abstracta Linguagem, Mente & Ação ISSN 1807-9792 Special Issue VII, pp. 17–21, 2014

Velleman on the Work of Human Agency

Tamar Schapiro

Stanford University
schapiro@stanford.edu

The disagreements I have with *How We Get Along* emerge against a backdrop of deep agreement. I admire Velleman's project of trying to understand the nature of the pressure that merits the name "practical reason." I also agree with him that self-consciousness is an essential feature of human agency, a feature that somehow figures into the explanation of why we are susceptible or subject to the force of practical reason. In addition I share his view that our agency is something we have to continually work to realize, and that the way we do this is by guiding ourselves according to its constitutive aim. And I find compelling the idea that somehow the realization of my agency both depends on and contributes to the realization of yours. In these brief comments, I will focus primarily on the third of these points, the idea that agency requires some sort of continual work. I want to know more about how the moral psychology underlying Velleman's view explains both the need for this work and its nature.

Velleman explicitly locates himself within the tradition in practical philosophy that sees task of agency as one of "overcoming doubleness." (89) But he takes his predecessors to have diagnosed the "characteristic failure of agency" incorrectly. According to Velleman, they conceived of the basic problem of agency as motivational conflict. Plato, Velleman writes, held that fundamental challenge is to avoid "dissension between the divisions of [the] soul," while Frankfurt saw it as a matter of avoiding the ambivalence that would result were we to fail to "identify decisively with either of the warring parties." (90) Velleman's main reason for denying that they have identified the characteristic failure of agency is that conflict and ambivalence are problems even nonhuman agents face. The role of the characteristic failure of agency, whatever it is, is to generate motivational pressure to avoid it, and it is this pressure that will ultimately explain our susceptibility to practical reasons. If nonhuman animals face the same pressure, then it cannot count as the source of specifically rational pressure.

Velleman goes on to suggest that the fundamental problem is not to overcome conflict or ambivalence but to achieve authenticity. The fundamental work of agency does not result in resoluteness about what to do, but rather in authenticity with respect to whatever response we might have to the situation we face. He writes,

> What is both peculiar to rational agents and a pitfall for them, I think, is reflective inauthenticity. Only a rational agent can play false with himself; and playing false with himself is the characteristic failure of agency – a failure of the psychological mechanisms by which agency normally works. (90)

Now to say that Plato is concerned with resoluteness rather than authenticity might be a bit unfair to Plato. The city-soul analogy in the *Republic* arguably aims to show how the members of the polity can act as one, where the emphasis is not only on the idea of coming

to resolution about what to do, but more fundamentally on the idea of unifying subagential sources of authority so that whatever one does carries one's authority as a whole agent. So if achieving authenticity is a matter of unifying subagential sources of authority, then this is indeed what Plato took to be the fundamental challenge. But Velleman does not conceive of the problem of authenticity this way. What is inauthenticity, on his view, and why does it constitute the characteristic failure of agency?

In these lectures, Velleman draws heavily on a theatrical analogy to illustrate what he has in mind. He claims that as agents, we are like actors improvising the characters that are ourselves. This is a powerful metaphor, because it captures the idea that our characters are, in some important sense, up to us.[1] But at the same time it suggests that despite this freedom, we are answerable to standards. There is a difference between good and bad improvisation. And the standard of improvisation is in some sense authenticity. To improvise your own character well, you have to actually inhabit your character as far as possible, being yourself even as you create yourself. You have to actually have the thoughts, feelings, and motives that you enact, and your way of enacting them has to actually reflect or express the person you are making yourself into.

There is a notion of integrity here, but the idea is not that we have to unify subagential sources of authority to achieve it. Nor is it the somewhat distinct but still familiar notion of integrity as a matter of simply living up to a normative self-conception associated with a freely chosen role or ideal (a "practical identity" in Korsgaard's terminology). (16, nt. 8) Rather, the idea is that we have to bring ourselves into accord with our enacted conceptions of ourselves, even as we bring our enacted conceptions of ourselves into accord with ourselves. The striving goes in both directions, so to speak. Importantly, the sense of "conforming" here is descriptive/explanatory rather than normative. We have to make sense of ourselves to ourselves, and we can do that both by shaping our self-conceptions to fit persons we are, and by shaping the persons we are to fit our self-conceptions. The point is not so much to be whole as to be true to yourself, and the sense of being true is not so much that of being faithful to your principles as it is of being faithful to the facts about yourself even as you inevitably transcend them.

