Special Issue VII
2014

ISSN 1807-9792

abstracta

Linguagem, Mente & Ação

http://abstracta.oa.hhu.de

d|u|p

abstracta
Linguagem, Mente & Ação

ISSN 1807-9792

Special Issue VII
2014

Contents

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Special Issue VII, pp. 3–15, 2014

Inescapability and the Analysis of Agency[1]

Phil Clark

University of Toronto
philip.clark@utoronto.ca

In *How We Get Along*,[2] Velleman pursues a Kantian strategy. The point of this strategy, he says, is "to show that we can account for the objectivity of morality without positing a normative reality of which judgments of right and wrong can be true (120)." To this end, Velleman tries to explain how the demands of practical reason can have a kind of objectivity even if there is no normative reality of which practical reason judgments can be true. The key is to recognize that while all reasons for action rest on motives that are present in those for whom they are reasons, some rest only on a motive that is necessarily shared by all who are subject to reasons. In this paper I argue that Velleman's theory of reasons, even if it is true, fails to support the hypothesis that there is some one motive that is necessarily shared by all those who are subject to reasons for action. To make good on his Kantian strategy, Velleman must establish this hypothesis in some other way.

I proceed by granting Velleman his theory of reasons. In that theory, action constitutively aims at behaving in a way that makes sense, and reasons for action are indicators of what would make for successful action. In granting this for the sake of argument I depart from other critical strategies, for instance that of Kieran Setiya, who thinks Velleman has misidentified the constitutive aim of action, and therefore rejects Velleman's theory of reasons.[3] Setiya does not question whether Velleman's theory of reasons, if true, would support his Kantian strategy, whereas questioning that is precisely my business here. My concern is also not that of David Enoch, who worries that even if Velleman is right about the constitutive aim of action, he can't explain why we should do actions rather than, say, shmactions, which are like actions except for lacking the aim that makes something an action.[4] Like Setiya, Enoch is questioning Velleman's account of reasons for action, whereas again I am just questioning whether that account, if true, would support the Kantian strategy.

To do this I first need to explain Velleman's Kantian strategy. After that I describe his theory of reasons, and finally I address the relation between the two.

1 The Kantian Strategy

A strategy is for doing something. What does Velleman want the Kantian strategy to do? The answer, I think, is that he wants it to resolve an apparent tension within his view. He wants to "explain how morality can be objective." And he does this by first explaining how

[1]For helpful comments I thank Donald Ainslie, Lauren Bialystok, David Dyzenhaus, Joe Heath, Tom Hurka, Gurpreet Rattan, Devlin Russell, and James Sherman.

[2]All citations in the text are to Velleman (2009). For the Kantian strategy, see also Velleman (2000: Chapter 8).

[3]See Setiya (2003: 376).

[4]See Enoch (2006) and (2011).

practical reason can be objective, and then explaining how practical reason supports morality. The problem is that he also accepts claims that can seem to rule out any sort of objectivism about practical reason. He takes these claims from Bernard Williams:

> I agree with Williams's premise that reasons for acting must be able to engage a motive that the agent has or could come to have through sound deliberation; and I do not wish to question the assumption that deliberation can convey him only from motive to motive, so that his current motives determine where he could rationally end up. But I reject Williams's conclusion, that reasons must therefore be geared to something subjective in the agent's psychological make-up (119-20).

Velleman grants that nothing can be a reason for action unless it either engages a current motive of the agent or engages a motive that could be derived from a current motive of the agent. He thus accepts that reasons must be "geared to" something in the agent's psychological make-up; they must be geared to some current motive of the agent. And this seems to spell doom for any thought that the demands of practical reason, and of any morality spun from them, might be objective. For surely an objective demand would be one that was there in the world quite apart from what the agent happens to want. So here is the problem. If practical reason is tethered to current motives of the agent, how can it be objective?

As Velleman notes, Williams's answer is that it can't be objective. For Williams the demands of practical reason are subjective, because they rest on current motives of the agent. But Velleman thinks this is too quick. Indeed he thinks Williams has shown it is too quick, but has not made enough of his own insight. For elsewhere Williams distinguishes two kinds of objectivity. A demand can be objective in the sense of being "woven into the fabric of the world," but it can also be objective in the sense of being "woven into the practical point of view (116)." The former is the sort of objectivity Mackie famously argued is not to be had in ethics, but the latter is the sort Kant famously argued is to be had in ethics. Velleman quotes Williams making this distinction:

> Consider another picture of what it would be for a demand to be 'objectively valid'. It is Kant's own picture. According to this, a demand will be inescapable in the required sense if it is one that a rational agent must accept if he is to be a rational agent. It is, to use one of Kant's favorite metaphors, *self-addressed* by any rational agent. Kant was wrong, in my view, in supposing that the fundamental demands of morality were objective in this sense, but that is not the immediate point, which is that the conception deploys an intelligible and adequate sense of objectivity. It seems to have little to do with those demands being part of the fabric of the world; or at any rate, they will be no more or less so than the demands of logic – which was, of course, part of Kant's point (115).[5]

Velleman's problem was to explain how the demands of practical reason could be objective given that they are tethered to current motives of the agent. His solution is that they can be objective in the sense of being woven into the practical point of view, even if it they are not objective in the sense of being woven into the fabric of the world. Practical reason and morality can exhibit Kantian objectivity even if they do not exhibit the sort of objectivity Mackie considered and rejected.

[5]Bernard Williams, "Ethics and the Fabric of the World," in Williams (1995: 172-81)

The Kantian strategy, then, is to explain how the demands of practical reason are woven into the practical point of view. Velleman's explanation is that there is a motive, namely the motive of self-understanding, that must be present in anyone who is susceptible to reasons for action. Unlike other motives, this is not a motive agents just happen to have. It is not contingent, he says, but inescapable. If there are demands of practical reason that are tethered to this motive and this motive only, then these demands will be objective, not the sense that they are there in the world apart from any motives of the agent, but in the sense that they are reasons for anyone who can have reasons at all.

> For suppose that there is a single motive that any reason must engage in order for an agent to act on the basis of it. A creature will need this motive in order to satisfy the prerequisite for being subject to reasons for acting – in order for there to be reasons *for* him, or *applicable to* him – but he will then satisfy that prerequisite with respect to any and all possible reasons. Variance in motivation will no longer entail that similarly situated agents can have different reasons, . . . it will entail only that some creatures but not others can have reasons at all, because only some can be motivated in the relevant way (120).

This is where Velleman parts company with Williams. He agrees that reasons for action must be geared to current motives of the agent, but denies that they must be geared to motives that can vary from agent to agent. Instead they may rest on the motive of self-understanding, which is inescapable. So the Kantian strategy is to make room for objectivity in practical reason by showing that there is a single motive that is necessarily shared by all who are susceptible to reasons for action. Of the motive of self-understanding Velleman writes:

> Without this aim to make us susceptible to reasons, we would be incapable of acting for reasons, would not be agents, and would therefore be exempt from the force of reasons altogether. … Hence reasons for acting depend for their influence on a motive, but it is not a motive dependent on our several subjective constitutions; it's a motive that provides our shared constitution as agents. Reasons are therefore objective, and their status as reasons can be established once and for all, by the philosophical analysis of agency. That's the Kantian strategy (146-7).

It is important to distinguish two things Velleman might be saying here. One is that the fact of a necessarily shared motive makes room for objective reasons. The other is that the fact of a necessarily shared motive guarantees that there are objective reasons.

The latter claim faces a serious objection. It could be that although agents necessarily share the motive of self-understanding, what makes for self-understanding always depends on desires the agent just happens to have. In that case there would be no reasons that held for all agents regardless of their subjective constitutions. To show that there are objective reasons, therefore, it is not enough to show that there is a motive that is necessarily shared by all agents. One must go on to show that there are reasons that rest only on that motive.

While Velleman does not explicitly distinguish the two claims, I am going to assume he intends only the former. What he takes to follow from the fact of a necessarily shared motive is that there may be reasons that are objective despite being tethered to current motives, not that there actually are such reasons.[6]

[6]A question thus arises. Does Velleman try to show that there are reasons that rest solely on the motive of self-understanding? As I read the book, the early chapters amount to a sustained argument that groups of humans each

Why should we think, though, that there is some one motive that any reason must engage in order for an agent to act on the basis of it? Perhaps, if I am to act for a reason, that reason must engage some motive or other, but why must it engage the same motive every time? When I close the door because there's a cold draft, that reason engages the motive of staying warm, and when I put on my glasses because I can't read the fine print, that reason engages the motive of reading the fine print, but these are different motives. Why think there must be some one motive that is engaged by both reasons? To answer this question, Velleman turns to his theory of practical reasoning, on which he says his version of the Kantian strategy relies (117).

2 Reasons for Action

Velleman builds his account of reasons for action on an account of action, much as he builds his account of reasons for belief on an account of belief. Where belief is concerned, he starts with a distinction between representing something as true and believing it. I can represent it as true that am slaying a dragon, perhaps in a daydream, without believing I am slaying a dragon. What then is the difference between believing something and merely representing it as true? Velleman's answer is that believing Q is representing Q as true with a certain aim, in particular the aim of truth. So his account of belief goes roughly like this:

> **Analysis of Belief)** For S to believe that Q is for S to represent Q as true in an attempt[7] to represent as true whatever really is true as regards whether Q

He then treats reasons for belief as indicators of what would make for successful belief. Thus for instance, someone's wearing a gold band will be a reason for thinking he is married just in case his wearing a gold band is an indicator that representing it as true that he is married would make for success in an attempt to represent as true what really is true as regards whether he is married. More generally, the account of reasons for belief goes something like this:

> **Theory of Reasons for Belief)** For P to be a reason for S to believe that Q (or to believe that not-Q) is for P to be an indicator of what would make for success in an attempt to represent as true whatever really is true as regards whether Q[8]

Velleman models his account of reasons for action on this way of understanding reasons for belief. He begins with an account of action, and then characterizes reasons for action as indicators of what would make for successful action.[9] Moreover, his approach to action parallels his approach to belief. For belief he started with a distinction between representing as

---

of whom succeeds in molding his or her behavior and attitudes to what makes sense will "in the very long run" develop scenarios or protocols – ways of getting along – that have a distinctively moral cast. In this way, practical reason turns out to be "pro-moral." (The argument culminates at 149-51.) I suppose this might be read as an argument that there are reasons that rest only on the motive of self-understanding.

