

Michael Bukoski

Self-Validation and Internalism in Velleman's Constitutivism

[Forthcoming in *Philosophical Studies*. Please cite published version.]

Abstract

Metaethical constitutivists explain reasons or normativity in terms of what is (putatively) constitutive of agency. In Velleman's paradigmatic constitutivist theory, that is the aim of self-understanding. The best-known objection to constitutivism is Enoch's shmagency objection: constitutivism cannot explain normativity because a constitutive aim of agency lacks normative significance unless one has reason to be an agent rather than a "shmagent". In response, Velleman argues that the constitutive aim is self-validating. I argue that this claim is false. If the constitutive aim of self-understanding structures an agent's practical deliberation as Velleman describes, then some correctly deliberating agents will regard that aim as normatively arbitrary or unjustified. Moreover, I argue that the self-invalidation of the constitutive aim undermines or significantly qualifies Velleman's claim that his constitutivism reconciles internalism with a kind of objectivity about reasons. Internalists typically hold that motives from which an agent is alienated do not generate reasons, but in cases of self-invalidation, agents are alienated from the aim of self-understanding in just this way. Thus, the larger cost of the constitutive aim's possible self-invalidation is that Velleman's constitutivism does not satisfy a plausible and widely accepted constraint on an internalist account of reasons.

Constitutivism is a metaethical theory according to which one can explain reasons or normativity in terms of what is constitutive of agency. In David Velleman's (1989; 2000c; 2006a; 2009) constitutivist theory, what is constitutive of agency is the aim of self-understanding: every agent has this aim just by virtue of being an agent.¹ He argues that this aim best explains a number of otherwise puzzling features of agency, including our apparently non-observational knowledge of our intentional actions and the openness of the future from the perspective of the deliberating agent.² Reasons for action, Velleman argues, are thus considerations in light of which an agent's actions would make folk-psychological sense to her, or (equivalently) in light of

¹ Korsgaard (2009), Smith (2012), and Katsafanas (2013) also defend detailed constitutivist theories. Ferrero (2009) and Silverstein (2012, 2015) defend the general constitutivist approach, or versions of it, without endorsing a particular constitutivist theory.

² For the former, see, e.g., Velleman (2004a). For the latter, see Velleman (2000a).

which she could understand herself acting (Velleman 2009, pp. 18-20).³ The most prominent objection to constitutivism is David Enoch's (2006) *shmagency objection*. The objection is that, even granting that a certain aim is constitutive of agency, constitutivists cannot explain its normative significance because they cannot explain why we have reason to be agents rather than "shmagents," non-agents similar to agents but lacking the constitutive aim of agency. In response, Velleman (2009, pp. 138-145) argues that the aim of self-understanding is what I will call *deliberatively self-validating*: an agent whose deliberation is structured by the aim of self-understanding will inevitably regard that aim as justified. Thus, the aim of self-understanding is at least justified according to *its own* standard.

In this paper, I argue that although Velleman may be correct that the aim of self-understanding is deliberatively self-validating when considered in abstraction from other features of an agent's character and circumstances, it is sometimes self-invalidating when those features are taken into account. In other words, the aim of self-understanding will lead some correctly deliberating agents to regard that aim as normatively arbitrary or unjustified. Although it will continue to motivate them—because it is constitutive of agency and they are agents—they will not regard it as determining their reasons. The possibility of deliberative self-invalidation undermines Velleman's response to the shmagency objection. Moreover, although Velleman (2000d; 2009, pp. 118-120) argues that a virtue of his constitutivism is that it reconciles internalism and objectivity about reasons, I argue that the possibility of deliberative self-

³ Velleman (2009, pp. 185-206) describes a second mode of self-understanding that he calls narrative understanding. Narrative understanding is the kind of understanding provided by a good story, which Velleman explains in terms of characteristic patterns of emotional arousal and resolution. For simplicity, in this paper I focus on folk-psychological self-understanding, as Velleman also does.

invalidation leaves Velleman committed to an unattractive form of internalism.⁴ Internalists typically hold that aims or desires from which an agent is alienated do not provide her with reasons, but in cases of deliberative self-invalidation, agents are alienated from the aim of self-understanding in just this way.

1 The Shmagency Objection

Let's grant for the sake of argument that Velleman's theory of agency is correct: that is, that the aim of self-understanding is indeed constitutive of agency. Why does this apparently merely descriptive fact about agency provide the basis for an account of reasons? This question motivates Enoch's (2006) well-known shmagency objection to constitutivism. Enoch grants for the sake of argument that agency has a constitutive aim, but he argues that this aim lacks normative significance unless we have reason to be agents rather than "shmagents," non-agents that closely resemble agents but lack agency's constitutive aim. He (2006, p. 178) imagines someone deliberating about what she has reason to do:

She then finds out that some parts of her psychological makeup are unique in that they are such that without them she would not have qualified as an agent at all.

Knowing that, is she supposed to be relieved? Why does it matter, as far as the question of normative arbitrariness is concerned, that some parts of her psychology have this necessary-for-agency status? Why shouldn't our agent treat

⁴ The kind of internalism at issue is what Darwall (1983, p. 54) calls *existence internalism*. It says that one has a reason to do X only if one would be motivated to do it under specified conditions. Existence internalism contrasts with *judgment internalism*, according to which, necessarily, if one makes a sincere normative judgment (at least under specified conditions), then one is to some degree motivated to act accordingly.

the motives and capacities constitutive of agency as normatively arbitrary?

The result, Enoch argues, is a dilemma for constitutivism. If we do have reason to be agents, then constitutivism cannot explain normativity because it cannot, without circularity, explain that reason. The explanation would be circular if the argument were that, *according to the constitutive aim of agency*, we have reason to be agents. It would then presuppose rather than explain the normative significance of the constitutive aim. But if constitutivism cannot explain the reason to be an agent, or if we have no such reason, then constitutivism cannot be the correct account of reasons.

One constitutivist response is to argue that shmagency is impossible. If so, one might think, then whether to be a shmagent rather than an agent is not a live question, so we do not need a reason to be agents rather than shmagents (Rosati 2016, p. 201 n. 71). Moreover, if shmagents are impossible, then they have no reasons (or reason-like “shmreasons”), so constitutivists can explain all normative reasons in terms of the constitutive aim of agency (Silverstein 2015, pp. 1136-1138). Interestingly, one can reject the possibility of shmagency in either of two ways. First, one might say that everything with the “surface” features of an agent (e.g., the capacity for something like practical deliberation and intentional action) is *ipso facto* an agent, so if we grant that a certain aim is constitutive of agency, then we grant that everything with the surface features of an agent also has that constitutive aim. In that case, shmagency is conceptually impossible. If opponents of constitutivism showed how the surface features of agency could be realized without the (putatively) constitutive aim, they would have shown that that aim is not constitutive of agency after all, because some possible agents (i.e., these ones) do

not have it.⁵

Alternatively, one could begin with an account of the surface features of agency, argue (as Velleman does) that a certain constitutive aim best explains these features, and then identify the class of agents as those beings with that constitutive aim, while leaving open the conceptual possibility of beings that realize the surface features of agency by an alternative mechanism. Velleman (2009, p. 143) endorses this route when he argues that, if there were shmagents, their reasons would depend on the aim constitutive of shmagency. At the same time, the constitutivist may insist, as Velleman (2009, pp. 143, 204) also does, that in all likelihood nothing could actually realize the surface features of agency without having the constitutive aim of agency, so shmagents are not actually possible; or at least, the burden of proof is on the opponent of constitutivism to show otherwise.

