7

Moving Stories: Agency, Emotion and Practical Rationality

Dave Ward

1   Introduction

How are agents made? Which psychological components, assembled with which structure, do we need to engineer a system that acts for reasons, rather than one that is merely shunted around by external forces? One kind of answer to these questions, associated with G. E. M. Anscombe (1957), identifies the crucial ingredient as epistemic—the qualitative difference between agents and mere movers or behavers depends on a distinctive kind of knowledge that agents have of their own actions. As an agent, your knowledge of mere happenings (like your stomach involuntarily rumbling) is observational, whereas your knowledge of your voluntary actions (like your going to the kitchen to make a snack) is non-observational—you have practical knowledge of your own activity through your very act of carrying it out. Let's assume this framework for now (we'll work through arguments in its favour below).

D. Ward (✉)
University of Edinburgh, Edinburgh, UK
e-mail: Dave.Ward@ed.ac.uk

How should the kind of practical knowledge distinctive of agency be analysed? In particular, what role if any do *emotions* play in enabling knowledge of this kind?

Over the past 30 years, J. David Velleman (1989, 2006a, 2009, 2014) has developed a sophisticated view of practical rationality which aims to address these questions. For reasons we'll set out below, he argues that the kind of practical rationality distinctive of agency depends on the motivational force of a drive towards self-consistency. To be an agent is to be a system that is driven to act in ways that make sense to oneself. Articulating the nature and operations of this drive towards self-consistency thus becomes the key task for a theory of human agency and practical reason. The main question with which this chapter is concerned is: What role do emotions play in structuring such a drive and its operations? As we will see, there are two interpretations of the nature of the drive towards self-consistency in Velleman's work, not always happily related. Agents might aim towards self-consistency by striving to act in ways that make causal-psychological sense—roughly, by acting in ways that maximize the coherence of their beliefs, desires, plans and other psychological states. Or they might aim towards self-consistency by striving to act in ways that make *narrative* sense—acting in ways that make their lives conform to the beats of familiar story structures. Velleman provides compelling reasons to think that emotions are implicated in narrative sense-making—the beats of familiar story structures are *felt* rather than thought through, so agency can consist in shaping your life in order to feel a particular way about it. How are causal-psychological consistency and narrative consistency related? As we'll see below, Velleman initially thought of the latter as an appealing way of redescribing the former, but later came to argue that each kind of consistency conveyed a fundamentally different kind of understanding. I will agree with Velleman that causal-psychological and narrative understanding are distinct, but argue that Velleman's view of their relations is unsatisfactory. Velleman comes to prioritize causal-psychological over narrative self-understanding in a way that jeopardizes the naturalistic credentials of his framework and undermines his claim that a drive towards narrative coherence can constitute a legitimate source of practical rationality. The positive suggestion this chapter aims to motivate

is that the problems facing Velleman's view can be overcome by reversing his order of dependence between causal-psychological and narrative self-consistency—instead of viewing narrative self-understanding as a sophisticated achievement resting on folk-psychological self-understanding, we should understand our drive towards rational self-consistency—and thus our status as agents—as resting on an emotionally structured bedrock of narrative competences.

2   Narrativity and Normativity

In 'The Self as Narrator', Velleman (2005) proposes a connection between narrativity and a kind of normativity which he thinks is distinctive of practical rationality. A self-narrating system—one that issues a running commentary on its own activities—comes to exhibit a form of practical rationality when its past self-narrations constrain its present action selection in a particular way. This is the sort of constraint that happens when, weighing up whether to go for a run or to the pub, I remember saying to myself that morning that I was turning over a new leaf—one involving less pub-going and more running—and factor that into my decision. A bit of self-narration from my past has come back to haunt my present deliberation.

Velleman's guiding example, adapted from Dennett's (1991) reflections on the links between selfhood and narrativity, concerns Gilbert, an intelligent robot that churns out a running commentary on his activities as he goes about his business. Dennett (op. cit., 1991; Dennett and Humphrey 1998) thinks that such epiphenomenal, stream-of-consciousness self-narration is all there is to selfhood. Consequently, Dennett holds that we should understand the self—both in Gilbert's case and our own—as a *narrative fiction*. The self is just the protagonist described or implied by the epiphenomenal tale spun by a self-narrating system. But this protagonist is fictional. When we look beyond or behind the narrative, Dennett argues, we will find nothing but a patchwork of special purpose subsystems, cleverly stitched together to issue in behaviour and self-report that gives the illusory appearance of a unified 'self' working behind the scenes.