As far as I can see, however, the argument that practical reason is pro-moral could succeed even on the assumption that all reasons rest on contingent desires of agents. Perhaps over time humans with different subjective constitutions, and hence with different reasons, would under pressure of the necessarily shared drive for self-understanding come to have subjective constitutions, and hence reasons, that favor acting morally. But this is consistent with the hypothesis that all reasons rest on contingent desires. It is not clear, then, how Velleman can reach the conclusion that there are reasons that rest solely on the inescapable motive of self-understanding.

[7]The attempt, Velleman says elsewhere, need not be an attempt on S's part. It might just be an attempt on the part of S's cognitive faculties. See his 2000: 184-5.

[8]This is an account of theoretical reasons for belief only. If there can be practical reasons for belief, of the sort envisaged in Pascal's Wager, then one will need a separate account of these.

[9]One might worry that the notion of an indicator is too close to that of a theoretical reason for belief to be useful in an account of theoretical reasons for belief. But even if that is right, the notion of an indicator might be helpful in account of reasons for action. On this point see Setiya (2003: 339-40).

true and believing. For action he starts with a parallel distinction between mere behavior and action. Behavior can fail to be action. Crying, for instance, "can be a completely involuntary outpouring of emotion," in which one "just lets oneself go." This is mere behavior. Alternatively, one can "indulge in a good cry," in which case one's crying is guided by an idea of what one is doing, a "conception of crying." Depending on one's conception of crying, one may sniffle into tissues, or rend one's garments, or shout out lamentations. This is more than mere behavior, for Velleman. It is action, because in it one brings the manifestation of the emotion into accord with a conception of what one is doing (10-11, 26).[10]

Just as belief is more than representation-as-true, then, action is more than behavior. And just as he explains belief as representing-as-true with a certain aim, so Velleman explains action as behaving with a certain aim. To act, he says, is to behave with the aim of behaving in a way that makes sense. So the account of action goes roughly like this:

> **Analysis of Action)** For A's φ-ing to be an action is for A to φ in an attempt to behave in a way that makes sense as regards φ-ing

Thus the bereaved person's behavior – her crying or not, and her crying in one way rather than another – counts as action just in case she cries, or doesn't cry, or cries this way or that, in an attempt to behave in a way that makes sense. If she shakes uncontrollably, and not in an attempt to cry in a way that makes sense, then her shaking is mere behavior, even if it is a manifestation of her grief.

But is Velleman seriously suggesting that when I close the door to stay warm, or put on my glasses to read the fine print, my ultimate goal is something other than staying warm or reading, namely self-understanding? No he is not. As Velleman sees it, we pursue ends like staying warm for their own sake, not for the sake of understanding ourselves. What makes us agents, though, is that we seek to pursue those ends in ways that make sense. In this respect, he says, self-understanding is like efficiency:

> We cannot pursue efficiency alone; we can pursue it only in the course of pursuing other aims, by seeking to pursue them efficiently. And in seeking to pursue them efficiently, we don't pursue them for the sake of efficiency; we pursue them for themselves, albeit with the additional aim of doing so efficiently. So it is with self-understanding (27-8).

Having described action as "behavior aimed at intelligibility, just as belief is acceptance aimed at truth (133)," Velleman goes on to characterize reasons for action as indicators of what would make for success.[11] So the account of reasons for action parallels the account of reasons for belief, roughly as follows:

> **Theory of Reasons for Action)** For P to be a reason for A to φ (or not to φ, or to φ this way or that) is for P to be an indicator of what behavior on A's part would make for success in an attempt to behave in a way that makes sense as regards φ-ing

[10]Note that mere behavior, for Velleman, can be more than mere bodily movement. It can be motivated, in this case by grief, and it can even be a case of taking means to a desired end. On the latter point see his 2000: 189-191.

[11]"Reasons," he says, are "considerations that indicate an action to meet a substantive criterion of aptness or correctness (124)." But "having an aim already establishes a criterion of success or failure, which in turn yields a criterion of correctness for whatever can promote or hinder success (136)." Moreover, "action constitutively aims at making sense (146)," and this aim establishes the criterion of correctness for action. To be a reason for or against an action, then, is to be a consideration that indicates of some behavior that it will promote or hinder success in an attempt to behave in a way that makes sense.

Velleman cashes out the relevant notions of understanding and intelligibility in terms of a non-normative reading of the expression "what makes sense." He notes that we are apt to take this expression as meaning "what's appropriate or right or best." In Velleman's view, however, there is another notion of intelligibility from which appropriateness, rightness and goodness derive (27). The expression "what makes sense," read this way, means what is explicable in folk psychological terms. To make sense of someone's behavior in this sense is to trace the behavior to its causes in that person's motives, traits, and other dispositions. So Velleman does not see action as an attempt to behave appropriately, under that description. He sees it as an attempt to behave in a way that "can be understood as caused by [one's] motives, habits, and other characteristics (185)." He first explains reasons for action as indicators of what will make for success in such an attempt, and then explains appropriateness in terms of reasons.

3 Inescapability and Objectivity

As I said, Velleman takes his version of the Kantian strategy to rely on his view of reasons for action. And one can see why he would. The Kantian strategy, recall, was to explain how the demands of practical reason could be objective, by explaining how they are woven into the practical point of view. The explanation was that the demands of practical reason rest on a current motive of the agent – the motive of self-understanding – that is inescapable. The Kantian strategy thus relies on that motive's actually being inescapable. But this is precisely what Velleman's theory of practical reason provides. If Velleman is right that every action necessarily aims at self-understanding, then the motive of self-understanding is inescapable for agents. And this gives the Kantian strategy what it needs.

Or so it seems. In fact I think that while there is a way in which the motive of self-understanding is inescapable, on Velleman's view of practical reason, this does not give the Kantian strategy what it needs.

To see this, we first need to distinguish two ways in which a motive can be inescapable as opposed to contingent. Consider an analogy with another sort of an attempt, not an attempt to represent the truth or to make sense of one's behavior, but an attempt to locate an object.

Suppose we refer to any attempt to locate something as a search. And suppose we coin the term searcher to stand for anyone who ever searches for anything. Now one thing we might want to say about searchers is that anyone who searches for anything must have the motive of locating that thing:

> **Local Search)** For any object the motive of locating that object is inescapable for anyone who searches for that object

This is certainly a claim about inescapability. What it says is that the motive of locating a thing is inescapable for those who search for that thing. Thus the motive of locating my keys is inescapable for anyone who searches for my keys, and the motive of locating your dog is inescapable for anyone who searches for your dog. But it doesn't follow that there is any motive that is necessarily present in all who search for anything. Certainly there can be searchers who don't search for my keys, and searchers who don't search for your dog. And as a general matter there may be nothing that all searchers must try to locate. If so, then we have as yet no reason to think there is any motive that all searchers must share. It may be that each search is propelled by a motive that is contingent for searchers. Each of those motives will be inescapable for those who search for that thing, but contingent for searchers.

If this is right then we need to distinguish the claim above from the claim that there is a motive that is inescapable for searchers. It is one thing to say the motive of locating an object

is inescapable for those who search for that object, and quite another to say there is a motive that all searchers must share. We can put the latter claim as follows:

> **Global Search)** There is a motive such that, necessarily, anyone who searches for anything has that motive.

To pursue the analogy, we now need to consider a specific version of this claim, namely the suggestion that the motive of locating objects is inescapable for searchers. This is analogous to saying that the motive of truth is inescapable for believers, and that the motive of self-understanding is inescapable for agents. To say these things is to say that there is a motive that is necessarily shared by believers or agents, and to say something about what that motive is. But what would such a motive be like? Here too the analogy may help.

What would it mean to say there is a motive that is necessarily shared by all searchers, namely the motive of locating objects? In fact it is not altogether clear what this could mean. This necessarily shared motive of locating objects cannot be the motive of locating my keys, or your dog, because these are motives a searcher could lack. They are contingent for searchers, not inescapable. And even if it were impossible to search for anything without searching for, say, the elixir of life, the motive that was necessarily shared would be the motive of locating the elixir of life, not the motive of locating objects.

A more promising picture is available, however. For suppose that although there is no particular object that I want to locate, I still want there to be an object that I locate. This desire could move me to pick some item, say the world's largest frying pan, and search for it, as a way of bringing it about that I locate something. In that case the motive of locating the world's largest frying pan is derived from the motive of locating objects. One might also think of egg hunts here, or of the modern game of geocaching.[12]

Plainly this is a non-standard way of coming to search for something. The more usual way is to search for things on the basis of a need or desire to locate those things, rather than from any general desire that there be things that you locate. But this example does give us a clear picture of what the motive of locating objects could be.

Now suppose we understand the motive of locating objects in this way. What then becomes of the idea that the motive of locating objects is inescapable for searchers? There are two things to note here. The first is that this idea does not follow from Local Search. The second is that the idea is not very plausible. I'll take these points in turn.