Yet the possibility of shmagency is not, I think, essential to the shmagency objection. Constitutivists need not be—in fact, are not—committed to the claim that we have reason to be or to remain agents (Korsgaard 1996, p. 161; Silverstein 2015, pp. 1134-1136). They typically agree that the constitutive aim of agency can sometimes prescribe suspending or terminating one’s own agency, as when one falls asleep or commits suicide. Were shmagency possible, it would be no objection if the constitutive aim sometimes favored becoming a shmagent.⁶ Thus, constitutivists could accept the possibility of shmagency (understood in the second way above).⁷

⁵ Rosati (2016, p. 202) may be endorsing this approach when she writes that “[i]f, with respect to a particular conception of agency, one can coherently imagine being ‘very similar’ to an agent, while lacking what is claimed to be constitutive of agency, then that would cast doubt directly on the account of agency itself”.

⁶ Velleman (2009, p. 143) notes that, if we are agents, then the criterion for answering the question “Why be agents rather than shmagents?” is given by the constitutive aim of agency. See also Velleman (2009: pp. 204-206).

⁷ A qualification is that, if it is possible to become shmagents and we have reason to do so, then we may not in the end (i.e., after the transformation) have reason (or shmreason) to do what the constitutive aim of agency requires. For constitutivists who aim to justify morality (or another set of substantive normative requirements) by showing that it follows from the constitutive aim of agency, this “escapability” of the constitutive aim would be a

Instead, what ultimately motivates the shmagency objection is the question of why this apparently merely descriptive feature of agency has the normative significance that constitutivists assign it when they explain an agent's reasons in its terms. That question remains pressing even if shmagency is impossible. So although this first constitutivist response clarifies what it might mean for an aim to be constitutive of agency, it does not address the main worry underlying the shmagency objection.

However, Velleman (2009, pp. 135-145) develops a further two-part response to the shmagency objection. First, he agrees that asking for a justification for the normative significance of the constitutive aim of agency is reasonable. He argues that, if his constitutivist conception of agency is correct, then from the perspective of a deliberating agent, this justificatory question will be answered affirmatively. The constitutive aim of agency is self-validating: it prescribes regarding itself as justified. One might object that this justification is circular because it presupposes that the constitutive aim of agency provides the standard for justification, which is precisely what the anti-constitutivist denies. So the second part of Velleman's response is to argue that the circularity is not vicious because that aim being constitutive of agency distinguishes it from other self-validating aims that lack normative significance. Its constitutive status is non-circularly established by philosophical arguments in action theory, not by any sort of self-validation. Thus, Velleman claims that establishing the normative significance of a constitutive aim requires meeting two conditions: (1) showing that the aim is indeed constitutive of agency, and (2) showing that the aim is deliberatively self-

problem, at least if we assume that the constitutive aim of shmagency does not also justify morality. However, though this result would show that the constitutive aim of agency does not provide the firm foundation for morality that many constitutivists seek, it would not show that the constitutive aim is not normatively significant for agents while they are agents, and that is the claim that the shmagency objection is meant to contest.

validating. The circularity of self-validation is not objectionable because self-validation is only a necessary condition, not a sufficient one.

Enoch (2011, pp. 213-217) is unimpressed with this response, and with good reason. Velleman presents the constitutive character of the aim of self-understanding as sufficient for prima facie normative significance and deliberative self-validation as a further necessary condition. In other words, if the constitutive aim were not self-validating, then that failure would undermine its prima facie normative significance, but the aim's self-validation is not what establishes its prima facie normative significance: instead, its being constitutive of agency suffices for that. This response will not impress those who, like Enoch (2006, pp. 177-178; 2011, p. 210), are willing to grant for the sake of argument that a certain aim is constitutive of agency, but who deny that it has even prima facie normative significance. The debate thus appears to have reached a stalemate. One way to make progress would be to find new arguments for or against the claim that a constitutive aim of agency has prima facie normative significance. This route is difficult because the gulf of metaethical intuitions between constitutivists and their opponents is wide; hence the stalemate. An alternative for opponents of constitutivism is to shift their critique to the self-validation condition, which Velleman accepts. If that condition is not satisfied, then whether a constitutive aim has prima facie normative significance is irrelevant because Velleman's response to the shmagency objection fails on other grounds. That is what I will argue.

2 Deliberative Self-Validation

Why does Velleman think that the constitutive aim of self-understanding is deliberately self-validating? He makes two key claims, which I will call *No Detached Deliberation* and *Self-*

Validation. The first claim, *No Detached Deliberation*, is that any deliberation about whether to endorse the aim of self-understanding must proceed in light of that very aim. Velleman (2004a, p. 293) says:

[T]he agent cannot attain a perspective of fully detached reflection on the aim constitutive of action, because it is an aim that must be operative in order for him to reflect, in the first place.

In other words, because the aim of self-understanding is constitutive of agency and practical deliberation is an exercise of agency, it is not possible to evaluate the aim of self-understanding except through deliberation that is structured by that very aim. It structures deliberation by providing the ultimate criterion by which one's deliberation proceeds. The aim is “sub-agential”: it guides deliberation at a level below consciousness awareness (Velleman 2000b: 21). If you are deliberating about whether to take a certain action, what implicitly determines your answer is whether that action would make folk-psychological sense to you in light of your character and circumstances. If you could understand yourself acting in that way, then you will decide to do it; if not, then not. Now, just as one can deliberate about whether to perform a certain action, one can deliberate about whether to regard the aim of self-understanding as normatively significant or justified. Given *No Detached Deliberation*, the criterion for such deliberation is also provided by the aim of self-understanding. It thus provides the criterion for its own evaluation.

However, Velleman (2004a, p. 293) recognizes a possible problem:

Yet the agent's inability to withdraw from his intellectual drives does not entail

that he must approve of them, and it certainly does not entail that he must approve of them as that by appeal to which considerations qualify as reasons for action.