In response, Velleman points out that self-narration need not be epiphenomenal. There are cases, including our own, where 'an autobiography and the behaviour that it narrates are mutually determining' (Velleman 2005, p. 211), such that a narrative produced at one time can have repercussions for subsequent deliberation and action selection. The narrative Gilbert produced about his activities in the morning might shape his action selection in the evening. If Gilbert self-narrates that he will soon go into the supply closet for fresh batteries, the memory of this proclamation might tip the balance in a subsequent competition between closet-going and energy-conserving subroutines vying for control of Gilbert's behaviour. In the same way, my morning proclamation that I am now more of a runner than a pub-goer can help determine my behaviour in the evening. According to Velleman, this kind of system or creature—one that is driven to shape its current action selection to cohere with its past acts of narration—is a genuine *agent*. And this specifies a sense in which such a system or creature has, or is, a *self*—as we will see below, Velleman argues that some of their behaviours qualify as exercises of agency by being genuinely *autonomous* or *self-governed*.

One intuitive way of differentiating actions from mere happenings is via appeal to the distinctive way in which the past shapes the present and the future in exercises of agency. Every happening in the universe unfolds according to causal laws, whereby the past and present shape the future. But some happenings are also *acts*, in which an agent can properly be held responsible for the way things unfold. In cases of agency the past and present shape the future not only via causal laws, but also by furnishing an agent with *reasons* to act in one way rather than another.[1] Velleman's suggestion is that we can often understand these reasons in terms of the way in which potential states of an agent's future stand in relations of *narrative coherence* to states of their past. When weighing up whether to run or to pub, my deliberation is shaped

---

[1]On Velleman's understanding this way of drawing the act/happening distinction coincides with Anscombe's epistemic strategy as outlined in the introduction. As we will see, for Velleman the reason-governed way in which the past shapes the future in agency is a function of the distinctive epistemic relation in which an agent stands to her acts.

not only by causal processes determining my current levels of thirst and fatigue, but also by the fact that I *told myself* this morning that I was now the sort of person who goes for a run more often than they go to the pub. Choosing the pub over the run would mess up the coherence of this self-conception. According to Velleman, the fact that these considerations of coherence are in play in motivating my choices can mark me out as an *agent* who is responsive to *reasons*, rather than a mere lump of organic matter, shunted around by external forces.

What exactly, you might be wondering, is narrative coherence? More on this soon. For now, note that feeling the motivational sway of considerations of narrative coherence is, for Velleman, sufficient but perhaps not necessary to qualify as an agent. This is because his earlier work argues, persuasively and influentially, for understanding agency in terms of susceptibility to the motivational sway of considerations of *causal-psychological* coherence—roughly, being motivated to act in a *rationally* coherent way. We will look at this view in a little more detail below. For now, the basic idea is that someone motivated by a drive towards causal/psychological coherence is thereby disposed to avoid acting in ways that are inconsistent with their plans, projects and self-conception, and to ensure that the motives that drive their eventual action are those that they, as a rational agent, have the strongest reasons to endorse.[2] It thus appears, on Velleman's picture, that there are two ways in which a system can qualify as autonomous, and thereby as an agent. Agents can be motivated by a drive towards narrative coherence, as when they try to act so as to make a coherent story out of the unfolding events of their lives. Or they can be motivated by a drive towards causal-psychological coherence, as when they try to maximize the rational consistency of their current activity with their plans, projects and self-conceptions. How are these two ways related to each other? On the face of it, Gilbert's declaration that he's about to go into the closet, and my declaration that I will err towards running over pubbing, could be understood as constraining our respective futures via a drive towards either kind of coherence. Indeed, Velleman initially supposed

---

[2]See e.g. Velleman (2006b, 2009, 2014).

that narrative coherence was simply an appealing way of redescribing causal-psychological coherence:

> My hope was that narrative self-understanding would be less objectionable to those who had previously criticized my view of autonomy as overly intellectualist, because it requires the agent to psychologize about himself. Nobody wants to think that he is studying himself like a rat in a maze; but everybody likes to think of himself as the novelist of his life and hence as the protagonist of his own story. (2007, p. 284)