The claim that the motive of locating objects is inescapable for searchers entails the claim that there is a motive that is inescapable for searchers, and adds a specification of what that motive is. So we can formulate it as follows:

> **Global Search 0.1)** There is a motive such that, necessarily, anyone who searches for anything has that motive, and it is the motive of locating objects.

And we can note that Global Search 0.1 entails Global Search.

We have already seen, though, that Global Search does not follow from Local Search. This was because Local Search could be true even if there was nothing every searcher had to search for. And this means Global Search 0.1 doesn't follow from Local Search either. For suppose it did. Then because Global Search 0.1 entails Global Search, Global search would follow from Local Search, and that gives a contradiction.

[12]Geocaching is the recreational activity of hunting for and finding a hidden object by means of GPS coordinates posted on a Web site. My authority is the *New Oxford American Dictionary, 2nd Edition*.

So Local Search does not entail Global Search 0.1. Or to put it in English, it doesn't follow, from the claim that the motive of locating an object is inescapable for those who search for that object, that the motive of locating objects, as currently construed, is inescapable for searchers.

But beyond this logical point, the two claims also differ in plausibility. Local Search is almost trivially true, whereas Global Search 0.1 is highly questionable.

As we've seen, the standard way of coming to search for something does not require the motive of locating objects, so understood. Normally, one just sees a need or conceives a desire to locate the object itself. On the face of it, then, there could be searcher who always came to search for things in this way, and never came to search for things in the non-standard way. And if that is so, then it is quite mysterious why there could not be a searcher who lacked the motive of locating objects. It may be common in humans to find some entertainment value in the sheer activity of locating things, but the claim in question is much stronger than that. The claim is that searching for anything entails the presence of that motive, so that there simply could not be a being that searched for things but lacked that sort of general attachment to the activity of locating things. That is what I am saying is implausible.

It is far from obvious, then, that the motive of locating objects is inescapable for searchers. But now what of the claim that the motive of locating an object is inescapable for those who search for that object? This seems a fair candidate for a conceptual truth. We simply stipulated that searching for something was attempting to locate it. Granted there is a lack of clarity around the notion of a motive, but on most any reasonable account a motive will be roughly an aim, goal or desire. And it is hard to see how one could attempt to locate an object without aiming to locate it, having the goal of locating it, or wanting to locate it.

So there are two morals we can draw from the analogy. One is that Global Search 0.1 does not follow from Local Search. The other is that Local Search is almost certainly true, whereas Global Search 0.1 is arguably false. These points matter because Velleman's theory of reasons rests on claims that are closely analogous to Local Search, whereas what he needs for his Kantian strategy is something analogous to Global Search 0.1. So if his theory of reasons delivers only the local kind of inescapability, then it does not deliver what the Kantian strategy needs. And that, recall, is what I aim to show in this paper.

To finish showing this I need to do two things. First I need to explain why Velleman's Kantian strategy requires the global kind of inescapability. And then I need to show that the theory of reasons delivers only the local kind.

Why then does the Kantian strategy need the global kind of inescapability? Why wouldn't local inescapability be enough?

Well, recall what the Kantian strategy is. The idea is to show that although every reason for action must rest on a current motive of the agent, there may be some reasons that rest only on a motive that is necessarily shared by all agents. For this to work, there must be a motive that is necessarily shared by all agents. In other words, the following claim must be true:

> **Global Action)** There is a motive such that, necessarily, anyone who does actions has that motive.

The reason local inescapability is not enough is that local inescapability does not entail global inescapability. We've already seen this for the case of searches. Now we need to see it for the case of action.

Just as one might search for one thing but not for another, so one might attempt to understand one's behavior in one respect but not in another. Perhaps I am currently making shrimp Alfredo, guided by a conception of making shrimp Alfredo, and I am also sighing, but not

guided by a conception of sighing. In Velleman's terms, my making shrimp Alfredo is an action, whereas my sighing is mere behavior. The difference, he says, is that I am making shrimp Alfredo in an attempt to behave in way I can understand, whereas although I am sighing, I am not sighing in an attempt to behave in a way I can understand.

I take it, though, that I am not making shrimp Alfredo in an attempt to make sense of all of my behavior. I am not, for instance, making shrimp Alfredo in an attempt to sigh in a way that makes sense. Indeed I may not even be aware that I am sighing. Rather, my aim in making shrimp Alfredo is to understand my behavior in respect of making shrimp Alfredo – my making it, or not making it, or making it this way or that, as the case may be. Or to put it another way, my aim is to exhibit shrimp Alfredo related behavior that I can understand. Thus if I stubbornly make shrimp Alfredo when it would make more sense make something else, or if I make it in the oven when it would make more sense to make it on the range, then my making it at all, or my making it in the oven, constitutes a failed attempt to exhibit shrimp Alfredo related behavior that makes sense. With this in mind we can formulate local inescapability for action as follows:

> **Local Action)** The motive of understanding one's φ-ing related behavior is inescapable for anyone whose φ-ing related behavior amounts to action.

And now we can see why, in the case of action, local inescapability does not entail global inescapability. It is because Local Action can be true even if there is no sphere of behavior that must amount to action for all agents. Suppose you have never heard of shrimp Alfredo, and are not currently engaged in any attempt to exhibit shrimp Alfredo related behavior that makes sense. That need not stop you being an agent, since you may be engaged in attempts to behave intelligibly in other respects. And for all Local Action says, it may be that every respect is optional in this way.

So Local Action does not entail that there is any motive that is necessarily shared by agents. For all it says, every motive may be contingent, in the sense that there could be an agent who lacked it. Global Action, on the other hand, says there is a motive that is not contingent in this sense. It says there is a motive that no agent could lack while still being an agent. Thus Local Action does not entail Global Action.

And for the same reason, Local Action does not entail the more specific version of Global Action that Velleman wants to emerge from the philosophical analysis of agency:

> **Global Action 0.1)** There is a motive such that, necessarily, anyone who does actions has that motive, and it is the motive of self-understanding

For suppose Global Action 0.1 did follow from Local Action. Then since Global Action 0.1 entails Global Action, Global Action would follow from Local Action. But we've just seen that Global Action does not follow from Local Action, so we have our contradiction.

Velleman's Kantian strategy consists in showing that there is there is a motive that no agent could lack while still be an agent, a motive whose presence is a prerequisite for being subject to reasons at all. And Velleman hopes to show this by showing that the motive of self-understanding is such a motive. His strategy, therefore, requires that both Global Action and Global Action 0.1 be true. But neither Global Action nor Global Action 0.1 follows from Local Action. So Local Action does not deliver what the Kantian strategy needs.

That was the first thing I needed to show. The second was that Velleman's theory of reasons, which I have here granted for the sake of argument, delivers only local inescapability. Given my reconstruction of Velleman's theory of reasons, this falls out fairly easily. I do grant, though,

that my reconstruction might be inaccurate, so that what I say here could be questioned on exegetical grounds.

Consider once more the two analyses on which the theory of reasons rests.

> **Analysis of Belief)** For S to believe that Q is for S to represent Q as true in an attempt to represent as true whatever really is true as regards whether Q
>
> **Analysis of Action)** For A's φ-ing to be an action is for A to φ in an attempt to understand her behavior as regards φ-ing

The former says the motive of truth about Q is inescapable for those who are in the business of forming beliefs about Q. It does not say there is any proposition that believers necessarily form beliefs about. The latter says the motive of understanding one's behavior as regards φ-ing is inescapable for those whose φ-ing is an action. It does not say there is any sphere of behavior with respect to which every agent's behavior must amount to action. There is nothing in Velleman's analyses of belief and action, then, that goes beyond local inescapability.

Nor do the theories of reasons that are built on these analyses go beyond local inescapability:

> **Theory of Reasons for Belief)** For P to be a reason for S to believe that Q (or to believe that not-Q) is for P to be an indicator of what would make for success in an attempt to represent as true whatever really is true as regards whether Q
>
> **Theory of Reasons for Action)** For P to be a reason for A to φ (or not to φ, or to φ this way or that) is for P to be an indicator of what behavior on A's part would make for success in an attempt to understand her behavior as regards φ-ing

The theory of reasons for belief analyzes reasons for belief on a topic in terms of what would make for success in an attempt at truth on that topic. It does not commit us to the idea that believers have any general motive of representing as true whatever is true.[13] It may be that humans sometimes do want there to be truths that they represent as true, much as humans sometimes want there to be objects that they locate. But nothing of the kind follows from Velleman's theory of reasons for belief.

Likewise, the theory of reasons for action analyzes reasons for φ-ing in terms of what would make for success in an attempt to understand one's φ-ing related behavior. It does not commit us to the idea that agents have any general motive of behaving in ways they will understand. Again, it may be common among humans to want there to be behaviors of theirs that make sense, and this could move them to pick a behavior that makes sense and do it, much as one might pick an object and set out to locate it. But nothing like this follows from the theory of reasons for action.

Velleman's theory of reasons, I conclude, delivers only local inescapability. What he needs for his Kantian strategy, however, is global inescapability. His theory of reasons, therefore, does not support his Kantian strategy.

That was what I set out to show. But it is worth emphasizing the limits of this conclusion. I have not questioned Velleman's account of action, or his theory of reasons for action. Nor have

[13] Velleman (2000: 252) makes a very similar point about belief. "What distinguishes believing a proposition from imagining it or supposing it," he writes, "is a more narrow and immediate aim – the aim of getting the truth-value of that particular proposition right … Belief is the attitude of accepting a proposition with the aim of thereby accepting *a* truth, but not necessarily with any designs on truths in general, or Truth in the abstract."