In other words, even if the aim of self-understanding provides the criterion for its own assessment, it is possible that the aim could fail that assessment. It is important to distinguish two kinds of justification here. One question is whether *having* the aim of self-understanding is justified. Do we have reason to have that aim rather than not? We granted Velleman's claim that the aim of self-understanding is constitutive of agency. If shmagency is impossible, then one will have an instrumental reason to be an agent if one has reason to do anything at all, because agency is a necessary condition for action. So it seems likely that, in most cases, because it makes folk-psychological sense for one to do something or other, now or in the future, the aim of self-understanding will prescribe having the aim of self-understanding. Moreover, we have already seen that it is no objection to constitutivism if the constitutive aim of agency sometimes prescribes suspending or terminating one's agency, as when one falls asleep or commits suicide.

The present question is whether the aim of self-understanding has non-instrumental normative significance—more precisely, whether it is, as Velleman says, “that by appeal to which considerations qualify as reasons for action.” Does it make folk-psychological sense to regard the aim of self-understanding as justified in this way?⁸ If not, then the constitutive aim of agency would be *self-effacing*: it would prescribe that the agent see that very aim as unjustified. A self-effacing normative standard is not always objectionable. Silverstein (2015, pp. 1134-1135) draws our attention to the so-called paradox of hedonism. Suppose that maximizing one's own

⁸ Unless otherwise noted, subsequent references to “regarding the aim of self-understanding as justified” refer specifically to regarding it as justified qua criterion for reasons, as opposed to regarding it as instrumentally (or otherwise derivatively) justified.

pleasure is the correct normative standard. It may be that self-consciously evaluating one's actions and commitments in this instrumental and egoistic way reliably causes one to be less happy than if one cared about other people and projects for their own sake. If so, then the hedonistic standard prescribes that one repudiate hedonism. That does not show that the hedonistic standard is false, but only that one has practical reason not to believe it.

Could one say, likewise, that if the constitutive aim of agency is not deliberately self-validating, that does not show that the constitutive aim is unjustified, but only that the agent has reason to regard it as unjustified? In principle, this response is unobjectionable. However, it is unavailable to the constitutivist in the present dialectical context. If the normative significance of the constitutive aim were already established, then its self-effacement would not undermine that significance. But what the shmagency objection questions is precisely whether a constitutive aim of agency *is* normatively significant. Velleman responds by arguing that the aim of self-understanding meets two conditions: (1) it is constitutive of agency and (2) it is deliberately self-validating. In this context, self-effacement would be a problem because it would mean that the constitutive aim does not meet Velleman's own self-validation condition. Moreover, the first condition (i.e., being constitutive of agency) is dialectically useless on its own, for the proponent of the shmagency objection already grants that the relevant aim is constitutive of agency. So the self-validation condition is essential to Velleman's response.

Velleman's (2009, p. 138) response to this threat is to invoke the second claim: *Self-Validation*. He writes:

The fact that you already have the aim constitutive of agency needn't blunt the force of your demand for a justification of it. Even if all of us share this aim, in

our shared capacity as agents, you can still ask whether we ought to have it, and why. To my ear, this question sounds like a demand for self-understanding. . . . So interpreted, the question demands that the constitutive aim of action be justified in relation to the criterion set by the aim itself. Such a justification is easy to give; for . . . trying to do what makes sense, makes sense for creatures like us, who are by nature driven to understand the world and are aware of ourselves as especially salient parts of it.

According to *No Detached Deliberation*, the criterion for whether the aim of self-understanding is justified is provided by that same aim. According to *Self-Validation*, the aim passes that test. Given that we have the aim of self-understanding, it makes folk-psychological sense to regard that aim as normatively significant or justified. Thus, according to Velleman, the aim of self-understanding is not only constitutive of agency, but also one that no correctly deliberating agent can fail to regard as justified upon reflection.

3 Local vs. Global Self-Validation

Are Velleman's two claims true? *No Detached Deliberation* follows from the claim that the aim of self-understanding is constitutive of agency, so we can grant it for the sake of argument. My target is *Self-Validation*, for Velleman's conclusion that the aim of self-understanding is deliberatively self-validating is rather quickly drawn.⁹ Let us distinguish between *local* and *global self-validation* of an aim. An aim is locally deliberatively self-validating if, if we abstract from other features of an agent's character and circumstances, an

⁹ As Enoch (2011, p. 221) also remarks.

agent whose deliberation is structured by that aim will, if deliberating correctly, conclude that that aim is justified. Some possible constitutive aims are not locally self-validating. For example, recall the hedonistic aim of maximizing one's own pleasure. If that aim is self-effacing given some ordinary assumptions about human psychology, then an agent who has that aim, but no other character or circumstances, should regard that aim as unjustified. However, the aim of self-understanding is plausibly self-validating under these conditions. If your only aim is to understand yourself, then it surely makes folk-psychological sense to endorse *having* the aim of self-understanding, because having that aim will generally make you more likely to achieve its object. Arguably, it also makes sense to regard the aim of self-understanding as justified. Such a person at least has no obvious reason to doubt or reject its normative significance.

What about global self-validation? An aim is globally deliberatively self-validating if an agent whose deliberation is structured by that aim will, if deliberating correctly in light of all relevant features of her character and circumstances, conclude that that aim is justified. Features of an agent's character and circumstances are relevant because they affect what it makes folk-psychological sense for one to do, and thus what the aim of self-understanding prescribes. As Velleman (2009, pp. 27-28) emphasizes, the aim of self-understanding is a second-order aim that shapes and guides the pursuit of our many first-order aims rather than replacing them with an exclusive first-order aim of self-understanding. For example, if you are thirsty, then it makes folk-psychological sense for you to get a drink—this is what we would expect you to do, and we would be surprised if you did not, all else equal—so getting a drink is something that you could understand yourself doing, so the aim of self-understanding motivates you to do it. Or if you are lazy, then it makes folk-psychological sense for you to procrastinate, so the aim of self-understanding would prescribe procrastination, unless some other feature of your character or

circumstances, such as your intimidating supervisor, makes it folk-psychologically intelligible for you to get to work instead. Thus, what the aim of self-understanding prescribes—that is, what it makes folk-psychological sense for one to do—largely depends on other aspects of one's character and circumstances. The question of global self-validation is the question of whether the aim of self-understanding is an aim that it makes sense to regard as justified, not only (as in local self-validation) in light of the aim of self-understanding itself, but also in light of one's character and circumstances more generally.