Velleman wants us to see that we routinely employ something like this standard when we go to the theater. We hold the actors, improvisational or not, to a standard that he characterizes as one of "intelligibility." Every player has to act "in character," meaning, he has to think and feel and do the things that it would make sense for him to do, given the way he has established his character up to this point. That we apply this standard to actors may just be a matter of the conventions of theatergoing, but Velleman suggests that it is rooted in, or at least that it mirrors, something deeper about the human condition. Our nature as self-conscious creatures, he claims, makes us audiences to ourselves. As such, we necessarily hold ourselves to the standard of acting in character, despite the fact that our characters are not scripted for us in advance.

Velleman goes on to argue that pressure to conform to this standard underlies and explains the normative pressure of practical reason. I am going to set aside this ambitious claim, in order

[1]What sort of agent counts as the analogue of a nonimprovisational actor? It might seem natural to say that creatures of instinct do, because their instincts constitute their characters, and they do not have reflective distance on those instincts. They do not have to create their characters the way we do. But Velleman's use of Anscombe to define a distinction between what he calls "reading" and "writing" a character's actions suggests to me that this is not his line of thought. (132) This makes it sound as though the analogue of the nonimprovisational actor would be an agent who does not have to work to achieve self-knowledge, because he already knows what he is going to do. Granted that this is not a real possibility, is it even conceptually coherent? And if no such agents could exist, then what sort of freedom is the idea of "improvisation" supposed to pick out metaphorically?

to focus on the step from being self-consciousness to being audiences to ourselves. Velleman's account of why we are, by nature, audiences to ourselves is this. First, we have a "voracious appetite" to understand things, where understanding is something we do, an activity of "making sense," rather than an experience we passively undergo. This appetite, which we call "theoretical reasoning," is an "intellectual drive" that we share as human beings.[2] Second, among the things that appear to us as objects of this intellectual drive are ourselves. We naturally strive to make sense of ourselves. It is therefore in this sense that we are naturally audiences to ourselves. (17)

Suppose this much is true. Would we then be audiences to ourselves in the sense required by the theatrical analogy? The theatrical analogy seems to presuppose that our interests as theatergoers are somehow at bottom expressions of our theoretical interests, but why should we think this is so? Suppose you go to the theater and you find yourself disappointed by the bad acting. Is the problem that you are puzzled by what the actors are doing? It seems a distortion to say that the problem is that you have found their performances unintelligible. You can certainly understand what is going on. In fact, we can suppose that you have tried your hand at community theater yourself, with rather embarrassing results, so you have all too intimate "knowledge" of what is going on. A convincing folk psychological explanation of what you are observing is readily available to you. Instead of feeling puzzled, you feel bored and disappointed. You had gone to the theater with the hope and expectation that the actors would make it possible for you to escape into the world of the play, to suspend your disbelief. This is arguably a distinctively aesthetic demand, not an instance of a general cognitive demand. And what we call bad acting is usually bad precisely because it consists in failure to meet this distinctively aesthetic demand. The same is true of a certain kind of bad fiction writing. If the novel aspires to realism, then we want the characters to be believable. But even if being true to real life is not the standard for all fictional characters, we do generally want them to be such as to allow us to escape into the world of the story. Obviously there is a lot more to say about aesthetic norms. The point is that it seems a distortion to claim that the audience holds the actor to a standard of "intelligibility" as such, where that is construed as a way of keeping at bay the kind of puzzlement we feel when we lack a folk psychological explanation of what is happening.

The fact that theater audiences apply distinctively aesthetic standards explains why Velleman's analogy can seem to make human agency look unattractively self-absorbed. The claim that we are somehow acting for ourselves seems to imply that what matters to us about what we do is whether seeing ourselves doing it gives us a satisfying aesthetic experience. Clearly that is not Velleman's intention. But his purpose is to use the analogy to show us that we are already familiar with the standard of intelligibility that he takes to be the standard of human agency. To the extent that I do not see the continuity between my interests as a self-conscious agent and my interests as a theatergoer, I have trouble seeing how it is that I place a demand on myself to act "in character." If I do not have a "drive" to suspend disbelief with regard to myself, then what could the work of agency then be, such that its standard is like that of acting "in character"?

As I mentioned, Velleman rejects the idea that to act in character is simply live up to an ideal conception of ourselves. Doing that takes work because it is our nature to be subject to motivational impulses that threaten to lead us to betray the principles with which we identify

[2]The term "theoretical reasoning" appears in the main passage I am paraphrasing (*How We Get Along*, p. 17). The term "intellectual drive" is used in the parallel argument in "The Centered Self," in his *Self to Self*, p. 259.

ourselves. On this view, one characteristic way of failing as an agent is *akrasia*. The akratic fails to keep herself on track with respect to her chosen ends. Some would say that she allows her lower motivational capacity to determine what she does, over the objection of the higher, thus failing to uphold the natural constitution of her soul.