I argued that his Kantian strategy must fail. I have merely questioned his attempt to ground the Kantian strategy in the philosophical analysis of agency. The idea that there is some one motive that "provides our shared constitution as agents," as he puts it, simply cannot be derived from what he says about agency. For all I have said here, there may yet be such a motive. Perhaps its existence can be shown in some other way.

Now for a bit of diagnostic speculation. I've argued that Velleman can't spin the motive needed for the Kantian strategy from his philosophical analysis agency, but why would it seem that he could?

One possibility is that anyone who expects to engage in attempts of one sort or another is likely to want those attempts to be successful. Thus for example, if you figure that in the future there will be things you need to find, you will care whether your searches for those things succeed. And then if someone asks you, "Do you care whether you ever locate anything?" your answer must be yes, because if you never locate anything that means none of your searches will be successful, and who wants that? So perhaps there is a desire that all searchers will share, namely the desire not to fail in one's searches. And perhaps there is a desire that all believers share, namely the desire not to make failed attempts at the truth, and a desire that all agents share, namely the desire not to make failed attempts to behave intelligibly. So it appears that there is after all a way to derive a shared motive from the idea that beliefs and actions fall in the category of attempts.

But even if this is right, it does nothing for the Kantian strategy. The reason is that on the Kantian strategy, the motive that is necessarily shared by all agents must be one that can ground reasons for action. So the motive must be in a position to explain the actions for which it provides reasons. It must be possible, Velleman thinks, to come to do the action by a process of deliberation that starts from the motive. And that means the motive must be one from which particular attempts to behave intelligibly can be derived.

As we saw in our discussion of the motive of locating objects, there is a way of understanding that motive that allows it to play this role. If we take the motive of locating objects to be a general desire that there be objects that you locate, that desire can explain the motives of particular searches. For it can move you to pick some object and set out to locate it.

But the sort of desire we have just stumbled on is not like this. If you foresee that there will be things you need to find, and that is why you care whether you ever find anything, then the explanation runs from the anticipated searches to the desire, not the other way around. This is not a desire from which the motives of particular searches can be derived. You just have to let the searches arise as needed, and then execute them as best you can.

For the same reason, the analogous desires are not properly positioned to explain why one seeks the truth on some particular topic, or makes some particular attempt to behave intelligibly. If you foresee that there will be questions for which you seek true answers, and that is why you care whether you ever find the truth, that is not a desire from which the motives of particular beliefs can be derived. And likewise if you foresee that there will be things you do in an attempt to behave intelligibly, and that is why you care whether you ever behave intelligibly, that again is not a desire from which the motives of particular actions can be derived.

Perhaps, then, it is this phenomenon that explains the temptation to think the philosophical analysis of agency can support the Kantian strategy. We feel, reasonably enough, that all of us care about the success of our attempts. But then we mistakenly infer that the motives of particular attempts can be derived from this desire.

Another possibility is that we mistake a shared property for a motive. Local inescapability does entail that something is shared. For instance it does follow from Local Search that there is something all searchers share. It follows that they all have the property of being such that there is an object that they have the motive of locating. It does not follow, though, that there is any motive that they share. So perhaps we feel, rightly, that there is something all believers share, and something all agents share, but mistakenly infer that what they share is a motive.[14]

4 Conclusion

The Kantian strategy derives no support from the theory of reasons. But so what? How does this result affect Velleman's ambitions?

The answer may depend on which ambitions we have in mind. The one on which I have focused is the project of accounting for the objectivity of morality without positing a normative reality of which moral judgments can be true. The idea was to reconcile Williams's point that all reasons for action must be geared to present motives with an objectivist view of those reasons. The solution was to explain how the demands of practical reason could be "woven into the practical point of view." And the explanation was that if we look at what it is to act, we will find that there is a motive that must be present in anyone who qualifies as an agent. I've argued that even if we grant Velleman's analysis of action, and the accompanying analysis what it is to have a reason for action, we do not thereby find that there is any such motive.

But perhaps there are other ways of showing that the demands of practical reason are woven into the practical point of view. One possibility would be to stop looking for a single shared motive, and make do with Velleman's theory of reasons on its own. Could we then reconcile Williams's idea with the objectivity of practical reason?

It appears that we could not. For unless there is a necessarily shared motive, all motives are contingent for agents. And in that case we are forced to choose between the two options Velleman wanted to avoid. We must either bind reasons to contingent motives or sever the connection with current motives altogether.

Another possibility would be to stop looking for a motive that is constitutive of agency, and look instead for a motive that is reliably shared by human beings. Suppose Velleman were to accept that there could be agents who made only local attempts to behave intelligibly, and who lacked any more general aim from which the motives of those attempts could be derived. He could still claim that we humans are not such agents.

And in fact we know he would claim this, because he does. He distinguishes two senses in which the aim of self-understanding is inescapable. It is constitutively inescapable for us as agents, he says (137). But it is also naturally inescapable for us as human beings, in the sense that we humans can't help seeking to understand ourselves. Or rather we can only opt out of this aim temporarily, by losing ourselves in an activity or in daydreams, or by going to sleep (137).

What I am now suggesting is that Velleman could drop the idea that there is a motive that is constitutively inescapable for agents, and pin his hopes on the claim that there is a motive that is naturally inescapable for us as human beings. Perhaps he could argue that the demands of practical reason are woven, not into the practical point of view *per se*, but into the human point of view. That might be a way of combining Williams's internalism with a kind of objectivism about practical reason.

[14] I thank Gurpreet Rattan for this suggestion.

To make this out, he would need to show not only that humans can't help doing behaviors in attempts to behave intelligibly, much as they can't help forming beliefs, but that there is some one motive that humans can't help having. He would also need to show that this was a motive from which particular attempts to behave intelligibly could be derived. And finally he would need to show that there are reasons for action that rest on that motive and no other.

There is evidence that Velleman has such a program in mind. For he argues at some length that empirical work in social psychology supports the hypothesis that humans have a "drive toward self-understanding (17, 64).[15]"

Another possibility, though one to which Velleman is clearly less inclined, would be to rethink the commitment to Williams's conclusion. Does it really follow, from Williams's premises, that all reasons for action rest on current motives of the agent? One might rather reject this conclusion, instead of looking for a kind of objectivity that coheres with it.[16]

References

Clark, Philip (2001) "Velleman's Autonomism," *Ethics* 111: 580-93

Enoch, David (2011) "Shmagency Revisited," in M. Brady ed., *New Waves in Metaethics* (Basingstoke: Palgrave Macmillan) 208-33

Enoch, David (2006) "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Agency," *Philosophical Review* 115, 169-98

Korsgaard, Christine (1986) "Skepticism about Practical Reason," *The Journal of Philosophy* 83, 5-25

Setiya, Kieran. (2003) "Explaining Action," *The Philosophical Review* 112:3, 339-393

Velleman, J. David. (2009) *How We Get Along*, Cambridge: Cambridge University Press

Velleman, J. David. (2006) *Self to Self: Selected Essays*. New York: Cambridge University Press

Velleman, J. David. (2000) *The Possibility of Practical Reason*, Oxford: Oxford University Press

Williams, Bernard (1995) *Making Sense of Humanity and Other Philosophical Papers 1982-1993*, Cambridge: Cambridge University Press

Williams, Bernard (1981) "Internal and External Reasons," in Williams, *Moral Luck*, Cambridge: Cambridge University Press, 101-13

[15]See also Velleman, "From Self Psychology to Moral Philosophy," in his (2006).

[16]Velleman (2000: 172 ff) considers Korsgaard's (1986) critique of Williams (1981). He reads her as severing the connection between reasons and current motives of the agent, and objects that she then faces a "burden of justification." For a reply to this objection, see my (2001).

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.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Special Issue VII, pp. 23–29, 2014

Velleman on Reacting and Valuing

Justin D'Arms

Ohio State University
darms.1@osu.edu

How We Get Along articulates an original and richly systematic view of ethics and moral psychology in a thoroughly engaging way. It is the first book I have read in many years that is rigorous and original enough to be a new contribution to the discussions of professional philosophers while being at the same time sufficiently approachable and interesting that I have recommended it to a intelligent non-academic who wants see what is going on in philosophy. I recommend it wholeheartedly.

Here I focus on the novel account of value and reasons for valuing that Velleman articulates. I will discuss some cases of 'wrong kinds of reasons' and of evaluative conflict that look like they might be counterexamples to Velleman's view. But my goal is not to establish that they are counterexamples. I want to use them as tools to get clearer about the theory. I am primarily interested in trying to put some of Velleman's interesting ideas to work by applying them to my problem cases, and in giving him an occasion to improve on my initial efforts to do so.

"Something's being valuable" Velleman says (37) "...consists in there being reasons for valuing it...." I am very sympathetic to this suggestion, as are many contemporary philosophers. I see Velleman's view as falling within the resurgent philosophical tradition that identifies being valuable (in some respect) with being such as to merit or make appropriate some (relevant kind of) evaluative attitude.[1] This notion of a response being appropriate is commonly understood in terms of there being reasons (or its being rational) to have it. Thus value is explained in terms of reasons for valuing, not the other way around. Which is as it should be, because, as Velleman says, "'value' is the term most in need of analysis."

The question then is how to understand reasons for valuing, and here philosophers attracted to the broad idea diverge. Velleman has his own proposal: reasons are considerations of intelligibility, and reasons for valuing are considerations whose regulative influence turns a mere reaction (such as amusement) into a valuing (such as finding something funny). In this respect, Velleman's account of valuing echoes his account of acting for a reason (which is reframed in the first chapter of this book in terms that readers of his previous work may find illuminating, as I did).

