup> It involves deliberately grounded guidance of action by the reason. How does this work?

Suppose someone intends to go to the store. Suddenly, she realizes she has forgotten why she intends to go to the store. She knows she is going to the store, but the intention to go to the store has disappeared. There is just inertial behavior. The obvious next psychological move is to deliberate about whether she ought to go to the store. On the basis of reflections like this a natural explanation of how compliance with reasons is possible is to claim that so long as one remembers one's deliberations and one knows the reason for which one is acting, one thereby complies with that reason.

This approach is problematic, though. For, if the intention moving one to act is not the one authorized by the deliberations one remembers, then one is not complying with the reasons considered in those deliberations. Rather, one is luckily doing something but erroneously associating that behavior with prior deliberations. For example, imagine an addict who deliberates about whether to take heroin. But then her addiction overwhelms her capacity for self-control, triggering in her an overwhelming desire to take heroin. As she is shooting up, she recalls her deliberations. This is not sufficient to manifest compliance with any reason she considered in those deliberations. For, since her action was caused by a short-circuit to her deliberative process, she did not comply with any reason at all. She at best conformed with a reason, in the same way that a rock in free fall in a vacuum conforms with Newton's equation describing gravitational attraction.

---

<sup>13</sup> For more, see John Gardner and Timothy Macklem, "Reasons," in *The Oxford Handbook of Jurisprudence and Philosophy of Law*, edited by Jules Coleman and Scott Shapiro (New York: Oxford University Press, 2002), pp. 440 – 475.

Gardner and Macklem talk of 'deliberately' acting for a reason and 'accidentally' acting for a reason instead of conforming and complying.

Consider an intention to pull the trigger of the gun. This is an incomplete account of this intention. For, my intention is to pull the trigger for the sake of demonstrating how to fire a gun. This intention is different from the troublemaker's intention to pull the trigger for the sake of making a startling noise. These intentions are similar in that they are intentions to pull the trigger, but they are also very different. For, they involve compliance with different reasons. My reason – to demonstrate how to fire a gun – makes my intention to pull the trigger quite distinct from the troublemaker's intention to pull the trigger. As I will discuss in more detail presently, the reasons for the intention were considered in the deliberations, and in being the substantive part of the process that authorized the intention, they are exactly what allow us to say that my intention to pull the trigger of the gun is different than the troublemaker's intention to pull the trigger of the gun.

Or consider the many reasons I have to go for a run along the canal this afternoon: it will afford me a beautiful view, it will allow me to clear my mind after writing all day, it will relax me, it will help me get fit, and so on. Suppose that I decide to go for a run and the reason for which I do so is that it will help me get fit. In particular, I think about how I am getting a little soft around the middle, I reflect on how I would like that process to stop or to be reversed, and finally I think about how regularly running can slow and eventually reverse that process. On the basis of this deliberation, I form the intention to go for a run in order to get fit. I am, of course, aware of all the other considerations supporting going for a run. But, deliberations regarding getting fit are what authorized my intention to run along the canal. The intention must somehow reflect this if there is to be compliance with the relevant reason.<sup>14</sup>

---

<sup>14</sup> This does not mean that one must see the intention itself as the reason for one's action. Rather, it is in virtue of *having* the intention that one acts for a reason, and therefore in virtue of the intention that one complies with a reason.



As suggested, we can explain how an intention reflects the reason for which one acts by appeal to deliberations and intentions functioning as the authoritative ‘practical voice’ of the agent. If intentions are understood as intentions to  $\varphi$  on the basis of deliberations  $D$ , where the ‘on the basis of’ relation is not a causal one but is instead one of authorization, an intention to  $\varphi$  is actually an *intention to  $\varphi$  as authorized by deliberations  $D$* . This is represented colloquially by reference to what it is for the sake of which one is doing something. Intending to go for a run for the sake of getting fit is just intending to go for a run as authorized by deliberations about how I am committed to getting fit and about how going for a run will get me fit.<sup>15</sup> This is, as suggested above, how reasons ‘get into’ intentions. It explains how agents can comply with as opposed merely conform to reasons. In this way, an appeal to deliberations and intentions functioning authoritatively not only helps to explicate how it is that one can be responsible for an action, it also explicates how one can act in compliance with reasons.

8.

There is a significant philosophical tradition according to which a hallmark of action is that “when someone is acting intentionally, there must be something he is doing intentionally, not merely trying to do, in the belief that he is doing it.”<sup>16</sup> David Velleman argues that this knowledge about what one is doing

---

<sup>15</sup> An upshot of this is that any time there is a bare intention to act as opposed to an intention as authorized by deliberations, that bare intention is not a full-throated expression of one’s agential capacities.