We all know, however, that great stories can contain gaps and inconsistencies. And any student of philosophy knows that great reasoning doesn't always make for great reading. Velleman thus rightly came to think that narrative coherence is distinct from causal-psychological coherence, and that the pursuit of each kind of coherence involves striving to confer a distinct kind of intelligibility on ourselves and our activities—more on which below. This complicates Velleman's view of the kind of practical rationality that is constitutive of agency:

> The upshot is that practical reasoning is fragmented into the pursuit of two distinct modes of self-understanding. I think that we aim to make sense of ourselves not only in the mode of causal explanation but also in the mode of storytelling. (ibid., p. 285)

How do these two modes of self-understanding relate to each other? And what, a reader of this volume might be wondering, does either of them have to do with emotions? While our ultimate goal is to make progress on the first question, the second question is easier—so let's start there.

3   Understanding via Emotional Cadence

As noted above, Velleman initially thought of narrative self-understanding simply as a folksy way of describing the causal-psychological self-understanding he thought was constitutively involved in practical

rationality. As we also noted, however, the virtue of sheer rational coherence looks different from the virtue of being a good story. Velleman (2003) argues that this is because stories convey understanding in way that is different from a mere description of causally connected events, and do so because they arouse and complete an *emotional cadence*. A description of a sequence of causally connected events becomes a story when it has a beginning, middle and end, with each of these stages understood in terms of their emotional structure. A story's beginning arouses an emotional reaction, which is developed or complicated by its middle, and resolved by its end. It thus conveys an understanding of its events by bringing its audience to *feel* a particular way about them:

> the emotion that resolves a narrative cadence tends to subsume the emotions that preceded it: the triumph felt at a happy ending is the triumph of ambitions realized and anxieties allayed; the grief felt at a tragic ending is the grief of hopes dashed or loves denied. Hence the conclusory emotion in a narrative cadence embodies not just how the audience feels about the ending; it embodies how the audience feels, at the ending, about the whole story. (ibid., p. 19)

A happy ending can help the audience to view the protagonist's trials and tribulations as character-building episodes on the road to her inevitable triumph; a tragic ending can help the audience to view the same struggles as hopeless attempts to delay a foregone conclusion. Emotions thus impart narrative understanding by shaping the audience's experience of events as they unfold, and by imposing a stable and coherent interpretation of events via the anticipation or undergoing of a concluding emotion that provides a unified perspective on the story's contents.

This view of narrative understanding relies on a pair of relatively uncontroversial claims about emotion that I will assume to be correct. The first is that emotions have an intentional component—they involve a way of being directed towards specific objects, or towards the world at large (see e.g. Lazarus 1991; Goldie 2000; Prinz 2004; Slaby 2008). This is one reason why the emotional character of stories imparts understanding—feeling emotions involves having our minds preferentially directed to particular aspects of situations and events. The second claim

is that at least some emotions (those implicated in narrative understanding) have an essentially diachronic and relational character (see e.g. Lewis 2005; Goldie 2012; Barrett 2017). That is, they are partly defined by the ways in which they unfold over time and relate to situations, behaviour, and other emotional states:

> Fear and anger, for example, can be elicited out of the blue… whereas disappointment, gratification and grief must develop out of some antecedent attitude that can be disappointed, gratified, or aggrieved… Emotions like hope, fear, and anger are by nature unstable, because they motivate behavior, or are elicited by circumstances, that ultimately lead to their extinction… By contrast, grief and gratification are stable, because their eliciting conditions and resulting behaviors are not conducive to change. (ibid., p. 15)

As a result, the emotional cadences that are constitutive of narrative understanding have important constraints on how they can and cannot unfold. Not just any patchwork of affective responses will do—the concluding emotion must be one that is capable both of evolving out of and resolving the range of emotions engendered by the story, and of providing a coherent perspective on the story's situations and events. We exercise narrative understanding when our grasp of the elements of a situation or sequence of events is shaped by an emotional cadence— either via occupying a stable emotional state contributing to a perspective which makes sense of a prior sequence of events or feelings, or anticipating how unfolding events and feelings will resolve into such a perspective.