> "Just as behaving becomes acting when it is regulated by the agent's conception of what it would make sense for him to do, so reacting becomes valuing

[1] Early versions of this idea can be found, for instance, in the fitting attitude theories of Franz Brentano (1889) and A.C. Ewing (1948). More recently it has re-emerged inT.M. Scanlon's (1998) buck-passing idea, in John McDowell's (1985) and David Wiggins' (1987) "sensibility theories", in Allan Gibbard's (1990) norm-expressivism, in the neo-sentimentalist views of Kevin Mulligan (1998), John Skorupski (2010) and D'Arms and Jacobson (2005), and in Elizabeth Anderson's (1993) rational attitude theory (to which Velleman notes affinities), among others.

> when it is regulated by the subject's conception of what it would make sense for him to feel." (40)

Velleman takes a view of emotional reactions reminiscent of Singer and Schacter's, on which feelings get enriched through interpretation from an initial, somewhat labile, sort of affective excitement that could become any one of a number of different emotions, depending on how the subject understands his situation and finds it intelligible to react to it. If the subject focuses on aspects of the circumstances to which he thinks it makes most sense to be disgusted, say, his initial reaction can be shaped into disgust by this conception of what makes sense. Such disgust would be what Velleman calls a 'guided response', which is not a judgment of disgustingness, but an affective or conative state that is sensitive to indications of its own appropriateness. Guided responses are reason-sensitive emotional responses, whereby one finds things to be valuable in ways that are distinct from mere feeling, but also distinct from evaluative judgment. These guided responses are the evaluations at the heart of Velleman's account of value. Something's being valuable in some respect, disgusting say, is constituted by whichever of its features are such that considering them could lead to a guided response of being disgusted by the thing in question.[2]

Merely reacting to something by liking it or being amused by it, is not yet finding it valuable—likeable, or funny—according to Velleman (35). Perhaps that is right. Certainly it is not yet judging it to be likeable or funny, as he notes. But in cases where one is amused at what one knows is not funny, why not think one is in some sense finding it funny without judging it so? You can be amused at weak comic material because you are giddy or stoned, and know it. But it's common to suppose that such reactions are recalcitrant—that is, that they are in some sort of tension with the judgment that it is not funny. Likewise for fear at what you think is not really dangerous, and so on. Our sense that these are *recalcitrant* responses at least suggests that they already involve finding things to be some way, evaluatively speaking, that one has concluded they aren't (funny, dangerous, etc.). If not, then it is hard to see why such combinations of affect and judgment seem to be rationally at odds with one another, and why we often seek to resolve them.

Perhaps Velleman would say something like this: by the time you reach full-fledged amusement at some determinate thing, as opposed to a labile hit of pleasure which might be interpreted as amusement, or a feeling of social connection, or whatever, then you have a guided response, not a mere reaction. That's because there are features of the thing that make your amusement intelligible to you. You might even be able to identify those features (Moe just keeps hitting Curly with the hammer in different places as Curly defends the previous target), even if you judge that this is silly, predictable slapstick that does not really merit amusement. In other words, your response is being guided by the features of what is happening that make amusement intelligible, even if you think it is ultimately unearned. At that point, I think he should say, you could be finding it funny, affectively, despite your contrary judgment.

But if so, then what you find funny, by way of your guided responses, is not necessarily funny even by your lights. It is not necessarily funny-for-you. Most obviously, it is not yet funny by the lights of your judgment. Slightly more interesting, it is not yet funny by the lights of your sensibility either. That is because locally, in a particular case, something you would not normally find funny can temporarily seem so. In some such cases, I'd say, you find it funny, though it isn't, even by your own lights—that is, by the lights of your sense of humor,

[2]See p. 40. Could lead in whom, you may be wondering? There are various answers to this question, leading to more or less objective conceptions of value. More on this shortly.

which is a more stable subset of your tendencies to find things funny. Your sense of humor does not include all such tendencies, because some amusement is due to obscuring factors that can lead you to find things funny (or unfunny) in predictable ways that nonetheless do not fit a sensible pattern of value.[3] You can fail to be amused by an excellent joke, delivered well at a party by someone you detest, to the great mirth of everyone else. But the joke is funny by your lights, because were anyone else to have told it you would have found it very funny. And you can be genuinely amused at a story told by your beautiful and interesting date, even though when you retell it to a friend the next day (as evidence that she was not merely beautiful and smart but funny too) you come to think that it was not so funny after all. (Suppose yourself, for purposes of this example, to be romantically unencumbered, and sexually attracted to women.) At first, in retrospect, you think it must have been the way she told it that made it so funny. On reflection, though, you realize that her delivery was not comically excellent either; it was flirtatious, and fun, and you went along with her amusement at the story, as one often can, if the underlying material is close enough to funny. You did not need to *feign* amusement, because your enjoyment of the evening and your underlying desire to make a connection with this lovely woman led you unselfconsciously to find funniness in what was really not a funny story, by your own lights. So, I conclude, genuinely finding something funny can and should be kept distinct from its being funny by the lights of your sensibility, and likewise, mutatis mutandis, for other guided responses and the values they help us find.

I have been helping myself to talk of something's being funny or disgusting "by someone's lights", which is one thing one might naturally mean by saying that it is funny for him. And ordinary thinking recognizes a distinction between what is funny, disgusting, admirable by someone's lights and what is funny in some objective way that purports to have a claim on everyone. The relations between these notions is vexed. On one hand, it is natural to think that to find something disgusting or funny is to find it (rightly or wrongly) to be a way that would merit disgust or amusement from anyone who got *en rapport* with it, as it were. On the other hand, there are cases of blameless difference where it is tempting to think that what is funny or disgusting for you need not be so for me. Velleman has some very interesting things to say about these more and less objective standards of evaluation.

We allow for individual differences, so that what is likable or admirable for you need not be so for me, he notes. Moreover, what it is intelligible for you to admire or be amused by is holistically interdependent on what other responses are intelligible for you to have. Because there are specific functional-explanatory connections between different responses and various actions and other responses, the intelligibility of a given response may be a pretty idiosyncratic matter—tied up with patterns of interest and projects which make certain things interesting or boring or insulting to you in ways that affect what you can sensibly find funny or offensive, and so on. Still, your actual likes and dislikes can fail to detect what is really likable from your perspective. So for all its quirks, your perspective must be something sufficiently stable to rule out certain responses as inappropriate by your own lights. (Whether Velleman thinks these inappropriate responses can be 'guided responses,' as opposed to mere reactions is not clear, but in light of the date case, I think he should allow that they can.) And we can criticize one another's sensibilities as if there were an objective criterion of good taste. But different values appear to differ in their susceptibility to such a criterion—we allow more leeway for tastes in liking than in admiration, for instance. How can the conditions of appropriateness be objective in some cases and relative to individual sensibilities in others, and differ in all the ways above?

[3]Here and going forward, I make use of some ideas developed at more length in D'Arms and Jacobson (2010).

> "The answer, I suggest, is that the fundamental standard of appropriateness for a response is its intelligibility, which is determined partly by the nature of the response itself and partly by differences among individual sensibilities, which can themselves be compared and criticized on grounds of intelligibility." (41-42)

What are these individual sensibilities, and how are they related to the reactions it is intelligible for a person to have? One might think that any idiosyncracies of yours that render certain guided responses intelligible for you and not for others are elements of your sensibility. But the date case makes that suggestion problematic. As I was imagining the case, it seemed most plausible to say that the date's story was not funny by the lights of your sense of humor—your comic sensibility. And yet it seems that your amusement is entirely intelligible to you, at least on one natural way of thinking about intelligibility. It is a guided response, which is sensitive to comically relevant aspects of the story, as evidenced by the fact that you laugh at the right parts, not just randomly as it goes along. Moreover, being amused by a story told on a date by someone you are attracted to is entirely intelligible in social and prudential terms. If you *can* be amused in such circumstances, you'd have to be a real high-church stickler about the funny to resist. So it seems, on balance, that amusement is the most intelligible response in such a case.

One might try saying: No, feigning amusement is the most intelligible reaction. That might be true if you could not muster authentic amusement—and that is a common predicament about responding to incentives to feel. But the point of this case is that often enough we can and do respond immediately in ways that are unselfconsciously sensitive to facts about how our responses will fit in our social lives—an important point that Robert Solomon (1976) used to emphasize. This is a fact we can notice in the abstract and embrace, since (for instance) feigning amusement is less likely to lubricate our social interactions than is actually being amused. Thoughtful people can recognize and embrace respects in which their social emotions are shaped by various aspects of their relations to others that go beyond evaluatively relevant features of the objects of those emotions.

If I am right to say that amusement is on the whole the most intelligible reaction in this case, and also that the story is not funny by your lights, this shows that your comic sensibility comes apart from how it is intelligible for you to feel. That calls into question whether intelligibility is indeed, as Velleman says, the fundamental standard of appropriateness for guided responses—even when one limits oneself to personal standards of appropriateness that are supposed to fix the less objective notion of funny-, likeable- offensive-to-me.

I see in the text hints at several lines of response that Velleman might adopt. One is to insist that amusement in the case at hand is not on the whole the most intelligible response, even though it is natural, adaptive, and the response one would choose to have if such things were voluntary. I don't like that reply, because it requires me to abandon strong intuitions about intelligibility as I think my way into his theory (I think amusement at the date's story is clearly intelligible). I then worry that verdicts about what responses are merited or appropriate are driving the notion of intelligibility, when the avowed goal was to work in the other direction. But perhaps more could be said about intelligibility to relieve one or both of these worries.