<sup>16</sup> Setiya, *Reasons Without Rationalism*, p. 26.

when one acts is in fact a constitutive aim of action. To some degree, we know ourselves *through* our actions.<sup>17</sup>

A crucial element of action that feeds into one's knowledge about what one is doing is the reason for which one acts. For example, I understand my running along the canal in terms of a particular reason for running along the canal, not in terms of just any reason for running along the canal. When I am running along the canal for the sake of getting fit, what I believe that I am doing is something rather different from what I believe I am doing when I am running along the canal for the sake of clearing my mind after writing all day. Our knowledge of our own actions is partially facilitated by our capacities to comply with reasons. But, if the previous section's conclusion was correct, one's knowledge of one's own actions is partially facilitated by the one's deliberations and intentions functioning as authoritative. So,

---

<sup>17</sup> See, generally, Velleman, *Practical Reflection*, Velleman, *The Possibility of Practical Reason*, and J. David Velleman, *How We Get Along* (New York: Cambridge University Press, 2009). Setiya says that "self-knowledge can be described as the constitutive aim of action: it is a goal towards which intentional action is always and essentially directed." (Setiya, *Reasons Without Rationalism*, p. 108) See also Harry Frankfurt's work, according to which behavior that flows from attitudes with which the agent 'wholeheartedly identifies' is the paradigmatic form of action. Identification with an attitude is a complex state of affairs, but it involves something like self-understanding. One does not *discover* that one is acting on the basis of attitudes with which one also happens to wholeheartedly identify. Rather, one acts with the knowledge that such-and-such is in fact that for the sake of which one is acting. See Harry Frankfurt, "Identification and Wholeheartedness" in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 159 – 176, and "Autonomy, Necessity, Love," in *Necessity, Volition and Love* (Cambridge: Cambridge University Press, 1999), 129 – 141. See also G. E. M. Anscombe, *Intention*, 2<sup>nd</sup> Edition (Ithaca, NY: Cornell University Press, 1963) and Sarah K. Paul, "How We Know What We're Doing," 9 *Philosophers Imprint* II (2009): 1 – 24.

deliberations and intentions functioning as authoritative is therefore crucial to the construction of one's knowledge of what one is doing when one is acting.

9.

Responsibility for an action, acting for a reason, and knowing what one is doing are central to paradigmatic instances of action. Assuming that agency is ever realized in its paradigmatic form, it has these characteristics. And even if agency is never realized in this paradigmatic form (since we are all imperfect), this is still the paradigm of agency against which our more typical imperfect is understood. This is why we should interpret the three just discussed interlinked phenomena in a somewhat parsimonious fashion by interpreting them in terms of deliberations and intentions functioning as authorities. One upshot of this is that it aligns the claim that deliberations and intentions function as authoritative with other, less controversial claims about what it is to be an agent, thereby domesticating what some may think are crazy propositions – that a defining function of deliberations is to authorize intentions and actions, and a defining function of intentions is to authorize actions – by showing the very useful roles these propositions can play in our accounts of three familiar aspects of agency. That is, insofar as one is comfortable with saying that people are responsible for their actions, that people can comply with reasons, and that one aim of action is self-knowledge, then one should also be comfortable with saying that the function of deliberations and intentions is to authorize actions.

10.

If deliberations authorize the formation of an intention, this authorization is more than mere permission. Where A authorizes B with respect to O, A has given B some (or all) of A's own authority with

respect to  $O$ . So, when deliberations authorize intentions to  $\Phi$ , they give intentions some (or all) of their authority with respect to  $\Phi$ -ing.

An important part of the authority deliberations and intentions have is the authority to *end* deliberations about the practical question under consideration. In particular, deliberations must have the authority end themselves, and the intentions they authorize must have the authority to *keep* deliberations closed. This is a crucial function of intentions because it allows deliberations to settle practical questions. This is one way to interpret one of Michael Bratman's central claims about intentions.<sup>18</sup> According to Bratman, existing intentions play a structuring role in our deliberations about how we are going to live. When one settles on a plan, that plan becomes, as Bratman puts it, a 'fixed point' for future deliberations. But, what *fixes* that point? It is not a merely causal matter, where out of physical necessity one reasons on the assumption that the intended action will occur. Rather, it is a normative matter. Intentions function as requirements not to re-open deliberations about how to live.<sup>19</sup>

Putting aside the appeal to Bratman, just consider the issue on its own for a moment.