Velleman (2003, 2009) has much of interest to say about how this view relates to theories of literary narrative, and the particular cadences distinctive of various literary and dramatic genres. Our focus here, though, is on Velleman's claim that we usually understand *ourselves* in narrative terms. This amounts to the claim that we are motivated to grasp the activities and episodes that make up our lives in terms of the kinds of familiar emotional cadences that we use to understand stories. In arguing that there is a narrative component to our self-understanding, Velleman is claiming that in making sense of our lives we look for

the kinds of emotional linkages and resolutions that are at work in successfully making sense of a story. To illustrate, consider the ways in which many puzzling patterns of apparent practical irrationality can be explained by distinguishing between these modes of self-understanding. Velleman suggests, for example, that we can account for the 'sunk cost' fallacy—'in vernacular terms, throwing good money after bad' (2007, p. 285), in terms of our propensity to seek emotionally meaningful payoffs to our projects and commitments:

> As any A.B.D. student knows, an endeavor abandoned in mid-course yields no emotional resolution, not even the resolution of disappointment. Continually abandoning one relationship or project for another would leave us not knowing (as we might say) what some parts of our lives had been about. And to know what parts of our lives have been 'about' is just to know how they fit into a story with an emotionally intelligible arc. (ibid., p. 286)

Likewise, Velleman suggests, our propensity to see misfortune as something that can be learned from can be understood in terms of a drive towards narrative self-understanding:

> If life hands you a lemon, the instrumentally rational course may be to throw it away and look for a kumquat instead. Why feel obliged to make something out of a lemon just because you've been handed one? The answer is that a misfortune can be given meaning by a narrative that incorporates it into the remainder of your life, during which its bitterness is still detectable but somehow sweetened. (ibid.)

In each of the above cases, an apparent failure of practical rationality is explained by supposing that our choices aim at an outcome that completes a familiar emotional cadence, thereby providing a perspective according to which prior events are rendered intelligible in terms of their place in a narrative structure. If our overriding practical interest were in totting up utility points then ditching a failing relationship or research project might often be our best option. But if we have a practical interest in attaining a kind of self-understanding that involves

threading the events of our lives into a unified narrative, then the time and suffering costs incurred by persevering needn't put us off. For if we can successfully fight our way to feeling good about our relationship or research, this affords us a perspective from which the struggles along the road are revealed as necessary landmarks or obstacles in the journey of our life. This is the sort of story we all know and understand, even if its instances don't always make strict logical sense.

Let me note a final point in favour of Velleman's view of the role of narrative understanding in our lives. Velleman's reflections on the narrative structure of practical reason can be placed within a long tradition of arguing that important aspects of our selfhood, or self-understanding, must be accounted for via appeal to narrativity (MacIntyre 1981; Taylor 1985; Schechtman 1996; Korsgaard 2009; Rudd 2012). As Jongepier (2016) notes, however, narrative accounts of selfhood face a dilemma—they appear to be caught between *chauvinism* on the one hand, and *triviality* on the other hand. The claim that humans necessarily understand their lives as a story courts accusations of chauvinism—critics of narrative accounts argue that other ways of living and understanding one's life are available, and that these may be more psychologically or morally healthy (Strawson 2004; Lamarque 2007). Galen Strawson, for example, claims 'I have no sense of my life as a narrative with form, or indeed as a narrative without form. Absolutely none' (2004, p. 433). Given the apparent existence of such non-narrative psychologies, Strawson and others argue that claims that people do or should understand their lives as they would understand a story 'are not universal human truths… Their proponents, the narrativists, are at best generalizing from their experience in an all-too-human way' (Strawson 2017, p. 124). In light of such accusations of chauvinism, a natural response is to claim that narrative self-understanding need only be *implicit*—somehow evinced in the structure of our lives or psychology, rather than being a self-consciously held schema for making sense of ourselves (e.g. Stokes 2010; Davenport 2012; Schechtman 2014). However, such accounts face the challenge of spelling out a non-trivial sense of narrativity that nonetheless accommodates the sincere reports of those who find no trace of a story-like structure in their self-understandings. If the narrative structure of Strawson's self-understanding

is buried so deeply that he can't find it within himself in almost 20 years of trying (Strawson 1999, 2017), why suppose it's there at all?