Global self-validation, not local self-validation, is what matters for Velleman's response to the shmagency objection. Velleman claims that every agent, if deliberating correctly, will regard the aim of self-understanding as justified. That conclusion follows from *No Detached Deliberation and Self-Validation*. Global self-validation collapses into local self-validation in the special case of someone whose sole aim is self-understanding—the *blank* agent, as it were, who is characterized by the constitutive aim of agency but nothing else. But virtually everyone has other features of their character and circumstances in light of which it makes folk-psychological sense for them to act in some ways rather than others. (Otherwise, everyone would always act identically.) So the relevant question is whether someone with character and circumstances in addition to the aim of self-understanding will, in her own correct deliberation about that aim, inevitably regard it as justified. That is the question of global self-validation.

Velleman does not explicitly consider whether the aim of self-understanding is globally deliberatively self-validating. Let us do so by first clarifying how, in Velleman's view, the aim of self-understanding evaluates possible actions, and then applying that apparatus to the question of whether to regard the aim of self-understanding as justified. We can distinguish four variables in practical deliberation: (1) the input, (2) the criterion, (3) the object of assessment, and (4) the

output. The *input* is the agent's character, including her desires, commitments, emotions, habits, and character traits, and her circumstances, including her immediate environment and her broader situation in life. The input also includes the aim of self-understanding itself, which every agent has. The *criterion* is also the aim of self-understanding, which Velleman interprets in terms of folk-psychological self-understanding and which (we have granted) structures an agent's practical deliberation. The *object of assessment* is whatever one deliberates about on a particular occasion. For example, you might deliberate about whether to procrastinate tonight, in which case you (implicitly) consider whether procrastinating makes folk-psychological sense given your character and circumstances. The fact that your deadline is a month away is a reason to procrastinate because it is a consideration in virtue of which procrastination makes folk-psychological sense. Finally, the *output* is the verdict reached by applying the criterion to the input and the object: that is, the verdict about whether the object makes folk-psychological sense, or (because making sense is a matter of degree) how much sense it makes. The aim of self-understanding then motivates the agent to act in the way that seems to her to make the most folk-psychological sense. If she deliberates correctly, then she acts in the way that actually does make the most folk-psychological sense for her to act, or in which she could best understand herself acting, given her character and circumstances.

Notice that the criterion of assessment is not instrumental rationality. Even in the case of local self-validation, the question is not exactly whether regarding the aim of self-understanding as justified is an effective means to achieving self-understanding. Instead, the question is whether regarding that aim as justified makes folk-psychological sense given that one has that aim. Instrumental rationality is relevant to folk-psychological intelligibility because taking the means to one's ends typically makes folk-psychological sense, but the former does not exhaust the latter

(Velleman 2009, p. 146 n. 33). Velleman (2004b, pp. 295-297) presents it as a virtue of his constitutivist account of reasons that it explains the role of instrumental reasoning while also making room for non-instrumental considerations that consequentialists ignore.

An example illustrates this process. Velleman (1989, pp. 254-257) considers a law student deliberating about whether to drop out to study English literature instead. This course of action is the object of assessment. The input is the person's character and circumstances. As Velleman (1989, p. 254) tells us, an academic career "better express[es] [his] deepest values," and his motives for becoming a corporate lawyer now strike him as "crass", although other traits, such as his contentiousness, pull in the opposite direction. Moreover, years of planning support his present position in law school, and following one's plans generally makes sense. The criterion of assessment is, as always, the aim of self-understanding. So we determine the output of correct deliberation by asking whether, for someone with this character in these circumstances, it makes folk-psychological sense to drop out of law school to study English. One way to answer this question is to ask whether it is predictable that this person would take that action. If it is, then it probably makes greater folk-psychological sense to drop out, so the aim of self-understanding would favor, and motivate, his doing so.

Let us now consider whether we can find a possible case in which the aim of self-understanding is not globally self-validating. This would be a case in which the object of assessment is the aim of self-understanding itself. We want to know whether it makes folk-psychological sense to regard that aim as normatively significant or justified, to "approve of [it] as that by appeal to which considerations qualify as reasons for action" (Velleman 2004a, p. 293). The criterion of assessment remains folk-psychological self-understanding. To show that the aim of self-understanding is not globally self-validating, the output in this case must be that it

makes sense to regard that aim as normatively arbitrary or unjustified. In other words, though one might not reject *having* the aim of self-understanding (because having that aim is a necessary condition of being an agent, and thus of acting at all), one might still reject the claim that that aim determines one's reasons. The only remaining variable is the input, so the question is whether a possible specification of the input (i.e., the agent's character and circumstances) has this result. Does it?

Here is one possibility. Imagine a committed utilitarian along the lines of the young John Stuart Mill. Her central project in life, to which she is deeply committed, is to maximize aggregate happiness. (She might believe that utilitarianism is the true moral theory, or she might take it merely as a personal ideal.) In light of that commitment, does it make folk-psychological sense for her to endorse the aim of self-understanding as having noninstrumental normative significance—that is, normative significance not derived from its contribution to maximizing happiness? I think not. For a committed utilitarian, it makes folk-psychological sense to evaluate objects by their contribution to aggregate happiness.¹⁰ For example, if she is deliberating about whether to vote for a candidate advocating an egalitarian economic policy (the object of assessment), then all else equal, it makes folk-psychological sense for her to do so if and only if she believes that so voting would maximize happiness. We would predict that the utilitarian would act in this way. If she did not, then we would be puzzled. We would try to explain her action by adjusting the inputs in our informal analysis of her situation, by calling attention to other features of her character or circumstances in light of which we should expect a different

¹⁰ Cf. Velleman (1989, pp. 304-305): “[I]f you think of yourself as a consequentialist, then that component of your self-conception will naturally affect how you deliberate. . . . [Y]ou will regard nonconsequentialist reasons as making no sense for you to consider, and you will shun them accordingly. You will take only consequentialist considerations into account, as the only ones that make sense for you.”

action from her. For example, maybe (unlike Mill) she is also rather lazy, in light of which it makes folk-psychological sense for her to sleep in rather than go to the polling station.

The aim of self-understanding is just another possible object of assessment, so we can determine the output of correct practical deliberation in the usual way. For the committed utilitarian, it makes folk-psychological sense, all else equal, to evaluate objects by their contribution to aggregate happiness. It makes sense for her to endorse *having* the aim of self-understanding for two reasons. First, assuming that Velleman's theory of agency is correct, she must have this aim in order to be an agent, and she must be an agent in order to take actions to promote happiness. So the aim of self-understanding has considerable instrumental value for her. Second, at least if she accepts a Millian conception of happiness as involving the exercise of her higher mental capacities, she needs to be an agent in order to have those capacities and thus to realize the higher pleasures that she regards as especially valuable. So the aim of self-understanding is valuable for her because it is a constituent part of what she regards as valuable. However, it does not make folk-psychological sense for her to regard the aim of self-understanding as non-derivatively normatively significant or justified; that is, as the criterion for reasons that Velleman claims it is. We would be surprised if she did. It would not make sense in light of her utilitarianism, just as it would not make sense for her to endorse egalitarian economic policies as intrinsically just rather than a means to maximizing happiness. Instead, all else equal, given her character, it makes folk-psychological sense for her to regard the aim of self-understanding as normatively arbitrary: that is, as a descriptive feature of agency that is derivatively valuable for the two reasons already mentioned, but not of fundamental normative significance. In this case, the aim of self-understanding is not globally deliberatively self-validating.

