

SELF-DETERMINATION, SELF-TRANSFORMATION, AND THE CASE OF JEAN VALJEAN: A PROBLEM FOR VELLEMAN

Christopher Evan Franklin

~ Penultimate Draft ~

Philosophical Studies 172:10 (2015): 2591–2598

For published version, go to

<https://link.springer.com/article/10.1007/s11098-014-0423-8>

According to reductionists about agency, an agent's bringing something about is reducible to states and events (such as desires and beliefs) involving the agent bringing something about. Many have worried that reductionism cannot accommodate robust forms of agency, such as self-determination. One common reductionist answer to this worry (which I call 'identification reductionism') contends that self-determining agents are identified with certain states and events, and so these states and events causing a decision counts as the agent's self-determining the decision. In this paper I discuss J. David Velleman's identification reductionist theory, according to which an agent is identified with his desire to make most sense of himself. I develop two constraints that an adequate identification reductionist theory must satisfy and show that Velleman's theory cannot satisfy both. In particular, I argue that Velleman's account founders on cases of self-determined self-transformation.

self-determination; identification; causal theory of action; Velleman

According to reductionists about agency, an agent's bringing something about is reducible to states and events involving the agent bringing something about. Reductionists maintain, for example, that Jones's intentionally raising his hand is nothing over and above Jones's mental states and events (such as his desire to ask a question and his belief that he can ask a question by so raising his hand) jointly causing his arm to rise (Davidson 1980; Bishop 1989; Mele 1992). Reductionists are not eliminativists about agential activity: rather they deny that an agent's causing something is a fundamental feature of reality. What is fundamental on the reductionist picture is not agential activity, but the causal activity of states and events.

Although reductionists may be able to account for minimal forms of agency—such as acting unwittingly or merely intentionally—they face serious challenges when attempting

to capture more robust forms of agency, such as self-determination. When we act unwittingly or automatically, it may well be that our activity as agents wholly consists in the activity of purposeful mental states and events that bring about such actions. But cases of self-determination seem different. In cases of self-determination we seem to experience ourselves as sources of activity over and above the activity of our mental states. This experience is most vivid in the case of motivational conflict. Consider the classic case of conflict between duty and desire. I know I should be more attentive to my children, but I am so tired today. Duty pulls one way, desire another. Suppose I make an effort to turn my attention to my children and succeed. In this case it does not seem that my decision was merely a function of my desires and beliefs. These attitudes were in conflict after all and I myself had to decide how to resolve the conflict, or so it seems. It is this experience of myself as seeming to play a causal role over and above the causal role of my desires and beliefs that suggests that I exercised the power of *self*-determination. As Michael Bratman notes, “The image of the agent directing and governing is, in the first instance, an image of the agent herself standing back from her attitudes, and doing the directing and governing” (2005, 195-196; cf. Velleman 1992, 124, 137-139). Reductionists about self-determination are committed to reducing the agent’s role in determining or governing her activity to the causal activity of states and events. But is such a reduction possible?¹

¹ The notion of self-determination I am after differs from the one recently discussed by Buss (2012). She gives a functional definition of self-determination or autonomy: “An agent stands in the self-governing self-relation that interests me here if and only if the roles she plays in forming her intention is such that if she has the general capacity to appreciate the force of moral requirements and if she has adequate opportunity to discern the moral significance of what she is doing, she is blameworthy if her action is morally wrong and praiseworthy if it is morally admirable” (2012, 649). As I understand the literature on identification, the notion of self-determination at play there is not as intimately connected to moral accountability as Buss’s notion. For example, Velleman is explicit that one can be morally accountable for an action even if one does not self-determine it. For Velleman, moral accountability requires the capacity of self-determination, not its exercise (1992, 127 n. 13). For this reason some of Buss’s objections to these accounts of self-determination miss their target (cf. Mitchell-Yellin 2014)

There are different ways for reductionists to reply to this problem, but our attention will concern the *identification reductionists'* proffered solution. According to identification reductionism, an agent's involvement in his self-determined actions is reducible to the causal activity of mental states and events with which he is *identified* (Frankfurt 1971). Standard reductionist models of agency have a difficult time accounting for self-determination because, so maintain identification reductionists, their reductive base is too sparse—including only desires, beliefs, and intentions.² What is needed to account for self-determined actions is an enrichment of the reductive base to include states with which the agent is identified.