Another line of reply is suggested by a very interesting discussion of conflict in which Velleman urges the availability of conflicting valuings (48-58). He argues that it makes sense that someone who had a child at too young an age can both think that she should not have done so, all things considered, and truthfully say that she does not wish she had not had her

child. He sees these stances as rationally conflicting, but also as potentially the most intelligible combination of reactions to have. He says that such conflicts would not make sense if our evaluative responses could reflect "antecedent value-properties supervening on nature" but, since there are no such properties, inconsistent valuings, though inimical in general to self-understanding, are not to be condemned as senseless or stupid (48). In this spirit, one might say something like this: my date's story was funny (-for-me), in that light, on that night. But that story is not a funny story. It is neither objectively funny (I do not criticize others for failure to appreciate it) nor is it funny-for-me.

Whatever one makes of the baby case, embracing the inconsistency seems to me sub-optimal in the case at hand.[4] A central point of thinking about our responses as organized around values, surely, is that it allows us to understand ourselves as responsive to relatively stable features of the world that we think supply good grounds for (or confer intelligibility upon) responding as we do. We draw the 'is Φ/seems Φ' distinction for value terms (funny, disgusting, offensive, shameful...) precisely in order to distinguish reactions that we think responsive to generally intelligibility-conferring features from responses due to other vagaries of our natures and circumstances which, though perhaps predictable and sometimes intelligible, are not the sorts of causes of response that help to collect together a coherent value. If the notion of funniness is to do any work of this sort, it seems to me, it should be wielded here precisely to deny that the very good reasons for being amused by the date's story are comic reasons.

A third response is to grant my claims, and say that the standards of appropriateness for responses that constitute something as valuable in a respect, for a person, are just one strand (albeit a very important one) of intelligibility-conferring considerations with respect to questions of how to react. This is, in effect, to distinguish the question of whether it is intelligible to be amused by something from the question of whether that thing is funny for one. The idea would be to seek a general pattern in the kinds of features that confer intelligibility on responses specifically by making something funny, or disgusting, or offensive, and identify these strands as the ones conferring the sort of intelligibility that renders the response appropriate (in the sense of 'appropriate' corresponding to values). This looks like the best option to me. But it incurs a debt that the first two responses avoid: to explain which kinds of general patterns of intelligibility of response figure in standards of appropriateness (i.e. of value-by-your-lights) and which ones don't. Our tendencies to succumb to emotional contagion and be amused by marginal comic material in contexts where it is socially adaptive to do so may be sufficiently robust to constitute a kind of pattern, after all. Likewise our tendency not to be amused by jokes told by people we dislike. But while these patterns may enhance, rather than diminishing, our overall intelligibility to ourselves, it makes better sense to think of them as involving reasons for response that do not figure in our standards of funniness, but are justified on other grounds.

I think that Velleman can embrace this answer. This is to accept that some considerations which bear on the intelligibility of amusement, disgust, and offense do not bear on the appropriateness of those responses—that is, on whether something is funny, disgusting, or offensive. But if that is indeed granted, I think we have less reason to accept some claims Velleman makes about the relationship between distinct values. And this will be my final point.

[4] I am undecided what to think about the baby case, but I see the attractions of Velleman's analysis. It is worth noting that the judgments that Velleman wants to embrace while treating as inconsistent there are not expressed in terms of some single evaluative predicate that registers the appropriateness of a relatively discrete response—like funny, offensive, etc.. If they were, the conflict might be starker and make less sense.

According to Velleman (42-3),

> . . . [T]he criterion of appropriateness for a response is holistically interdependent with those for other responses, as are the corresponding values. . . . We say, "That's not funny," though sometimes we are laughing as we say it; and we may then add, "So why am I laughing?" This rhetorical question confirms that the unfunny is that which we don't understand laughing at. The reason we don't understand laughing at something is not that it is unfunny; rather, we don't understand laughing at it because it is boring or offensive or disgusting. . . and the resulting incongruousness of laughing at it is the reason why we think it isn't funny, despite our laughter. . . . Thus what it makes sense to be amused by depends in part on what it makes sense to be disgusted, bored, or offended by."

As I read this, the idea is that when the offensiveness or disgustingness of the material sufficiently reduces the overall intelligibility of laughing at it, that makes the material unfunny (for you)—even if you are laughing at it. Note, however, that another option would be to allow that the material is genuinely funny by your lights, it's just that funniness is not the only relevant value. Sometimes it would be more intelligible to withhold amusement *even though it is appropriate*, because other appropriate responses would better reflect one's considered views (or one's overall affective perspective) about the respective importance of the funniness and the offensiveness of the material. Having acknowledged that some considerations that render a response more or less intelligible do not thereby render it more or less appropriate, we can now resist a certain sort of evaluative holism. We can allow that, in some cases, even though the joke's offensiveness makes amusement not intelligibly accessible, it does not render amusement inappropriate—it does not make the joke unfunny. (In other cases the joke is just offensive, not funny, but that, I am suggesting, is not the only possibility.)[5]

This is not simply a point about amusement and the funny. The issue is general, because it concerns the claim of holistic interdependence between values in the first quoted sentence above. Certain combinations of responses are hard to have at the same time: amusement and offense, pity and indignation, fear and unflinching determination, admiration at one aspect of a person and contempt at another. This means that responding to the considerations that make one of them appropriate may be simply incompatible with responding to those that would make another appropriate. One question to ask in such cases is which response is most intelligible. If that question has a clear answer, then having any response that is incompatible with the intelligible one is not intelligible—or at least, not *as* intelligible. What I want to insist upon, though, is that it is always a further question whether the response that is thereby rendered less intelligible is also rendered inappropriate, or even less appropriate. Sometimes, I think, the answer will be no. In those cases we have conflicts of value which are such that we can't respond to all of them. In light of his willingness to countenance drastic kinds of conflicting value judgments in the baby case, perhaps Velleman will embrace this too. But if so, I think the claims of holistic interdependence among value are at least misleading; and more attention should be paid as well to the independence of distinct values. In short, once the appropriateness of responses is distinguished from their intelligibility, we cannot infer the holistic interdependence of values from the holistic interdependence of intelligibility conditions of response.[6]

[5]There is a footnote (FN 9 p. 43) in which Velleman allows for the possibility of "sick or offensive humor." But I don't understand it well enough to explore how it might help with the issues I am pursuing here.

[6]This paper was supported by a grant from the John Templeton Foundation.

References

Anderson, Elizabeth (1993). *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Brentano, Franz (1969) [1889]. *The Origin of our Knowledge of Right and Wrong*, ed. Oskar Kraus and Roderick Chisholm, trans. Roderick Chisholm and Elizabeth Schneewind. London: Routledge & Kegan Paul.

D'Arms, Justin and Daniel Jacobson (2005) "Anthropocentric Constraints on Human Value." *Oxford Studies in Metaethics*, 1: 99-126.

D'Arms, Justin and Daniel Jacobson (2010). "Demystifying Sensibilities: Sentimental Values and the Instability of Affect" *Oxford Handbook of Philosophy of Emotion*, ed. Peter Goldie. New York: Oxford University Press.

Ewing, A. C. (1948). *The Definition of Good*. London: Routledge & Kegan Paul.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.

McDowell, John (1985). "Values and Secondary Qualities." in *Morality and Objectivity*, ed. Ted Honderich. London: Routledge and Kegan Paul.

Mulligan, Kevin (1998). "From Appropriate Emotions to Values," *The Monist* 91,1: 161–188.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Skorupski, John (2010). *The Domain of Reasons*. Oxford: Oxford University Press.

Solomon, Robert (1976). The Passions. (New York: Doubleday)

Wiggins, David (1987). "A Sensible Subjectivism?" In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell, 1987.

abstracta Linguagem, Mente & Ação ISSN 1807-9792 Special Issue VII, pp. 31–38, 2014

Symposium on *How We Get Along*
Responses to Critics

J. David Velleman

New York University
jdvelleman@nyu.edu

How We Get Along begins with a sob: you are crying uncontrollably. But then you get a hold of yourself and settle down to have a good cry. I ask: what has changed? Whatever it is, it's what makes the difference between behavior that is out of your control and therefore not an action of yours, on the one hand, and behavior that is in your control and therefore an action, on the other.

I contend that whereas the uncontrolled crying is simply the manifestation of hurt or grief, the controlled crying manifests something further, namely, your awareness of the hurt or grief, and your resulting perception of crying as what it makes sense to do. Had you been unable to think of what you might be crying about, you would have asked yourself "Why am I crying?", and your tears would have tapered off. Because you do know what you're crying about, however, you keep crying, but now with the concurrence of that self-understanding, which transforms your crying from mere behavior into an action.

What differentiates action from other behavior, then, is that it is guided by a disposition to do what makes sense to you — a disposition, I suggest, that arises from a drive on your part toward self-understanding. The drive toward self-understanding is not the primary motive of your actions; it is rather a second-order motive with the respect to the manner in which you act on other motives. Grief is your primary motive; the drive toward self-understanding can only oppose or reinforce your grief.

When I say that the drive toward self-understanding is what differentiates action from mere behavior, I mean that it is responsible for the features that are distinctive of action: the way we know what we're doing when we act; the way our actions are up to us, in the sense that we choose between alternatives; the way we commit ourselves to future actions in advance; and the way that our actions are guided by reasons. These features were absent from your passive crying. You may not have realized at first that you were crying; it wasn't up to you whether to cry; and although something had caused you to cry, you weren't crying for that reason. Then you saw what there was to cry about and you had a cry for that reason. It was up to you whether to have a cry, and when you went ahead, you knew what you were doing.