It does not follow that our utilitarian is no longer motivated by the aim of self-understanding. If that aim is genuinely constitutive of agency, then, “like it or not,” she will find herself doing what makes folk-psychological sense to her even when this fails to maximize aggregate happiness, such as when she lazily slept in rather than voting (Velleman 2009, p. 137). But she will not always *endorse* this influence. She will regard it as akin to an ineradicable (albeit generally desirable) desire that sometimes motivates her in ways of which she does not approve. Given her deep utilitarian commitments, it makes folk-psychological sense for her to regard that aim in this way. Thus, the aim of self-understanding motivates the deliberative rejection of its own normative significance.

One might object that, because the agent relies on the aim of self-understanding to guide her deliberation even as she reasons her way to regarding it as unjustified, she remains implicitly committed to its normative significance.¹¹ By repudiating it *because* doing so makes folk-psychological sense given her character and circumstances, she implicitly treats the aim of self-understanding as the criterion for determining what she has reason to do. However, this result is inevitable because the present charge is that the aim of self-understanding is deliberately *self*-invalidating, not that it fails to meet some external standard. Thus, although her conclusion is inconsistent with (in Velleman’s view) the implicit commitments of any practical deliberation, it is nevertheless the product of correct deliberation.¹² This result undermines Velleman’s response to Enoch’s shmagency objection. If it were sufficient to justify the aim of self-understanding that

¹¹ I thank an anonymous reviewer for raising this objection.

¹² One might worry that, if the agent regards the aim of self-understanding as unjustified, and if she knows that this aim structures her practical deliberation, then she should regard all of her practical conclusions as unjustified, including her endorsement of her utilitarian and other commitments. If so, then the self-invalidation of the constitutive aim could lead a correctly deliberating agent to general normative skepticism. Whether it does so depends on what it makes folk-psychological sense for someone with her character and circumstances to do in this situation, and that question may have no clear answer.

agents implicitly treat it as justified when they deliberate, then Velleman could have stopped with *No Detached Deliberation* and dispensed with *Self-Validation*. Instead, he argues that, although the aim of self-understanding is not justified by the sort of irreducibly normative standards that non-naturalist metaethicists like Enoch favor, it is justified according to itself. I have argued that this further condition is not always met.

Of course, utilitarian commitments are nothing special; many other possible inputs have the same result. One might be a nihilist for whom it makes folk-psychological sense to reject the normative significance of any aim, including the aim of self-understanding. One might be a theist for whom it makes folk-psychological sense to regard the aim of self-understanding as justified only if one believes that God wishes one to do so. And so on. These problematic inputs are not constitutive of agency, so the aim of self-understanding will not fail to be globally self-validating in every case. Some inputs will lead a correctly deliberating agent to regard the aim of self-understanding as justified. But that aim is globally self-invalidating in some cases, and that is enough to show that Velleman's defense of its normative significance does not succeed on its own terms, because the aim is supposed to be deliberatively self-validating as such, for all possible agents.

Indeed, this conclusion should be obvious. If Velleman's conception of agency is correct, then the reader's own practical deliberation is structured by the aim of self-understanding. So consider: do you endorse that aim as the criterion for normative reasons? Do you agree that you always have most normative reason to do or feel what it makes most folk-psychological sense for you to do or feel, given your character and circumstances? You most likely do not, because you most likely have some incompatible values or commitments in light of which rejecting the purported normative significance of the aim of self-understanding makes sense. So either

Velleman's conception of agency is false and your practical deliberation is not structured by the aim of self-understanding, or that aim is not globally self-validating.¹³ If the former, then Velleman's constitutivism obviously fails. If the latter, then the argument of this section is correct, and Velleman's theory does not satisfy *Self-Validation*.

4 General vs. Particular Self-Validation

I have argued that the aim of self-understanding is not globally self-validating because, for some possible sets of character and circumstances, it makes folk-psychological sense for an agent to regard that aim as normatively arbitrary or unjustified. These cases are counterexamples to Velleman's claim that the aim of self-understanding is deliberatively self-validating. Now, Velleman might concede that correctly deliberating agents can reject the normative significance of the aim of self-understanding considered thus abstractly, but insist that they will nevertheless endorse the action it prescribes on each particular occasion. In other words, an agent might claim to doubt the normative significance of the aim of self-understanding, but in practice she will inevitably agree that, for each action that aim prescribes, she has most reason to perform that action. Call the former *general self-validation* and the latter *particular self-validation*.

Whether particular self-validation without general self-validation would satisfy the self-validation condition is unclear. However, we can set aside that issue because, as I will argue, Velleman's constitutivism does not guarantee particular self-validation either. An aim is particularly self-validating if an agent whose deliberation is structured by that aim endorses each

¹³ Actually there is a third possibility: you might be deliberating incorrectly. The observed behavior of actual agents is only a heuristic in this context because an agent might mistakenly regard the aim of self-understanding as unjustified even though regarding it as justified would make more sense given her character and circumstances. My earlier argument is not vulnerable to this objection, however, because that argument was based on directly applying the aim of self-understanding to some possible cases. Unless my arguments about what it makes most sense for them to do are mistaken, the agents in my examples are deliberating correctly.

action that it prescribes, in the sense that she regards herself as having most reason to perform that action (rather than some alternative action). Thus, the aim of self-understanding is particularly self-validating if it makes folk-psychological sense to endorse acting in any way that it makes folk-psychological sense to act. This condition may seem easy to satisfy. For won't whatever features of one's character and circumstances in light of which it makes sense to perform an action also be features in light of which it makes sense to endorse performing that action?

We can approach the problem by way of an objection that Velleman (2009, pp. 31-33) considers to his account of reasons. The objection is that it sometimes makes folk-psychological sense to act in ways that manifest one's "admitted vices," such as laziness, even though one does not have reason so to act (2009, p. 31).¹⁴ For example, it makes folk-psychological sense for a lazy person to procrastinate, but it surely does not follow that she has a reason (let alone *most* reason) to procrastinate, just because she is lazy.