Identification reductionists have two central commitments: (i) that there are states with which the agent is identified and (ii) that some of these states are among the causal etiology of all self-determined actions. When an agent is identified with a state, then that state has authority to speak for the agent, so that the causal activity of the state counts as the causal activity of the agent. This notion is best illustrated by an example in which identification is absent. Consider Harry Frankfurt's case of the unwilling addict. The unwilling addict is an addict because he has addictive desires. He is an *unwilling* addict because he wants *not* to have these addictive desires. When the unwilling addict acts to satisfy his addiction, he performs an action such that, although he desires it, he does not *really* want it. Frankfurt contended that in such a case "the unwilling addict may meaningfully make the analytically puzzling statement that the force moving him to take the drug is a force other than his own" (Frankfurt 1971, 13). At the heart of the addict's alienation is passivity with respect to his action: he is overcome by, and thus passive with

² Bishop 1989 offers an example of a reductionist model that restricts the reductive base of states and events to desires, beliefs, and intentions, even when seeking to account for self-determination.

respect to, the desire and the action it leads him to. The desire's leading the addict to action is more of an intrusion upon than expression of his agency. When we act from alienated motives our agency is hijacked and thus the force moving us to act is a force other than our "own."

While Frankfurt's example provides some sense of what is meant by 'identification', we are still in need of a more detailed analysis. Crucially, any such analysis must satisfy the following constraint:

Alienation constraint: if an agent S is identified with a mental state or event E at time t , then S is not alienated from E at t .

We can test the adequacy of an analysis of identification by its extensional adequacy with respect to clear cases of alienation. Any analysis of identification that entails that the unwilling addict is identified with his addictive desires is, for that reason, inadequate.

There is a second constraint that identification reductionism must satisfy:

Scope constraint: for each of S self-determined actions, there must be a mental state or event among the causal antecedents of the action with which S is identified.³

Since identification reductionists are committed to reducing all the activities of self-determining agents, an adequate theory must show that for *every* context in which an agent performs a self-determined action, there is some state or event with which the agent is identified and that brings about this action.

According to J. David Velleman (1992, 1996, 2000a, 2009), there is a single state with which an agent is identified and is behind all self-determined action: namely the desire to act

³ It is important to note the order of the quantifiers in the scope constraint, the existential quantifier being within the scope of the universal. This principle does not entail that for every self-determined action there is the *same* state that the agent is identified with, but only that for every self-determined action, there is *some* state or other that the agent is identified with.

in accordance with reasons.⁴ In what follows I will argue that Velleman's account cannot satisfy both constraints and in particular founders over cases of self-determined self-transformation. But let us first attend to the details of Velleman's theory. Unlike Frankfurt's (1971) higher-order desires, the desire to act for reasons is not simply one desire among many, but rather is a desire constitutive of self-determining agency.⁵ According to Velleman, part of what it is to be a self-determining agent is to be motivated to act by what one takes to be the strongest reason (Velleman 1992, 141-142). To alienate oneself from this state requires that one cease to guide one's life by reasons, and to cease to guide one's life by reasons is to cease to be a self-determining agent. As Velleman observed, "the sense in which an agent cannot disown his desire to act in accordance with reasons is that he cannot disown it while remaining [a self-determining] agent" (Velleman 1992, 142). Therefore, it may well seem that this desire does have authority to speak for the agent.