Deliberations, if they are to produce actions, must come to an end at some point. They cannot just end willy-nilly, though. That would be like treating a legislature being blown up while considering some bill as

---

<sup>18</sup> Bratman writes: "...unless and until I do give up or reconsider my prior intention, its role in my means-end reasoning will be to set an end for that reasoning and not just to provide one reason among many... to see my intention as providing just one reason among many is to fail to recognize the peremptoriness of reasoning-centered commitment." Bratman, *Intentions, Plans, and Practical Reasoning*, p. 24, footnote removed.

<sup>19</sup> The nature of this requirement is a matter of great debate. Obviously I am arguing for the view that intentions are reasons, not that they trigger, for example, mere Broomean normative requirements. But, so far, I've only been attempting to show that intentions function as reasons, not that they are reasons.

no different than the legislature validly passing that bill, transforming it into law (I am assuming away the peculiar practice of the presidential veto).<sup>20</sup> Just as a legislature has the authority to declare its deliberations over by passing a bill, the capacity for practical deliberation must have the authority to ‘declare’ its deliberations over by forming an intention. The passage of the bill is the authorization of the bill into a law, and the completion of deliberations with a decision is the authorization of that decision as an intention.

Upon reflection, this is quite clear. Deliberations often have merely causal endings – one is interrupted by some task mid-deliberation, one is startled by something, one simply runs out of steam before making the decision. But, these endings do not *authorize* anything. We must pick up where we left off if whatever behavior flows from the deliberation is to be a full-throated instance of agency. On the other hand, intentions do not rule out dispreferring the intended option relative to another option. One can also evaluate an option as morally best without also intending that option.<sup>21</sup> One might be criticizable for not intending to do what one judges one is morally required to do. But, this criticism is not grounded on any account of the nature of intentions. It is instead grounded in the claim that when one judges that the best thing to do is to  $\phi$ , then one will form an intention to  $\phi$ , and presumably will abandon intentions

---

<sup>20</sup> This does not mean that the law cannot be repealed or amended. It just means that for now the question is settled and for now all future rule making proceeds on the assumption that this law is a fixed element of the overall body of law.

<sup>21</sup> This is not a claim about reasons- or motivation-internalism. For taking oneself to have a reason to do something, or being motivated to do something is not the same thing as being committed to doing that thing. One can see oneself as having a reason to give to charity, one can thereby be motivated to give to charity, and yet still lack the kind of commitment to give to charity that generates the action of giving to charity. This could be weakness of will.

to act in ways incompatible with  $\phi$ -ing. That is not a thesis about the nature of deliberations and intentions. It is a thesis about the nature of moral judgment.

Additionally, an intention is not just the latest all-things-considered judgment about what it would be best to do. For, if an intention were no more than the latest all-things-considered judgment about what it would be best to do, not only would intentions be indistinguishable from evaluations, but akratic intentions would be impossible (since an akratic intention is an intention to act against what one judges to be the best course of action).

An intention is also not merely a summary of the latest deliberations. Just as one can summarize one's evidence for some proposition without making a judgment about whether that proposition is true, one can summarize the case for  $\phi$ -ing without thereby committing to  $\phi$ -ing. Summaries of deliberations have on their own no immediate significance. If one  $\phi$ 's on the basis of a summary, one would do so rationally. For, there would be nothing internal to one's reasoning that indicated that these were the reasons that settled the question on how to live.

Finally, an intention does not forever block re-opening deliberation. One remains open to at least some considerations in favor of suspending the intention and re-opening deliberations.<sup>22</sup> Again, consider the analogy with law. A state may pass a law requiring people to drive a certain speed. This does not mean that the state will never review this law and perhaps amend or repeal it. But, the state has to reflect on *whether* to amend the law before considering *how* to amend the law. The same point goes with intentions: we have to reflect on whether to change course in our lives before we can decide how to change course. So,

---

<sup>22</sup> In "Rational Resolve," Richard Holton discusses this point in the context of dispositions not to reconsider. See also Richard Holton, *Willing, Wanting, Waiting* (Oxford: Oxford University Press, 2009), chapters I, 6, & 7.

one function of an intention to  $\varphi$  is to require that the agent, before re-opening deliberations about whether to  $\varphi$ , consider whether to do so.

The central message of this section, then, is that for deliberations to have the function of authorizing actions, they must have the function of (i) requiring their own completion by way of the formation of an intention that in turn has the function of (ii) requiring the action, and that intention must also have the function of (iii) requiring that deliberations about the intended action are not re-opened.