Velleman argues that alleged counterexamples of this sort divide into two categories, neither of which is ultimately problematic. First, the agent might endorse the character trait in question, in which case she would not characterize herself negatively as "lazy" but instead positively or neutrally as (e.g.) "easygoing". Velleman regards it as no objection to his account that an easygoing agent has reason to take time off rather than working hard, all else equal. Second, the agent might indeed evaluate her character negatively as "lazy". But then whatever explains her negative evaluation—her professional ambition, say, which her laziness frustrates—will also be such that it makes folk-psychological sense for her to overcome her laziness rather

¹⁴ Even someone who thinks that an agent's reasons depend on her ends or desires should agree that lazy agents do not always have most reason to be lazy. When laziness interferes with an agent's pursuit of her own more valued ends, it is a form of instrumental irrationality.

than procrastinating. In that case, Velleman's theory avoids the counterintuitive implication that the lazy person has most reason to procrastinate. So purported counterexamples about the folk-psychological implications of bad character traits resolve upon closer inspection into either of two unobjectionable scenarios.

However, this treatment fails in some cases. It fails because whether to perform an action and whether to endorse that action are different objects of assessment and a given input of character and circumstances might be such that it makes folk-psychological sense to perform an action but also to regret rather than to endorse it. This discrepancy is possible because the folk-psychological connection between someone's character and her actions is not always mediated by endorsement. Sometimes it is. In some cases, it makes folk-psychological sense for someone with a certain character trait to perform a certain action because someone with that trait would *endorse* that action, and people typically act in ways that they endorse. Thus, an ambitious person typically endorses working hard, so it makes sense for her to work hard. But in other cases, it makes folk-psychological sense for someone with a certain character trait to perform a certain action simply because someone with that trait would be *motivated* to perform that action (without endorsing it). For example, it makes sense for a socially anxious person to avoid stressful social situations such as parties or dates even if she does not endorse acting in this way and wishes that she could overcome her anxiety and be more sociable instead.

Because the folk-psychological connection between character and action can be mediated by motivation without endorsement, a “bad” character trait can motivate an action even though a “good” character trait leads one to regard that action as unjustified. For example, someone's laziness might have a strong motivation-mediated folk-psychological connection with procrastination but no endorsement-mediated connection, while her professional ambition has

only a weaker endorsement-mediated connection with working hard. All else equal, it makes most folk-psychological sense for her to procrastinate and then to berate herself for laziness.¹⁵ Or someone might be strongly motivated to help others even though she wishes that she were not so self-sacrificing, with the result that it makes most folk-psychological sense for her to help, against her better judgment, and then to regret her soft-heartedness. That this sort of behavior is, by all appearances, not particularly unusual among the actual agents we encounter provides some confirmation of the analysis. So Velleman's account of reasons does not avoid the counterintuitive implication that people sometimes have most reason to act in ways that manifest their "admitted vices".

By itself, this result is some reason to reject Velleman's constitutivist account of reasons: it has some very implausible implications about the reasons people have. What are the implications for our question about particular self-validation? We were considering the objection that, although the aim of self-understanding may not be globally self-validating when considered generally, it remains particularly self-validating in the sense that, for any action it prescribes, a correctly deliberating agent will regard that action as justified. The preceding discussion shows that this claim is false because, with the right input, the aim of self-understanding can prescribe performing one action but simultaneously prescribe refusing to endorse that action (and endorsing a different action instead). Recall our lazy utilitarian. It may make folk-psychological sense for her to stay in bed rather than going to the polling station even if she thinks that voting is the happiness-maximizing action in her circumstances. But it may also make folk-

¹⁵ Velleman (2009, pp. 32-33) also argues that someone in this conflicted state could make her life simpler and therefore more favorable to self-understanding if she eliminated her laziness. But this move does not help against the objection because it leaves intact the objectionable claim that the agent has most reason to be lazy *now*, given her present character. Moreover, it may not be possible for her to become less conflicted if she is in what Velleman (2009, p. 33) calls a "rational dead end," in which making the effort required to reform her character does not make folk-psychological sense in light of her present, lazy character.

psychological sense for her to regard her laziness as a vice or failure, not something to which she has most reason to accede, though the aim of self-understanding motivates her to stay in bed regardless. And she may feel this way about many other actions that the aim of self-understanding motivates her to perform.

Thus, not only does the aim of self-understanding lead some possible agents to regard that same aim as normatively arbitrary or unjustified in general (so it is not generally self-validating), it also leads some possible agents to regard some of the particular actions that the aim of self-understanding prescribes and motivates them to perform on particular occasions as unjustified. As before, utilitarianism and laziness are nothing special. Similar outputs follow from many different inputs of character and circumstances.

5 Self-Validation and Internalism

So far I have argued that, although Velleman claims that the aim of self-understanding is deliberately self-validating, this is not the case for some possible (and actual) agents. Although such agents continue to be motivated by that aim—because we are granting that it is constitutive of agency, it could cease to motivate them only if they ceased to be agents—they do not regard it as normatively significant. Instead, they believe that the aim of self-understanding sometimes leads them to act or feel in ways that they do not have reason, or at least most reason, to act or feel. They wish that they could act differently. Moreover, this attitude is not irrational. According to the criterion set by the aim of self-understanding, the role of which in structuring an agent's deliberation we have granted, regarding that aim as normatively unjustified is what these agents have most reason to do, for this is what it makes folk-psychological sense for them to do, given their character and circumstances. That the aim of self-understanding is deliberately self-

invalidating in some cases is a problem for Velleman because he relies on *Self-Validation* as one part of his response to Enoch's shmagency objection.

However, one might question how damaging the prospect of deliberative self-invalidation really is. If it undermines one argument for constitutivism, could the constitutivist simply give another? For example, Silverstein (2015) rejects the self-validation condition. He acknowledges (2015, pp. 1134-1135) that deliberation could lead an agent to the conclusion that she should not be an agent in general or on some particular occasion.¹⁶ But he claims that the metaphysical status of the constitutive aim (i.e., as constitutive of agency) secures its normative significance regardless of whether the aim is self-effacing in practice. Self-validation is "beside the point" (Silverstein 2015, p. 1142 n. 9). Moreover, Velleman himself offers other arguments for constitutivism. He claims that only constitutive standards can avoid either an endless regress of justification or an arbitrary halt (2009, pp. 125-128) and, against metaethical non-naturalism, that the idea of intrinsically justified standards of correctness for action is incoherent (2009, p. 145).¹⁷ He also argues (2004b, pp. 295-297) that his account of reasons is more extensionally adequate than instrumentalist alternatives.

The merits of other arguments for constitutivism are beyond the scope of this paper. However, I will argue that the prospect of deliberative self-invalidation is a theoretical liability for constitutivists who, like Velleman, favor an internalist conception of reasons. Internalists (e.g., Williams 1981) say that an agent's reasons depend in one way or another on her subjective

¹⁶ I am not sure whether Silverstein has in mind someone who doubts the normative significance of the constitutive aim or merely someone who doubts that she has reason to remain an agent (which is compatible with recognizing the normative significance of the constitutive aim given that she *is* an agent). But I take him to be committed to the view that the former sort of doubt is also no threat to constitutivism: "the authority of agency's constitutive norm does not require the validation of any norm—even itself," because "it depends instead on the deep metaphysical connection between agency and normativity" (2015, p. 1141).