Before we can assess this proposal, we need to clarify the nature of the attitude that Velleman has seized upon. The desire is not the desire to act in accordance with reasons so described, but rather the desire "to act in accordance with considerations of some particular

⁴ One might reasonably think Velleman (1992) denied that we are identified with this desire. Consider: "If there is...an attitude [that plays the self-determining agent's causal roles], then its contribution to the competition's outcome can qualify as his—not because he identifies with it but rather because it is functionally identical to him" (Velleman 1992, 142). It seems that Velleman is denying that a state's playing my causal roles counts as my playing my causal roles only if I am identified with it. But in fact there is no disagreement here. Just prior to the passage quoted, Velleman argues that an agent cannot "disown" the desire to act for reasons and remain a self-determining agent (Velleman 1992, 141-142). The reference to "disowning" strongly suggests the notion of identification I am deploying. In claiming that a self-determining agent cannot disown this motive and remain a self-determining agent, Velleman is, I believe, claiming that the self-determining agent is essentially *identified* with such a motive. It may well be that Velleman is operating with a somewhat different notion of identification than I am. But what is clear is that Velleman thinks that it is crucial that the state that plays my causal or functional role is one that I cannot (in some sense) disown and it is this very phenomenon that I have used the notion of identification to capture.

⁵ In this way it seems that Velleman's account is immune to Watson's (1975) original objection to Frankfurt's (1971) account. The desire that plays the causal roles of the self-determining agent for Velleman is not a *mere* desire, but a desire that is constitutive of one's self-determining agency.

Let me note here that my presentation of Velleman's account differs from his own. Where Velleman writes of 'an agent' I write of 'a self-determining agent.' This is merely a verbal difference. By 'an agent' Velleman means 'a self-determining agent.' Throughout his article Velleman uses 'action' and 'agency' to mean 'action *par excellence*' and 'full blooded [i.e. self-determining] agency' (cf. Velleman 1992, 124).

kind, which [happen] to be the kind of consideration that [constitutes] a reason for acting” (Velleman 1992, 141).⁶ And which considerations are these? They are the considerations that enable us, or contribute to enabling us, to “make sense of ourselves”: “Considerations weigh in favor of an action, I propose, insofar as they contribute to an overall understanding of the action, given how the agent conceives of himself and his situation” (Velleman 2009, 19). The desire to act in accordance with reasons can thus be more precisely described as the desire to act for considerations that allow one to make sense of oneself.

Velleman compares rational agency to improvisational acting. An improvisational actor seeks to respond in the drama in ways that are “in-character”—in ways that we as the audience would think make sense given the features of the character, such as his beliefs, desires, and personality traits. The notion of “making sense” here is not normative, but folk psychological. Velleman explains:

The understanding that must be possible, if an action is to make sense coming from the character, is a folk-psychological understanding that traces the action to its causes in the motives, traits, and other dispositions of the character. (Velleman 2009, 13)

The self-determining agent, as an improvisational actor, guides his behavior by his conception of what would make sense for him to do, given who he is, in the drama of life.

The desire to act in accordance with reasons is the desire to act in accordance with the *best* reasons, and thus I assume that it is the desire to act in accordance with considerations that enable one to make *best* or *most* sense of oneself that Velleman has in mind. Let the desire to act in accordance with reasons, then, be understood as the desire to

⁶ Velleman prefers this account for two reasons. First, he contends that a *de dicto* reading would be too conceptually demanding, requiring that self-determining agents have the concept of acting for a reason, and a specific conception of what counts as a reason and makes some reasons better than others (Velleman 2000a, 15). Second he contends that “if autonomous action were behavior guided in part by the desire to act in accordance with reasons so described, then an agent could never autonomously do something other than what he believed he had most reason for doing” (2000a, 14 n. 20).

act in accordance with considerations that enable the agent to make best sense of himself (Velleman 2009, 133).

Does this desire satisfy our two constraints? I submit that it does not. Cases of moral transformation suggest that it is possible to perform a self-determined action even though doing so would not enable one to make most sense of oneself. In such cases the agent seeks to embark on a new course of life, one that is deeply at odds with his previous cares, settled plans, and habits of mind. Consider the case of Victor Hugo's Jean Valjean, who after unjustly serving a prison sentence of nineteen years has developed into a bitter, cynical, and hateful man. After being welcomed into a local priest's home and given food and shelter, Valjean repays the priest's kindness by assaulting him and robbing him of his only monetarily valuable possession: his sister's beloved set of silver china. While fleeing, Valjean is arrested by the gendarmes and brought back to the priest, who Valjean claims freely gave him the china. To Valjean's astonishment, the priest confirms Valjean's lie and chides Valjean for forgetting the silver candlesticks, which he then gives to Valjean.