I assume at this stage that if something functions as a requirement to do something, then it functions as a *reason* to do that thing. Now, for the sake of considering an objection, let us put aside cumbersome talk of function. Here is the objection: I have failed to appreciate wide-scope/narrow-scope distinctions. I am arguing for a narrow-scope reading – an intention to  $\varphi$  is a reason not to reopen deliberations and a reason to  $\varphi$ . But, the objection goes, *all* that I have shown is that in virtue of intending to  $\varphi$ , one has a reason not (to re-open deliberations about whether to  $\varphi$  & to intend to  $\varphi$ ). But, this is wrong. For, I *first* argued that deliberations authorize their intentions. Deliberations producing intentions are reasons for having those intentions. In virtue of deliberating about whether to  $\varphi$  and then, on the basis of those deliberations, forming the intention to  $\varphi$ , one has a reason to intend to  $\varphi$ . This has nothing to do with the question of whether I should re-open deliberations. Furthermore, that I have deliberated about whether to  $\varphi$  and then formed the intention to  $\varphi$  *cannot be changed*. These events are in the past. And they are the reasons for me to intend to  $\varphi$ . The die has been cast: one has a reason to intend to  $\varphi$ . So, one

cannot discharge the first wide-scope requirement by giving up the intention to  $\phi$  without falling afoul of this reason that one intend to  $\phi$ .<sup>23</sup>

Returning now to function talk, we can state the conclusion of this part of the paper: deliberations about whether to  $\phi$  yielding an intention to  $\phi$  together function as a reason to  $\phi$ , and as a reason not to re-open deliberations about whether to  $\phi$ .

## II.

A several points in this essay, I've pushed an analogy between intentions and legislation (and so between practical deliberations and legislative deliberations). At this stage, I want to tighten that analogy. There is a classic distinction in the philosophy of law, pithily presented by Thomas Hobbes: "Law in generall, is not Counsell, but Command. . . addressed to one formerly obliged to obey him."<sup>24</sup> Hobbes's aims here are to distinguish two ways in which agents can influence the way another agent lives her life and then to analyze law in terms of one of them. In particular, Hobbes is distinguishing counsel, or advice, with (legitimate) commands, or (legitimate) orders. Law, in virtue of it aiming to be the final word on how the subject ought to live, is a kind of command, and not a form of counsel. For, counsel simply aims to inform someone of what the best course of action would be. Even when one receives counsel that is extremely thoughtful and well-informed, the counsel does not have any special practical claim to fix one's

---

<sup>23</sup> A similar view in epistemology was recently defended in "The Conflict of Evidence and Coherence," forthcoming in *Philosophy and Phenomenological Research*.

<sup>24</sup> Thomas Hobbes, *Leviathan* (1651), C.B. Macpherson, ed. (Baltimore: Penguin Books, 1968), ch. 26, p. 312. See also Thomas Hobbes, *On the Citizen (De Cive)* (1642), R. Tuck and M. Silverthore, eds. and trans. (Cambridge: Cambridge University Press, 1998), ch.14, para. I.



course of actions. One must, in order to act on the basis of the counsel, *further* reflect on its wisdom, and then on the basis of those deliberations, make a decision about how to live. Counsel therefore does not end deliberation – it merely invites further deliberation and may or may not be followed by decision, much less a decision in line with the recommendations of the counsel.

Commands, on the other hand, operate differently (I am concerned only with valid commands now).<sup>25</sup> Commands are rational interventions into a deliberative process. They aim to rationally settle what course of action the subject ought to take. So, commands do not purport to offer just any old reason in favor of some course of action. For, if that is all they did then commands would simply be counsel.

What commands do is end deliberation, where this ending of further deliberation is not just a short-circuiting of further deliberation as might occur when one's deliberations about how to deal with

---

<sup>25</sup> In what follows, I summarize the canonical account of commands, as found H.L.A. Hart, “Commands and Authoritative Reasons” in H.L.A. Hart *Essays on Bentham* (Oxford: Oxford University Press, 1982). A more complex and somewhat more tendentious account of commands can also be found in Joseph Raz, *Practical Reason and Norms* (London: Hutchinson & Co, 1975) and Joseph Raz, *The Authority of Law* (Oxford: Oxford University Press, 1979). Nothing I say here commits me to Raz’s “service conception” of authority or to his normal justification thesis. I am, though, committed to the most general contours of his account of the relationship between authoritative directives and second-order reasons. This is the notion that authoritative directives have two different functions in practical reasoning: they function as ‘first-order’ or ordinary reasons to act as commanded, and they function as ‘second-order’ reasons to disregard at least some of the reasons speaking for and against performing the commanded action. An important feature of Raz’s view is that this phenomenon is all over the place. For recent discussion, see Joseph Raz, *Between Authority and Interpretation* (Oxford: Oxford University Press, 2009), pp. 141ff.