¹⁷ For criticism of the latter claim, see Enoch (2011, p. 223-227).

motivational states, such as her aims or desires. An advantage of internalism is its ready explanation for why agents are so often motivated by their reasons rather than indifferent to them. Its disadvantage, at least in the eyes of many (e.g., Parfit 1997), is that because people's aims and desires are contingent and interpersonally variable, the reasons they generate will have those same features, with the result that some possible agents do not have the moral or prudential reasons that, intuitively, everyone does have. An attraction of constitutivism, according to Velleman (2000d; 2009, pp. 118-120), is that it reconciles internalism with a kind of objectivity about reasons.¹⁸ Constitutivism is an internalist theory because it explains reasons in terms of a motivating aim, but it avoids the characteristic disadvantage of internalist theories by appealing to an aim that is not contingent or interpersonally variable but instead constitutive of agency itself. Thus, Velleman argues that we should accept a constitutivist account of reasons at least in part because it explains how reasons can be both internal and objective.¹⁹

What is the problem? What we might call *simple internalism* is the view that an agent has a reason to do X if and only if (1) she has an intrinsic pro-attitude toward doing X or (2) doing X is instrumental to something toward which she has an intrinsic pro-attitude. Different internalist theories identify different pro-attitudes as relevant; familiar candidates are desires (e.g., Brandt 1979) and second-order desires (e.g., Lewis 1989, pp. 115-116). Most internalists reject simple internalism in at least two ways. First, they argue that some pro-attitudes that an agent does not actually have nevertheless provide reasons for her because she *would* have them after some

¹⁸ Other proponents of constitutivism (e.g., Katsafanas 2011, pp. 625-628) as well as critics (e.g., Enoch 2011, pp. 208-209) have cited the reconciliation of internalism and objectivity as an important motivation for constitutivism. Korsgaard's (1996, pp. 16-18) project of finding a justification for morality that succeeds in addressing every agent by virtue of appealing to considerations related to each agent's own identity reveals, to my mind, a similar ambition. Korsgaard (2009) develops her view in an explicitly constitutivist direction.

¹⁹ Velleman's accommodation of traditional externalist views is limited insofar as he claims that the aim of self-understanding is "merely pro-moral," encouraging the development of moral ways of life without guaranteeing that every agent always has reason to act morally (2009, p. 2). See also Velleman (2006b).

process of idealization. For example, she might have those pro-attitudes if she were fully informed and rational (Smith 1994). Second, they argue that some of an agent's actual pro-attitudes do not generate reasons—for example, if they result from inadequate information or mistaken reasoning. What matters here is that another typical basis for regarding an agent's actual pro-attitudes as not reason-generating is that she is *alienated* from them.

The classic example of alienated desire is the unwilling drug addict (Frankfurt 1971, pp. 12-13). The unwilling addict desires her drug, but she repudiates that desire: she wishes that she did not have it and, typically, prefers not to satisfy it (except perhaps for instrumental reasons, such as to relieve the pain and distraction it otherwise causes). It does not seem plausible to say that she has a (non-instrumental) reason to take the drug. Accordingly, internalist or subjectivist theories of all sorts, including well-being (e.g., Dorsey 2015, p. 6), value (e.g., Lewis 1989, p. 115), and reasons (e.g., Hubin 2003, pp. 326-329; Frankfurt 2004, pp. 15-16) generally exclude alienated pro-attitudes from the set of pro-attitudes that they take to generate reasons, either by explicit stipulation or, more commonly, by explaining reasons in terms of those pro-attitudes (e.g., second-order desires instead of first-order desires, or values instead of desires) taken to represent an agent's non-alienated interests.

This restriction also follows from an important philosophical motivation for internalism: the thought that an agent's reasons, good, or well-being should “fit” her in such a way that she finds it engaging rather than alienating. Railton (1986, p. 9) gives the canonical statement:

[W]hat is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware. It would be an intolerably alienated conception of someone's good to

imagine that it might fail in any such way to engage him.

Call this thesis, generalized to apply to reasons, the *non-alienation condition*. The unwilling addict may find the prospect of taking her drug motivationally compelling, but she does not find it normatively compelling or attractive. What she finds attractive is the prospect of being rid of that desire. So if internalism about reasons is motivated at least in part by the non-alienation condition, then excluding alienated pro-attitudes from the set of pro-attitudes that generate reasons is justified.²⁰ In a similar way, Hubin (1999) defends internalism on the grounds that it best explains the normative “force” of reasons for agents, by speaking to the concerns of those agents rather than trying to browbeat them with considerations they regard as irrelevant. From the perspective of the unwilling addict, a putative reason to take the drug does not have normative force—she sees taking the drug as compelled, not justified—so Hubin’s argument also justifies the claim that pro-attitudes from which an agent is alienated do not generate reasons, as he argues elsewhere (Hubin 2003, pp. 326-329). At the least, even if alienated pro-attitudes can generate some reasons, they should not determine what an agent has *most* reason to do. They should not outweigh her non-alienated interests, at least in important cases.

Now, the problem for constitutivism is that, if the constitutive aim of agency is sometimes deliberately self-invalidating, then some agents will be alienated from that aim in the relevant sense. Recall our lazy utilitarian. Like the unwilling addict, she feels the motivational force of

²⁰ Some influential early statements of internalism (e.g., Williams 1981; Korsgaard 1986) do not explicitly rule out the possibility that alienated pro-attitudes generate reasons. These authors may have simply not considered this issue, or they may have omitted discussion of it to focus on the larger contrast between internalism and externalism. Williams is moved by concerns about alienation in other contexts (e.g., 1973, pp. 116-117), and Korsgaard (1996, pp. 93-94) is very clear that, at least in her own view, desires not endorsed by the agent do not generate reasons. In any case, what matters is that the non-alienation condition is plausible and widely accepted among contemporary internalists.

the aim of self-understanding, but she sometimes wishes that she could act otherwise. She differs from the addict in that, because she wants to continue being an agent, she does not want to stop having that aim. Her alienation is not thoroughgoing. But she is nevertheless alienated from it to the extent that it leads her to act in ways that she regards as unjustified, such as when she lazily stays in bed rather than doing what she regards as more important work. Staying in bed does not “fit” her; it does not strike her as normatively compelling or attractive; it is at odds with her values. She is still motivated to do it, but so is the addict motivated to take her drug. To say that she has most reason to stay in bed must strike her as perverse.