This powerful confrontation with mercy and love forces Valjean to make a disquieting decision. Hugo describes the inner battle as follows:

Obscurely he perceived that the priest's forgiveness was the most formidable assault he had ever sustained; that if he resisted it his heart would be hardened once and for all, and that if he yielded he must renounce the hatred which the acts of men had implanted in him during so many years, and to which he clung. He saw dimly that this time he must either conquer or be conquered, and that the battle was now joined, a momentous and decisive battle between the evil in himself and the goodness in that other man. (Hugo 1862, 116)

Valjean realizes (although only dimly) that he is faced with a decision between two ways of life: the way of hate and the way of love. Clothed in cynicism that was woven through years of suffering injustice, the way of love seems naïve and hazardous. But he cannot dismiss the power of the priest's act that easily—it is a light in his darkness. He realizes that this decision will have a decisive role in shaping how the rest of his life unfolds. As we know, Valjean chooses the way of love and lives a life that inspires us to give and forgive.

Although Hugo withholds details that would allow us to decide with certainty whether Valjean's decision was self-determined, it seems possible to fill in the details of the story to render his decision genuinely self-determined. Assuming then that Valjean's decision is self-determined, we need to ask: did it make most sense of him, given his motives and dispositions of character, to choose the way of love? I do not see how we can say that it did. What moved Valjean was a confrontation with love: the source of transformation was, in this way, external to him. He did not become a new kind of a person and then choose the way of love; he chose the way of love in order to become a new kind of person. It was his contact with goodness, a kind and degree of goodness hitherto unknown to him, that drew him to the way of love. What would have made most sense of Valjean, the hardened, embittered, and hateful man, was to laugh and rejoice—laugh at the priest's naïveté and rejoice at his newfound wealth. This is what an audience would expect and this is part of why Valjean's story is so powerful—it is a welcomed exception to a sad rule of life.

There is of course a sense in which Valjean's decision makes most sense: it is a wise or good or beautiful decision. How could anyone come in contact with such love and not be transformed? But the sense in which it would make sense for Valjean to make the decision is not, and cannot be, the sense that is at play for Velleman. Making sense for Velleman is to be understood, rather, as what would amount to being in-character, as it were, for Valjean.

It is what we, the astute reader of Hugo's novel, would predict Valjean to do.⁷ Admittedly, judging intelligibility is a subtle matter and there will be cases in which it is just not clear to the agent, or to us the audience, which action makes most sense. But if the notion of making sense is to have any determinate content, then it cannot be the case that the first step along the path of moral transformation is *always* the action that, at the time, makes most sense of the agent.⁸ Precisely because it is a step towards moral transformation, it goes against the agent's settled and dominant motives and traits, and so cannot (or at least will not always) make best sense of him. To take the step towards moral transformation is to act *out* of character, and thus to act in a way that makes less sense of oneself than one would have had one acted on the other available options.

To drive this point home, consider Velleman's (2004) remarks about a different case of self-transformation: the case of the Mafioso. The Mafioso endorses a practical identity constituted by "principles of perfect loyalty to the mob and perfect ruthlessness to outsiders" (Velleman 2004, 304). Suppose the Mafioso is ordered to make a hit on yet another innocent bystander: a city official who is unwilling to be bribed. Although the Mafioso has repeatedly followed similar orders in the past, he finds himself in doubt about what to do. Should he give up his practical identity and commit himself to a tamer, kinder way of life? Suppose the Mafioso decides to give up his practical identity as a mobster and embark on a new, morally superior way of life. Here is what Velleman has to say about such

⁷ That the notion of making sense is akin to a kind of predictability on the basis of a character's psychological make-up and situation is made clear in Velleman 2009, 132. When we are seeking to determine what would make sense of a character we are seeking to predicate what the character would do given the kind of person he is and the kind of situation he is in.