some threat are short-circuited by fear of that threat. Commands give subjects both a reason to do as commanded and a reason not to deliberate further about how to live.<sup>26</sup>

The reasons given by commands are limited and defeasible. My sergeant's command only has scope over some parts of my life, namely the parts of my life over which my sergeant has authority. My sergeant's command that I drop and give her twenty gives me a reason to do twenty push-ups, but my sergeant's command that I divorce my spouse does not give me a reason to divorce my spouse. For, my sergeant has no authority over that part of my life.

Commands can also be defeated by other considerations. For example, suppose my sergeant gives me orders to fire my weapon at a potential threat. This gives me reason to fire my weapon. But if I am certain that the target is an innocent child, this reason is presumably overridden by other considerations (or outweighed, or whatever metaphor one wants to use).

Commands also do not rule out continuing evaluation about the merits of  $\phi$ -ing or fantasies about not- $\phi$ -ing, and they do not require changing considered preferences about  $\phi$ -ing versus alternatives. One can be commanded to  $\phi$ , thereby have reason to close deliberations about whether to  $\phi$  and reason to  $\phi$ , but *still* rationally reflect on the reasons that ground  $\phi$ -ing. But, a valid command to  $\phi$  is sufficient to give one reason to  $\phi$ . All the while, one may still reflect on why it is good to  $\phi$ , consider what the superior alternatives to  $\phi$ -ing might be, and fantasize about doing something other than  $\phi$ -ing.

Another very important feature of commands is that they do not change the nature of  $\phi$ -ing itself. Commands do change  $\phi$ -ing's public relation to people's attitudes and it is this, among other things, that

---

<sup>26</sup> For more on the 'command model' of the law, see Gerald J. Postema, "Law as Command: The Model of Command in Modern Jurisprudence" II *Philosophical Issues* (2001): 470 – 501.

helps to make commands what they are.<sup>27</sup> So, authoritative commands generate reasons to  $\phi$  that are not grounded in the merits of  $\phi$ -ing. Legal theorists refer to the reasons generated by commands as *content-independent reasons*.<sup>28</sup>

Content-independent reasons appear to be widespread. People can rationally agree to submit a dispute to a common judge and thereby have reason to do whatever it is that the judge orders them to do with regard to that dispute. The authority of the reason to act as the judge orders is grounded in the subjects' agreement to submit to the judge's decision, not in the wisdom of doing whatever it is that the judge orders. This same kind of state of affairs can be realized within any scheme in which a practical authority would be reasonable (e.g., an orchestra in need of a conductor, a team in need of a coach or manager, a boat in need of a captain, etc.). Since there are often very good reasons for setting up practical authorities (such as that ignorance, partiality, and shortsightedness often make it prudent to take oneself out of the decision making procedure), we would expect content-independent reasons to be widespread.

---

<sup>27</sup> For more, See David Enoch, "Authority and Reason Giving" 89 *Philosophy and Phenomenological Research* 2 (2014): 296 – 332.

<sup>28</sup> There is a vast literature on the relationship between law and content-independent reasons. As noted above, the modern headwaters of this literature is H.L.A. Hart's work. See especially, in *Essays on Bentham*, "Commands and Authoritative Legal Reasons" and "Legal and Moral Obligation." For a useful overview, see Scott Shapiro, "Authority," in *The Oxford Handbook of Jurisprudence and Philosophy of Law*, edited by S. Shapiro, et al. (Oxford: Oxford Univ. Press, 2002): 382-439. See also George Klosko, "Are Political Obligations Content Independent?" 39 *Political Theory* 4 (2011): 498 – 523. Klosko explicitly discusses several other instances of content independent reasons, with special prominence given to promissory reasons as content independent.

I2.

So far, I've argued on functional grounds that our intentions inherit the authority of our deliberations. I then argued that one aspect of this authority is that intentions are reasons not to re-open deliberations about how to live. I argued that this is also a feature of the authority of commands. I then pointed out that commands also function as a content-independent reason to act. Since intentions share the backward-facing function of commands, why deny that they share the forward-facing function of commands? After all, deliberations that yield commands aim at a very similar thing as do deliberations that yield intentions, namely, settled plans about how to live. So, by analogy, I conclude that intentions function as commands one gives oneself on the basis of one's deliberations. Or, to put a Kantian spin on it: Intentions are forms of self-legislation. In virtue of this, it follows that intentions function as reason to act as intended.