Velleman describes *akrasia* as involving a different kind of failure. He writes, "If [an agent] misunderstands himself, he may try to enact a disposition he doesn't have – in which case he may fail to carry it off, thereby suffering *akrasia*, or weakness of will." (15, nt. 7) This makes it sound like the failure involved in *akrasia* is simply a failure to attribute to yourself the motives you actually have. But Velleman denies that this is necessarily a failure. He affirms that certain kinds of "wishful thinking" can be entirely compatible with the requirements of agency. (91-92) Sometimes we should cultivate a false belief that we are brave for the sake of becoming brave. So attributing to ourselves the motives we actually have is neither necessary nor sufficient for acting in character in the relevant sense. But neither is it necessary or sufficient to attribute to ourselves the motives we would like to have. Sometimes doing this just counts as inauthentic.[3]

Given the weakness in the theatrical analogy, and given that Velleman explicitly rejects these alternative construals of what it might mean to act "in character," I admit I do not have an intuitive handle on what the standard is supposed to be. More to the point, I do not have an intuitive handle on the nature of the work of agency, such that this is its standard.

Perhaps the following line of thought would help. Imagine two scenarios. In the first, I act akratically and in the second I act enkratically. Suppose I am determined to grade this stack of papers. In the first scenario, I sit down at my desk and after an hour I find that all I have done is browse the web, answer e-mails, and shop online. In the second scenario, I sit down and my desk and after an hour I find I have finished my grading competently and efficiently. It is natural to say that in the first case I fail to act with integrity while in the second I succeed. At least this is the case if I chose my end in an authentic way, however that might be construed. Can we also say that in the akratic case, I have failed to make myself intelligible to myself, whereas in the enkratic case I have succeeded? It does seem natural to think that my response to the first scenario might be to say, "I don't know what I've been doing!" And it would be odd to think that I would respond the same way to the second scenario. I might be surprised to have worked so efficiently, having predicted I would be very distractible. But I would not express this surprise by saying, "I don't know what I've been doing!" If *akrasia* necessarily involves some kind of reflective opacity, if there is essentially a tension between being akratic and being clear-eyed, then could clear-eyedness in this sense be the kind of intelligibility Velleman thinks is at stake in every action?

I do think that clear-eyedness in this sense might well be at stake in every action, but I do not see evidence that this is what Velleman has in mind. Consider the view (which I will not argue for here) that *akrasia* necessarily involves self-deception at some level. If this is true, then there is a sense in which *akrasia* necessarily involves a failure to be intelligible to ourselves. The point is not that we fail to be predictable to ourselves; we might well predict our own weakness of will. It is rather that we fail to be transparent to ourselves. The way I would imagine this is that the motivational parts of ourselves that we have to unify in order to act enkratically are not communicating properly with one another. The sense of intelligibility is

[3]Velleman does provide a principle for distinguishing between authentic and inauthentic wishful thinking. If a false belief about ourselves will in fact be self-fulfilling, if it will in fact lead us to act in ways that engender in us the motive we lack, then the wishful thinking is compatible with authenticity because it ultimately results in the right kind of match between what we are and what we think we are. (91-92) But why describe this wishful thinker as acting authentically instead of describing her as sacrificing a degree of authenticity in the present for the sake of (perhaps greater) authenticity in the future?

that of transparency or publicity among the parts of the soul. I am not going to try to fill out the moral psychology that would support this suggestion. I will just testify that in my experience, the moments when I sense in a deep way that I do not know what I am doing are moments where, so to speak, "on hand doesn't know what the other is doing." I believe this sort of alienation is at stake in the examples of Freudian puzzlement that Velleman sometimes invokes. But nothing in Velleman's official moral psychology, as far as I know, commits him to there being motivational parts of the soul in either Plato's or Freud's sense. At least, he seems to think that the notion of intelligibility he is concerned with can be grasped independent of this assumption. So a general lack of textual support for this suggestion makes me think it will not be helpful in understanding how Velleman conceives of the relevant notion of intelligibility. Moreover, if this were the relevant notion of intelligibility, it would be something we value because we have to unify ourselves in the traditional sense. We would value intelligibility as transparency among our motivational parts because we value our integrity, conceived along something like Platonic lines as the unification of these parts. Since Velleman rejects that ideal, it seems even less plausible to attribute to him the corresponding notion of intelligibility. I am still wondering, then, exactly what the work of human agency is, and why acting "in character" counts as a determinate way of doing it.