Thus, in cases of deliberative self-invalidation, the reasons generated by the aim of self-understanding should not be regarded as internal reasons. It follows that Velleman’s account of reasons is not an internalist account, or at least not one that satisfies the non-alienation condition. Nor is there an obvious alternative for Velleman. It would not be plausible to say that agents alienated from the aim of self-understanding have no reasons. And because Velleman takes the constitutive status of an aim to be essential to its normative significance, he cannot instead explain these agents’ reasons in terms of other (non-constitutive) pro-attitudes from which they are not alienated.

So the deeper threat posed by deliberative self-invalidation is that it undermines what many, including Velleman, regard as an important virtue of constitutivism: its promised reconciliation of internalism and objectivity. If agents can be alienated from the aim of self-understanding, then Velleman’s constitutivism does not yield a plausible form of internalism, which must (among other things) satisfy the non-alienation condition.²¹ Of course, even if

²¹ Velleman’s constitutivism violates the non-alienation condition in a second respect as well. Recall the objection that Velleman’s account of reasons implausibly allows “bad” character traits such as laziness to provide reasons

successful, this argument is not a refutation of constitutivism as such. My argument has been restricted to the aim of self-understanding, so the putatively constitutive aims of agency favored by other constitutivists, such as Korsgaard (2009) and Katsafanas (2013), may indeed be deliberately self-validating. That question is beyond the scope of this paper. Constitutivists could also defend a form of internalism that rejects the non-alienation condition, or reject internalism altogether. The limited conclusion of this section is that those who find internalism attractive at least in part on the basis of the non-alienation condition (1) should not be impressed by Velleman's claim that his constitutivism reconciles internalism and objectivity and (2) should require, of other internalist constitutivist theories, that they show how their proposed constitutive aim avoids deliberative self-invalidation.²²

Acknowledgements I am grateful to David Enoch, Connie Rosati, Mark Timmons, and an anonymous reviewer for very helpful feedback on this paper. Any mistakes remain my own.

to act in ways that manifest those traits. A version of this objection applies to character traits from which the agent is alienated. For example, although the unwilling addict is alienated from her desire for the drug, it may nevertheless make folk-psychological sense for her to take the drug, because the alienated desire has (what I called) a strong motivation-mediated folk-psychological connection with that action. So even if an agent is not alienated from the aim of self-understanding itself, she may be alienated from other motives that, in conjunction with that aim, generate the reasons that Velleman's account says she has.

²² As some already do; e.g., Katsafanas (2013, pp. 204-207).

References

- Brandt, R. B. (1979). *A theory of the good and the right*. New York: Oxford University Press.
- Darwall, S. L. (1983). *Impartial reason*. Ithaca, NY: Cornell University Press.
- Dorsey, D. (2015). Idealization and the heart of subjectivism. *Nous*, DOI: 10.1111/nous.12130.
- Enoch, D. (2006). Agency, shmagency: why normativity won't come from what is constitutive of agency. *Philosophical Review* 115(2), 169-198.
- Enoch, D. (2011). Shmagency revisited. In M. S. Brady (Ed.) *New waves in metaethics* (pp. 208-233). New York: Palgrave Macmillan.
- Ferrero, L. (2009). Constitutivism and the inescapability of agency. In R. Shafer-Landau (Ed.) *Oxford studies in metaethics*, vol. 4 (pp. 303-333). New York: Oxford University Press.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1), 5-20.
- Frankfurt, H. G. (2004). *The reasons of love*. Princeton, NJ: Princeton University Press.
- Hubin, D. C. (1999). What's special about Humeanism. *Nous* 33(1), 30-45.
- Hubin, D. C. (2003). Desires, whims, and values. *Journal of Ethics* 7(3), 315-335.
- Katsafanas, P. (2011). Deriving ethics from action: a Nietzschean version of constitutivism. *Philosophy and Phenomenological Research* 83(3), 620-660.
- Katsafanas, P. (2013). *Agency and the foundations of ethics: Nietzschean constitutivism*. New York: Oxford University press.
- Korsgaard, C. M. (1986). Skepticism about practical reason. *Journal of Philosophy* 83(1), 5-25.
- Korsgaard, C. M. (1996). *The sources of normativity*. New York: Cambridge University Press.
- Korsgaard, C. M. (2009). *Self-constitution: agency, identity, and integrity*. New York: Oxford University Press.

- Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society, Supplementary Volumes* 63, 113-137.
- Parfit, D. (1997). Reasons and motivation. *Proceedings of the Aristotelian Society, Supplementary Volumes* 71, 99-130.
- Railton, P. (1986). Facts and values. *Philosophical Topics* 14(2), 5-31.
- Rosati, C. (2016) Agents and “shmagents”: an essay on agency and normativity. In R. Shafer-Landau (Ed.) *Oxford studies in metaethics*, vol. 11 (pp. 184-213). New York: Oxford University Press.
- Silverstein, M. (2012). Inescapability and normativity. *Journal of Ethics and Social Philosophy* 6(3), 1-26.
- Silverstein, M. (2015). The shmagency question. *Philosophical Studies* 172(5), 1127-1142.
- Smith, M. (1994). *The moral problem*. Malden, MA: Blackwell Publishing.
- Smith, M. (2013). A constitutivist theory of reasons: its promise and parts. *Law, Ethics, and Philosophy* 1, 9-30.
- Velleman, J. D. (1989). *Practical reflection*. Princeton, NJ: Princeton University Press.
- Velleman, J. D. (2000a). Epistemic freedom. In *The possibility of practical reason* (pp. 32-55). New York: Oxford University Press.
- Velleman, J. D. (2000b). “Introduction.” In *The possibility of practical reason* (pp. 1-31). New York: Cambridge University Press.
- Velleman, J. D. (2000c). *The possibility of practical reason*. New York: Oxford University Press.
- Velleman, J. D. (2000d). The possibility of practical reason. In *The possibility of practical reason* (pp. 170-199). New York: Oxford University Press.
- Velleman, J. D. (2004a). Précis of 'the possibility of practical reason'. *Philosophical Studies*

121(3), 225-238.

Velleman, J. D. (2004b). Replies to discussion on *the possibility of practical reason*.

Philosophical Studies 121(3), 277-298.

Velleman, J. D. (2006a). *Self-to-self: selected essays*. New York: Cambridge University Press.

Velleman, J. D. (2006b). Willing the law. In *Self-to-self: selected essays* (pp. 284-311). New York: Cambridge University Press.

Velleman, J. D. (2009). *How we get along*. New York: Cambridge University Press.

Williams, B. (1981). Internal and external reasons. In *Moral Luck* (pp. 101-113). New York: Cambridge University Press.

Williams, B., and Smart, J. J. C. (1979). *Utilitarianism: for and against*. New York: Cambridge University Press.