We now have a complete picture of the normative function of deliberations and intentions: together, they function as reasons not to re-open deliberations, and as reasons to do as intended.

I3.

So, are our deliberations really authoritative or do they merely function as authoritative? This question can be answered by determining whether agents have the authority to govern themselves through their deliberations. If they do not, then deliberations and the intentions they yield function as reasons, but they aren't really reasons. If agents do have the authority to govern themselves through their deliberations, then the deliberations and the intentions they yield not only function as reasons, they are reasons.

I4.

The burden is not on me to establish that agents have the authority to govern themselves. The presupposition in favor of authority over oneself is as good as any starting point in practical philosophy. It is a lot more perverse to deny that we ought to be self-governing than to claim that intentions are reasons. Nonetheless, it is worthwhile to reflect briefly on some reasons why we ought to have the authority to govern ourselves.

First, we ought not be thoroughly deferential to others.<sup>29</sup> Even if we regularly invite and consider others' counsel about how to live, we ought to deliberate about whether to follow this counsel. Even when we abandon authority over ourselves, we should deliberate about whether to do so. And, when we do abandon authority over ourselves, we always reserve some authority to reconsider that decision and to take back control of our lives. This chimes with both the *prima facie* value of individual liberty and the supposed authority of reason. Both of these partially constitute the Enlightenment vision of self-creation through one's actions. On this view, people ought to develop their capacities to guide their lives via healthy deliberation (including developing habitual intentions as a result of healthy deliberations) at least partially because that is what it is to have a life of one's own. Furthermore, if we assume that one has both a good (but not perfect or even the best possible) grasp of the facts and a well-functioning capacity for deliberation and decision, then basic egalitarian commitments about the equal liberty of all suggest that it is *prima facie* best for each to exercise her own healthy, mature agency.

---

<sup>29</sup> For more, see Thomas E. Hill, Jr. "Servility and Self-Respect" in Thomas E. Hill, Jr., *Autonomy and Self-Respect* (Cambridge: Cambridge University Press, 1991), 4 – 18. For a fine discussion of this issue regarding moral judgment, see Julia Driver, "Autonomy and the Asymmetry Problem for Moral Expertise" 128 *Philosophical Studies* 3 (2006): 619 – 644.

One cannot object to this kind of authority simply by pointing to exceptions to its value. Circumstances requiring total submission to others' judgments about how to live do occur, but they are quite unusual. They typically involve stark and unfortunate limitations on an agent's capacity to decide for herself how to live. For example, total lack of information and terror immediately after being diagnosed with a frightening illness can be grounds for ceding authority over one's life to a loved one and a doctor. But, this exception does not support a general principle militating against such authority. Perhaps those who completely lack certain capacities ought to defer substantially more to others in order to determine how to live.<sup>30</sup> But, such cases – young children, the severely mentally disabled, etc. – are exceptions and not grounds for generalization to others in different conditions.

There are also cases, such as emergencies, when it is best to defer to others in order to achieve some valued end. Some might even extend such cases into the political and treat the state as serving this role.<sup>31</sup> Thus, one ought to defer to rules and laws when doing so will serve the good. This may be the way to go in certain cases, e.g., cases in which what is good is not controversial and how to achieve that good is a matter of rare or significant technical expertise. But, either when what is valuable is contested or when

---

<sup>30</sup> But see Jaworska, "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value" and Jaworska, "Caring and Full Moral Standing." But see Friedman, *Autonomy, Gender, Politics*: "If what someone adaptively prefers and chooses is behavior so servile that she ceases to act according to her own deeper concerns in any sense and becomes slavishly obedient to others instead, or becomes subject to their coercive interference with whatever subsequent choices she tries to make, then she loses autonomy in a content-neutral sense." (p. 25)

<sup>31</sup> For more, see David Estlund, *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2008).

many non-rival values can be jointly pursued, it is not obvious that deference is *required*. At the very least, a case must be made for deference in place of authority over self.

So, since we ought to have authority over ourselves, and since this is partially constituted by our intentions having authority – by their functioning as self-legislation which requires their functioning as content-independent reasons both not to re-open deliberations and to act as intended – *and* since our intentions *in fact* perform this function, then our intentions are reasons.

15.