⁸ I do not mean to suggest that Velleman's notion of making most sense must entail that for every situation, there is a *single* action that makes most sense of the agent. Presumably there are many circumstances in which there is not single action that makes most sense, but rather a range of options. However, I do insist that if this notion is to have any determinate content, it cannot be that in *every* case of self-determined moral transformation, the transformative action is among the range of options that makes most sense of the agent.

a possibility: “even in the heat of the moment, the mobster might simply step out of his bind: The scales might fall from his eyes, and he might drop his gun and walk away, never to return to his life of crime....In [this] case, I would say, the mobster would not be acting on the balance of reasons that were currently available to him” (Velleman 2004, 310-311). Given the mobster’s practical identity (his character, motives, values, beliefs, etc.), acting in a way that brings about self-transformation does not make most sense of him: the mobster has most reason *not* to change his practical identity. Nonetheless, the first act in the process of self-transformation is not unintelligible. While the mobster is ignoring the balance of reasons in favor of a smaller subset of his reasons, the mobster is still acting for reasons: his action makes sense—it is intelligible. Nevertheless, this action, this first step, Velleman argues, “is not a rational step for the agent to take, all things considered” (Velleman 2004, 311); it is not an action that makes *most* sense of the agent.

Velleman agrees, then, that in cases of self-transformation, the first step is not (always) a step that makes most sense of the agent. But the case of the Mafioso, like the case of Valjean, seems to be a paradigm of self-determined action. The Mafioso is not passive with respect to this first step. He is not overwhelmed by emotion or blinded by desire. He clearly recognizes (perhaps for the first time) the despicableness of his life and decides that enough is enough: it is time to become a new kind of person and live a new kind of life. Surely there is nothing about this case (or Valjean’s) that *forces* us to deny that the Mafioso’s action is self-determined. I conclude that Velleman’s account violates the scope constraint because it cannot account for (all) self-determined actions of self-transformation.

I suspect that some of the plausibility of Velleman’s account results from an ambiguity in his description of the desire that he contends is constitutive of self-determining agency. Velleman almost always describes it as ‘the desire to make sense’ (or ‘the desire to

act in accordance with reasons'). When taken at face value, it does seem impossible to perform a self-determined action contrary to this desire. Self-determined actions must be intelligible; they need not be good or right, but they must be intelligible. Valjean and the perverse agent's actions make sense in this minimal way. However, when taken in this minimal sense, the desire is too weak: acting on the desire to make sense is not sufficient for performing a self-determined action. The unwilling addict's consuming his drug of choice makes sense in this minimal way (he is an *addict* after all), but clearly his action is not self-determined. Thus, Velleman needs it to be the case that it is the desire to perform the action that makes *most* sense that an agent is identified with. But, as we have seen, this motive is not behind all self-determined action. Thus Velleman's account cannot satisfy both constraints.

Bibliography

- Bishop, J. (1989). *Natural agency: An essay on the causal theory of action*. Cambridge: Cambridge University Press.
- Bratman, M. (2005). Planning agency, autonomous agency. In James Taylor (Ed.), *Personal autonomy* (pp. 195-221). New York: Cambridge University Press.
- Buss, S. (2012). Autonomous action: Self-determination in the passive mode. *Ethics* 122, 647-691.
- Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.
- Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68, 5-20.
- Hugo, V. (1862). *Les misérables*. London: Penguin Books.
- Mele, A. R. (1992). *Springs of action*. New York: Oxford University Press.
- Mitchell-Yellin, B. (2014). In defense of the Platonic model: A reply to Buss. *Ethics*, 124, 342-357.
- Velleman, J.D. (1992). What happens when someone acts? *Mind*, 101, 461-481. Reprinted in his 2000b, (pp. 123-143).
- . (1996). The possibility of practical reason. *Ethics*, 106, 694-726. Reprinted in his 2000b, (pp. 170-199).
- . (2000a). Introduction. In his 2000b, (pp. 1-30).
- . (2000b). *The possibility of practical reason*. New York: Oxford University Press.
- . (2004). Willing the law. In Peter Baumann and Monika Betzler (eds.), *Practical conflicts: New philosophical essays*. Cambridge: Cambridge University Press. Reprinted in Velleman 2006, (pp. 284-311).
- . (2006). *Self to self: Selected essays*. Cambridge: Cambridge University Press
- . (2009). *How we get along*. Cambridge: Cambridge University Press.
- Watson, Gary. (1975). Free agency. *Journal of Philosophy*, 72, 205-20.