I initially compare this case to the crying of a "method" actor, who actually feels the emotion fictionally felt by his character and then channels it into behavior that is "in character" because it makes sense as the way his character would cry. I then complicate the comparison by pointing out that you are not following a script: you are therefore more like a improvisational method actor, feeling the relevant emotions but making up your part of the whole episode, still within the constraint of acting in character, lest the improvisation fail to make sense.

Finally, I introduce one more complication. You are not acting out a fiction, working up reactions to fictional circumstances with a method actor's tricks and then acting them out in a character that's fictional. You are improvising your actual self, reacting to your actual circumstances, in part spontaneously but also guided by how it makes sense to react; and doing what actually makes sense for you to do given those reactions. Your "good cry", I claim, is an authentic improvisation, fueled by motives that are both realistic and real, but also guided by your perception of how it makes sense to feel and to manifest your feelings.

In the remainder of the book, I develop this theatrical analogy into a theory of individual and collective practical reasoning — individual in that each actor must do what makes sense for his character to do, collective in that what makes sense for each depends on what the others are doing. Like a troupe of improvisational actors, I suggest, members of a community or society must develop shared scenarios on which their joint improvisations are based, as variations on a theme. These scenarios are their shared way of life. And I suggest that their way of life will take on the shape of a morality because of the constraints imposed by exigencies of collective improvisation.

Response to Phil Clark

My theory of action has a metaethical implication. Because self-understanding is a second-order aim of every action, it provides a normative standard for action as such, irrespective of any particular action's first-order aims. Self-understanding is a normative standard for action as such in the same way as truth is a normative standard for belief as such, irrespective of what any particular belief is about. If someone's movements aren't aimed at intelligibility, in addition to his first-order aims, then he isn't acting. If he is acting, then his movements are aimed at intelligibility, and so the standard of intelligibility applies. Thus, the standard applies to any and every action.

As Phil Clark says, this theory of action and its companion metaethics are meant to implement what I call the Kantian strategy. Kant tries to derive an objective practical norm from a standard that is inescapable from the practical perspective, in that anyone must be committed to it in deciding to act. I do likewise, though I call it an aim and I identify it as intelligibility rather than universalizability. These differences are minor, but Clark thinks that there is a more significant difference. For he thinks that I fail to derive an objective practical norm from my theory.

The problem, according to Clark, lies in the specification of the aim that is inescapable from the practical perspective. It is not the aim to do intelligible things; similarly, the aim of belief is not to believe true things. A rational believer doesn't go around accumulating as many true beliefs as he can. The constitutive aim of belief is to believe things only if they are true, but their being true is not sufficient; they also have to bear on a topic on which he has reason to hold an opinion. Similarly, a rational agent doesn't go around doing as many intelligible things as he can. The constitutive aim of action is to do things only if they make sense, but their making sense isn't sufficient; they also have to be things that he has some motive for doing.

Clark's objection is that aiming to do things only if they make sense isn't inescapable from the practical perspective. An agent could have a specific aim, with respect to each action, to do *it* only if it makes sense. Clark compares these aims to the aims of finding particular things when one is looking for them. Not only would it be ridiculous to aim at finding lots of things, as it would be to aim at doing lots of intelligible things or believing lots of truths; it would even be odd to have the general aim of finding things if one is looking for them. All one needs,

when searching for something, is the aim of finding that particular thing. And that specific aim is sufficient to constitute one's activities as a search. So long as one is trying to find something in particular, one is engaged in a search, whether or not one has the general aim of finding things that one looks for.

The upshot, Clark contends, is that the aim of being intelligible in doing a particular thing should be sufficient, according to my theory, to constitute that thing as an action. Just as a belief that *p* requires no more than the aim of believing *p* only if *p* is true; so performing action *a* should require no more than the aim of doing *a* only if *a* makes sense. Hence action requires no universal aim that is present in every case; it requires only particular aims, specific to the actions undertaken by particular agents on particular occasions. All that can be derived from those aims are particular standards, different standards for different agents and different actions. Objectivity has therefore not been attained.

Clark illustrates his argument with an example. Imagining himself cooking shrimp Alfredo, he says, "[M]y aim in making shrimp Alfredo is to understand my behavior in respect of making shrimp Alfredo – my making it, or not making it, or making it this way or that, as the case may be. Or to put it another way, my aim is to exhibit shrimp Alfredo related behavior that I can understand." Here Clark assumes that the aim of understanding his shrimp-Alfredo–related behavior is not an instance of the more general aim of understanding whatever he does. He assumes, in other words, that the relevant aim is not the result of substituting "making shrimp Alfredo" for *x* in an aim with the content "For all actions *x,* to do *x* only if I understand *x*-ing", since the latter would be a general aim of the sort that he denies my theory requires.

Clark believes that his description of the case follows from my analysis of action, which he paraphrases like this:

> **(Analysis of Action)** For A's φ-ing to be an action is for A to φ in an attempt to behave in a way that makes sense as regards φ-ing.

But this is not my analysis. My analysis, if transposed into Clark's language, would not include the last three words, "as regards φ-ing". It would say merely "For A's φ-ing to be an action is for A to φ in an attempt to behave in a way that makes sense." Given that A *is* φ-ing, of course, his attempt to behave in a way that makes sense becomes an attempt to φ (or to stop φ-ing) in a way that makes sense — but only given that he is φ-ing or has at least decided to φ. Before he decided to φ, intelligibility with respect to φ-ing was not his aim. Indeed, my analysis is that in deciding to φ rather than χ or ψ, A had the aim of doing whichever one made more sense, an aim that was not specific to intelligibility with respect to φ-ing.

Why does Clark assume that my analysis posits the more specific aim? I don't think that this assumption flows from the analogy between action and belief. Although my belief that *p* has the aim of accepting *p* only if *p* is true, that specific aim results from substituting *p* for *x* in a general aim with the content "For all *x*, believe *x* only if *x* is true". What favors Clark's assumption is rather the case of searching, where the aim of finding *c,* given that I am searching for *c,* is not an instantiation of a more general aim. That is, I don't have the aim of finding whatever I search for. Actually, the case of searching differs from that of believing or acting in an additional respect. In the latter cases, my aim has conditional content, and so it can be satisfied, on the one hand, by believing what's true or doing what I understand, or on the other hand, by not believing what's false or doing what I don't understand. In the case of searching however, the condition attaches to my having the aim, not to its content. If I am merely considering whether to look for *c* without yet having decided or started to look for it, I

may not have any aim at all with respect to finding *c*, not even an aim conditional on looking for it, and so I have no aim that can be satisfied by my not looking for it. By contrast, as soon as the question arises whether to believe *p*, I have the aim to believe *p* only if *p* is true; and as soon as the question arises whether to do *a*, I have the aim to do it only if it will make sense.

The aim posited by Clark wouldn't be sufficient to constitute behavior as action, because it would not, for example, play any role in a choice among alternative actions, one of the roles that a theory such as mine requires it to play. Nor would it be explicable as a reflexive application of our general aim to understand the world around us. That's why my theory posits a more general aim that is indeed common to all actions.

Response to Tamar Schapiro

Tamar Schapiro takes on the book's organizing metaphor, which is a special case of improvisational acting, in which the agent improvises the role of being himself — an improvised self-enactment. In order to be authentic, the self-enacting agent must portray who he really is; but he can do so by really being who he portrays. As Schapiro sums it up:

> 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.

"[T]he idea," in other words, "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."

This process of reciprocal self-conception and self-enactment results from our occupying two roles simultaneously, the role of improvisational actor and the role of audience to the improvisation. The audience wants to understand the actor's performance, and the actor suits his performance to what the audience will understand, because he is, after all, identical with the audience who wants to understand it. If a self-enactor is to avoid inauthenticity, however, he has to *be* the character, not just act like him, and he has to understand who he *is*, not just who he's acting like. He can do this by enacting traits and attitudes that he has anyway; but he can also do it by acquiring traits and attitude, for example, through the well-known attribution mechanism, by which his inchoate reactions will crystalize into particular attitudes as he manifests them in the corresponding actions.

Schapiro worries that my theatrical metaphor mis-describes the phenomena of ordinary agency. Her worry is that our primary interest as the audience to a performance is not merely to understand it; our primary interests are aesthetic.

I agree. My analogy between agency and theatrical acting is not meant to be perfect. The actor and audience of a theatrical performance have many interests not shared by an agent with respect to his own behavior.

Even so, I think that Schapiro underestimates the theatrical audience's interest in understanding. She says that the audience is primarily interested in escaping into the world of the story through the suspension of disbelief. But whether they can suspend disbelief depends on whether the characters are believable, which depends on whether the actors' behavior is understandable in light of their characters' traits, attitudes, and circumstances. In a scripted drama, of course, the believability of the action has already been secured by the author: the

actors are responsible only for their gestures, postures, facial expressions, tone of voice, tempo, and so on. But, again, these features of their performance make their characters believable by making sense in light of the attributes and actions that the author has assigned them. An actor's vocal, facial, and bodily expressions have to be intelligible as manifestations of the motives and emotions behind his character's actions. That's why an actor often asks the director "What's my motivation?"

In any case, the analog of human action in my argument is not scripted drama but improvisational theater, where the actors are responsible for the action itself, not just the manner in which they enact it. They're responsible for what they say, not just how they say it. And in this case, intelligibility is more salient as an interest for both actor and audience, because the risk of incoherence is greater. As in the case of scripted drama, intelligibility subserves other interests that aren't often operative outside the theater — humor, for example — but it remains a pre-condition for satisfying those interests.