Does this mean that intentions to act wrongly can be content-independent reasons to act wrongly? If so, there must be something wrong with this view. But, there is nothing particularly unique about this sort of problem. For, *all* sources of content-independent reasons for action, such as promises and the law, face this challenge. The natural way to respond to the problem of evil promises or evil commands is to carve out exceptions to the authority of such phenomena: evil promises may be promises but they are not reasons because they are evil, evil laws may be law but they are not reasons because they are evil, and so on. If this is the start of a philosophically acceptable response to this problem when it arises in relation to promissory obligation, legal obligation, and the like, then it should be an acceptable start in the present context.

Furthermore, promises and laws do not have to be evil to cease to have application as reasons for action. As already noted, sometimes in emergencies, one ought to cede authority over self to others. This point generalizes: there can be emergency-based suspensions of any normal source of authority. If there is an emergency and obeying the person most capable of properly organizing those affected would contravene a promise or the law, there is a strong moral case for treating those promises or laws as being overridden.

This same line of argument can be applied to intentions as reasons. We might argue that our intentions are no more undefeatable or non-overrideable reasons to  $\phi$  than are promises to  $\phi$  and laws requiring  $\phi$ -ing. If standing requirements of morality can trump promises and laws, then they can do the same to intentions-based reasons. So, the mystery of morally objectionable intentions being reasons to do wrong is no different than the mysteries of morally objectionable laws and morally objectionable promises being reasons to do wrong. The problems faced by this account of intentions are the same ones faced by *any* account of content-independent reasons for action.

In fact, this account of intention-as-reasons may be stronger in the face of this objection than many accounts of promises-as-reasons or laws-as-reasons. For, the authority of intentions rests on the value of authority over self. If we ought not be authoritative over ourselves when it comes to doing evil, then evil intentions ought not to function as reasons. The person who is so evil as to deliberate about how to be evil and then to decide to act evilly on the basis of those deliberations, has, as John Locke puts it in the *Second Treatise on Government*, “renounced reason, the common rule and measure [of] mankind” and so is more like “a lion or a tiger, one of those savage beasts with whom men can have no society nor security.”<sup>32</sup> This is why evil people are to be managed like dangerous wildlife. Furthermore, since this reason to manage evil people is agent-neutral and not agent-relative, evil people themselves have a reason not to treat themselves as authorities over themselves.<sup>33</sup>

---

<sup>32</sup> Locke, *Second Treatise*, §11.

<sup>33</sup> For more on this topic, see the classic essay Gary Watson, “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme,” in Fischer and Ravizza, *Perspectives on Moral Responsibility*, pp. 119–148.



Finally, unlike cases such as promises and rules in which suspension of the requirement to act is not subject to the authority of the promisor or the rule-bound subject,<sup>34</sup> each agent has authority over herself and so has the authority to abandon her own intentions. Earlier we asked when it would be appropriate to re-open deliberations about how to live. Here is one answer: one ought both to ask whether to re-open deliberations, and actually to re-open deliberations when one's intentions are evil (and certainly when one *knows* one's intentions are evil!).<sup>35</sup> This point applies to changes in circumstances as well. Such changes can require abandoning intentions. For example, one might intend to go in the evening to the gym. But, as the evening approaches, one starts feeling sick. If these circumstances support re-opening the question about whether one ought to go to the gym, then one has strong reason to abandon one's intention to go to the gym and to deliberate, once again, about whether to go to the gym. But, if circumstances *don't* change and one simply abandons one's intention out of laziness or distraction, then one is criticizable for having certain character defects: a lack of will-power or a lack of resolve, and so on.<sup>36</sup> At least part of the

---

<sup>34</sup> Except when the rule-bound agent is also a special kind of rule-applying agent such as a legislator or judge in a common law state.

<sup>35</sup> At this stage, my main difference with Bratman's views in "Intentions, Practical Rationality and Self-Governance" are at their starkest. Bratman argued that non-modifiable intentions to  $\phi$ , even evil intentions to  $\phi$ , that are essential to one's self-constitution are reasons to act on the intention to  $\phi$ . In contrast, I argue that only *non-evil* intentions to  $\phi$  that are product of sound deliberation, regardless of whether those intentions are modifiable (and in fact especially in the case of modifiable ones since those are the ones that are best expressions of our capacity for deliberation), are reasons to  $\phi$ . I am also able, unlike Bratman, to locate familiar partners in guilt for exceptions to the rule that intentions are reasons.

<sup>36</sup> For more, see Richard Holton, "Rational Resolve."

explanation for why these are defects is that they involve the agent failing to do what she has reason to do in virtue of her intention to do it.

I6.

It seems, then, that one important ground for denying that we ought to have authority over ourselves is when the deliberations from which the authority flows are somehow defective. This is unsurprising. If our capacity for deliberations about how to live are the headwaters of our agential authority, then if these waters are corrupted, so too is the capacity for agency. In these conditions, we ought not to be authorities over ourselves.

It is not easy to spell out the standards of non-defective deliberation.<sup>37</sup> What makes insane or corrupted deliberations insane or corrupted?<sup>38</sup> What counts as an irrational connection between deliberations and intentions? Small logical errors are probably not sufficient to render us insane or deeply irrational. This is especially the case since deliberations typically are enthymematic and so are sensible only on the basis of certain presuppositions by the agent. If some of those presuppositions are false, that

---

<sup>37</sup> That is the project of some of the essays in this volume.

<sup>38</sup> See, e.g., Richard Holton, "Intention and Weakness of Will," *Journal of Philosophy* 96 (1999): 241 – 262. For more on the problem of what makes some reasoning good and other reasoning bad (usually in the epistemic context), see, e.g., Paul Boghossian, "Blind Reasoning" *Proceedings of the Aristotelian Society (Supplementary Volume)*, 77 (2003): 225 – 248, David Enoch and Joshua Schechter, "How Are Basic Belief-Forming Methods Justified?," *Philosophy and Phenomenological Research* 76 (2008): 547–579, Ralph Wedgwood, "Primitively Rational Belief-Forming Processes," in Andrew Reisner and Asbjørn Steglich-Petersen (eds.) *Reasons for Belief* (Cambridge: Cambridge University Press, 2011), pp. 180–200, and Sharon Berry, "Default Reasonableness and the Mathoids," *Synthese* 190 (2013): 3695–3713.

doesn't thereby render the deliberations insane, corrupt or utterly irrational. This is so even if one's decisions are not the best possible ones. Even being in the grip of malformed preferences or some false beliefs doesn't make the agent's loss of authority over herself acceptable.<sup>39</sup> Unless those preferences and/or beliefs are especially troublesome (and perhaps not even then), it is almost always better for the agent to be in control of her own life.

So, while meeting at least some requirements of reasoning may be a necessary condition for the value of self-governance, those requirements cannot be so stringent as to make it such that almost anyone subject to some form of false consciousness or mistaken patterns of reasoning ought to be slaves to others' commands or to socially enforced patterns of activity. Achieving the impossibility of perfection in deliberation and decision cannot be a necessary condition for self-governance's controlling value.

So, we have two ends of a spectrum of deliberative health: on the least healthy end, we have a stream of thought composed of completely disconnected propositions yielding an intention to do something unrelated to that stream of thought, and on the healthy end we have a long train of reasoning involving true propositions, and whose form carefully respects whatever norms of rationality there may be. Where on this spectrum deliberation becomes so corrupted as to no longer count as being healthy is as difficult a question as the question of where on the spectrum of corporeal constitution a body becomes so diseased as to no longer count as healthy. There is not space here to resolve this issue. All we can conclude at this stage is that it is better than not to be an agent who deliberates in a more or less intelligible fashion

---

<sup>39</sup> Marilyn Friedman, in *Autonomy, Gender, Politics* (NYC: Oxford University Press, 2003), writes: "Even adaptively deformed preferences can be the basis of autonomous behavior if they represent what someone reaffirms as deeply important to her upon reflective consideration and she is able to act effectively on those concerns." (p. 25)

about how to live and ends those deliberations with intentions to act, where those intentions function as reasons not to re-open deliberations and as reasons to act as intended.<sup>40</sup>

I7.

A constitutive feature of authority over oneself is one's deliberations and intentions functioning as content-independent reasons not to re-open deliberations and as content-independent reasons to do as intended. Since we in fact have authority over ourselves, then our deliberations and the intentions they produce are content-independent reasons not to re-open deliberations and as content-independent reasons to do as intended. If I am right about all this, then, the rejection of bootstrapping therefore requires the rejection of the value of a person having authority over herself.

---

<sup>40</sup> We can still gain some insight on healthy agency by appreciating a few of its limits. For example, deliberating under the presupposition that one is trapped in one's life –that there are no alternatives to how one is living or has lived – goes some distance towards loss of the capacity for self-governance. In whichever domains this depressing view of life applies, one lacks a healthy capacity for self-governance. In these domains one cedes authority over one's life to facts of how one currently and has recently lived. Past patterns – facts about what one has done recently – unjustifiably become normative. But, that is simply to see authority over one's life as residing somewhere outside oneself.