Schapiro interprets me as drawing an overly close analogy between human agents and an audience to their actions, so she raises the objection that agents have no interest in achieving the suspension of disbelief with respect to themselves. I not only concede this point; I positively assert it, since authenticity in action requires forming beliefs that will be true of oneself — "in character" with respect to one's actual character, in a broad sense that includes one's occurrent attitudes as well as enduring traits. Suspending disbelief would open the door to deception — in this case, self-deception — which is hardly conducive to autonomous agency. As Schapiro herself puts it ""[T]he 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."

That people are strongly motivated to bring themselves into accord with their enacted conceptions of themselves has been demonstrated time and again by social psychologists over the course of more than fifty years. It is perhaps the most robust finding of social psychology, since it is evidenced by such phenomena as the well-known cognitive-dissonance effect. That effect is not, as some people think, a response to discordant beliefs; it is a response to discord between one's attitudes and one's behavior, as manifested in the way people come to believe things that they have been induced to assert, or to want things that they have been induced to choose. Psychologists have also demonstrated that people are strongly motivated to bring their behavior into accord with their self-ascribed characters — for example, by manifesting emotions that they have been induced to attribute to themselves. Of course, the researchers who demonstrate these processes do so by manipulating subjects into believing things of themselves that aren't true — at least, not yet — and acting out attitudes that they don't yet have. But the processes don't come to light when they are operating normally. Normally, they lead people to bring their characters, self-conceptions, and behavior into genuine accord, thereby closing any gaps through which the processes themselves might show. Experimental manipulations are therefore necessary to uncover them.

This and other self-consistency phenomena do not require self-absorption, as Schapiro supposes. On the contrary, they are the result of largely unconscious processes. If they weren't, we would already have known about them without the help of social psychologists. My view is that practical reasoning is the conscious tip of this unconscious iceberg. In order to illustrate how it can work without self-directed attention, I quoted an extended passage in which a philosopher expressed thoughts that were ordered not logically but *psychologically*, so as to

express an intelligible process of thought [pp. 20-21]. As the author wrote this passage, he managed to make psychological sense to himself without attending to his psychology.

Schapiro concludes her commentary by examining a sentence about *akrasia* in one of my footnotes. There I wrote that if an agent mis-characterizes his own motives, "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" [p. 15, fn. 7]. Schapiro says, "This makes it sound like the failure involved in *akrasia* is simply a failure to attribute to yourself the motives you actually have." But, she complains, I don't treat this mis-attribution as a failure, because I say that we should engage in wishful thinking, for example, by believing we are brave in order to steel our nerves

What I wrote, however, is that *akrasia* occurs when we extend the habit of effective wishful thinking to cases in which the wish won't come true:

> Without intentionally trying to exploit the power of wishful thinking, the agent can fail to distinguish between the cases in which his thoughts will be self-fulfilling and the cases in which they won't. Accustomed to thinking ahead of the facts about himself and relying on them to catch up, he can fail to notice when they are no longer following" [p. 92].

In other words, the mis-attribution is indeed a failure if the agent fails to act accordingly and, by that means, to bring his character into accord with it.

Schapiro says, "It does seem natural to think that my response [to my own *akrasia*] might be to say, 'I don't know what I've been doing!'" Schapiro doesn't recognize this view of the case as mine. "I do think," she says, "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." It is precisely what I have in mind, and I am glad to find that Schapiro agrees.

Response to Justin D'Arms

Justin D'Arms correctly explains how my theory of self-improvisation extends beyond actions to reactions, and through them to values. What's valuable, I say, is what it makes sense to value — to desire, in the case of the desirable; to admire, in the case of the admirable; to detest, in the case of the detestable; and so forth.

What it makes sense to value is prior in the order of determination to what is valuable. For example, something funny makes sense to laugh at, not because it's funny, but because it's like other things that make one laugh and unlike things that make one wince or gag; because it's shocking to one's expectations, amenable to one's prejudices, titillating but not gross; and so on. I then distinguish between merely laughing at something and finding it funny. Finding something funny is being sensitive, in laughing at it, to whether doing so makes sense. It is, in other words, laughter guided by the thing's being funny, though it doesn't rise to the level of a judgment to that effect. D'Arms explains this view admirably and invites me to join him in considering some problem cases.

In one sort of case, you laugh at a lame joke because you are drunk, while judging that the joke isn't funny. By my definition, finding the joke funny would require laughing at it partly because of the perception that laughing at it made sense. Since you see that this joke isn't funny, your laughter must not amount to finding it funny, so defined. D'Arms points out that your laughter and your judgment are "rationally at odds", and he argues that they wouldn't be at

odds unless your laughter was to some extent evaluative, which would make it a "finding" of humor in the joke. So my definition of "finding funny" seems too restrictive.

In my view, however, the laughter and the judgment are at odds, not because the laughter is evaluative, but because, in light of your judgment, the laughter doesn't make sense. That's all it takes for them to be at odds. No "finding" needed.

In the next case, you laugh at a joke told at a party by a woman to whom you are attracted, but when you retell the joke the next day, you find that it isn't funny. Did you find it funny the night before, in my sense of "find funny"? It all depends. You may have found it funny by mistake, because you misperceived it as having qualities that made it intelligible to laugh at. Perhaps you perceived it as clever, novel, erudite when in fact it was hackneyed and low-minded. You mistook pretty for witty. That can happen (not to me, dear). In the light of day, you see your mistake, and so you rightly find the joke unfunny. You find it to have different qualities on different occasions because one "finding" was based on erroneous perceptions.

D'Arms favors a version of the story in which you realized at the time that you were laughing at the joke partly because you were attracted to the teller. Laughing would then have made sense to you in light of circumstances, and its making sense would have guided you to have a good laugh, just as the intelligibility of crying guided you to have a good cry. Do I want to say, then, that you found the joke funny?

First consider a slightly different case, to which D'Arms also alludes. Sometimes the humor lies not just in the joke but also in the delivery. When the joke falls flat the next day, you say, "Well, it was funny when *she* told it." In that case, what was funny the night before was not the joke itself but the whole performance, joke plus delivery. If you were guided by a perception of the performance as making sense to laugh at, then you indeed found it funny — the whole performance, not just the joke. What you subsequently find unfunny is a different performance.

Now return to the case in which you laugh in part because of the joke-teller's attractions, which also help to make it seem intelligible to laugh at. Should I treat this case like the last, by saying that you found something funny, namely, the joke plus the teller's attractions? Surely, her looks were not funny-making.

Faced with the last two cases, I might respond that the delivery of utterly unfunny content can also make you laugh — Mel Brooks reading the phonebook, for example — whereas good-looking women elicit a smile but never so much as a chuckle. But that distinction seems irrelevant to the question what you found funny in this particular case, where a woman's looks were partly responsible not only for making you laugh but also for guiding your laughter. A better response is that you laughed *at* the delivery as well as the joke in the first case, whereas you laughed at the joke alone in the second, the teller's looks being a circumstance rather than a target of your laughter. The distinction between the circumstances of a response and its target is psychological: there is a psychological fact as to what you are laughing at as opposed to other things that are causally implicated in your laughter. The same distinction has to be drawn in the case of many responses: your indigestion is partly responsible for your anger, but you're angry *at* the other driver; you're afraid *of* the dog, but also partly because of the darkness.

More pressing than the question whether you found the joke funny is whether it *was* funny by your lights, that is, deserving of laughter from someone with your sense of humor. As D'Arms points out, this case shows that what makes sense for you to laugh at need not be what is funny for you, or by yours lights, since you ultimately concluded that the joke wasn't funny but that, under the circumstances, laughing at it made sense. The case therefore appears to refute my

view that what's funny is what it makes sense to laugh at, or more generally, what's valuable is what it makes sense to value.

D'Arms offers me a solution to this problem. The solution is that we can find regularities in how our responses are regularly affected across the board by such things as priming, emotional contagion, heightened arousal, intoxication, and the like. We learn that having laughed at the last joke disposes us to laugh at the next, that having been angry at one meeting disposes us to be angry at the next, and in general that responses tend to carry over from one target to another. We learn that drink makes us quicker to laugh, but that also it makes us quicker to cry or to lose our tempers. We learn to recognize the damping effect of depression, the amping effect of anxiety, effects that apply to many responses alike. All of these psychological mechanisms, once recognized, can contribute to the intelligibility of a particular response on a particular occasion, but they can also be factored out to yield a judgment of what response makes sense other things being equal — "equal" meaning apart from effects not peculiar to that response. Whether a particular telling of a joke is funny is determined by all of the circumstances. But whether the joke itself is funny depends on whether it is in itself such as to make laughter intelligible, where "in itself" means apart from regular sources of interference. The joke is funny only if laughing at it makes sense other things being equal. In sum, the joke in itself wasn't funny because laughing at it made no sense apart from circumstances that alter many responses in predictable ways.

Having offered me this solution to the current problem, D'Arms worries that it undermines my remarks about the interdependence between different responses with respect to their intelligibility. Laughter is dampened by disgust, I say, and so laughing at a joke makes less sense if it is disgusting. Shouldn't I now say that the disgust can be factored out, so that the joke turns out to be funny other things being equal?

No, I shouldn't. If an anodyne joke were told over a disgusting meal, the dampening effect on laughter might be factored out as distortion, and the joke itself could still be such as made sense to laugh at, hence funny. But in this case the joke itself is disgusting, and so it is not funny in itself. We can say that it would be funny if it weren't so disgusting, but we cannot say that it is funny though not when told in those circumstances.

No author could ask for more attentive or more constructive commentators than Phil Clark, Tamar Schapiro, and Justin D'Arms. Their commentaries have given me a welcome opportunity to rethink, revise, and clarify the views presented in *How We Get Along*. I am grateful to them and to the editors of *Abstracta* for a thoroughly enjoyable and enlightening philosophical exchange.