Velleman's account simultaneously avoids the charges of chauvinism and triviality. Conceiving of narrative understanding in terms of grasping emotional cadences suggests how someone might exercise practical rationality in a way that aims at narrative coherence, but do so without having an explicit sense of their life as a continuous unfolding story. If Velleman is right then the question raised by those, like Strawson, who find a narrative self-conception implausible is whether their practical reasoning is often or always motivated by a drive to resolve an emotional cadence. We might be convinced by Velleman's substantive arguments that such drives play an important role in an agent's psychology while denying that this fact must be introspectively obvious to us. The claim that the involvement of emotional cadences is what distinguishes narrative understanding from mere causal-nomological understanding has likewise required substantive argument. Velleman's claim that human agency involves narrative understanding in the form of a drive to act in ways that resolve emotional cadences is thus non-trivial. It is non-chauvinist insofar as it is compatible with granting full human agency to those who have simply failed to notice this structure in their own psychology, so long as that structure does in fact obtain. Velleman's view does imply, however, that the absence of this narrative strand of practical rationality constitutes a defect:

> [C]aring about narrative coherence is part of human rationality. Someone who manages his life as he would a stock portfolio is not just unsentimental; he's lacking in a virtue of practical reason that's available to human beings, who can understand their lives in both causal-psychological and narrative terms. (2007, p. 286)

But this position rests on more than the hasty extrapolation from quirks of individual psychology which grounds Strawson's accusations of 'chauvinism'. If Strawson or a kindred spirit finds no trace of a drive towards narrative coherence in their psychology, the burden of proof now appears to rest with them—they must explain away the appearance that there is something deficient, with respect to the virtues that characterize

human practical rationality, in managing the projects and commitments of one's life like a stock portfolio.[3]

So the sincere self-reports of Strawson and his ilk need not count against Velleman's claims about the role of narrative understanding in our lives. Moreover, as we have begun to see above, and will now consider further, Velleman argues persuasively that such understanding plays a crucial role in a form of autonomous agency that is plausibly a constitutive aspect of human personhood.[4]

<p align="center">*   *   *</p>

So far I have mostly been approvingly summarizing some lines of Velleman's thinking about practical rationality. In a moment I want to begin make some trouble for him. Recall that Velleman originally began appealing to narrative self-understanding as a homely redescription of the causal-psychological self-understanding he argued elsewhere was a constitutive motive of practical rationality. But, as we saw in the last section, there are good reasons to think that understanding events, including the unfolding events of our lives, in narrative terms is distinct from understanding them in terms of mere causal and rational consistency. Velleman thus comes to think that the psychology of human agents is fractured into two 'independent and potentially competing modes of practical reasoning' (2007, p. 287), aimed at two distinct modes of self-understanding. In itself, there's nothing wrong with this. Perhaps it's not nice to be told that we have a fractured psychology, but sometimes the truth hurts. And given the widespread appearance

---

[3]Velleman (2009) appears to have changed his mind on this score, claiming that the drive towards narrative self-understanding is an optional supplement to the drive towards causal-psychological self-understanding for practical reasoners, 'at least to some extent' (p. 204). Below I argue that this is a mistake—it is more plausible to construe causal-psychological understanding as built on a foundation of narrative understanding.

[4]The degree of compatibility between Velleman's views on practical rationality and selfhood and the narrative theories cited above is a delicate issue. A key tenet of Velleman's work on personhood is that 'self' is a multi-purpose reflective pronoun, rather than the designator of a single entity or topic of enquiry (Velleman 2006a). In this chapter our concern is with the mode of reflexive self-understanding that Velleman argues is constitutive of agency. But Velleman is at best agnostic about the relationship between the self-understanding constitutive of agency (understood, roughly, in terms of a drive to act in ways that make sense to ourselves) and the kinds of self-conception with which narrative theorists are frequently concerned (e.g. conceiving of oneself as a loving spouse, committed parent, or hopeless loser). See e.g. Velleman (2001a, 2007, 2009).

of practical irrationality, perhaps the diagnosis of a fracture should not surprise us.

However, the admission that we're dealing with two modes of practical rationality rather than one appears to undermine some important aspects of Velleman's view. First, once it becomes clear that narrative self-understanding is distinct from causal-psychological self-understanding, the appeal to narrative self-understanding can't serve its original purpose of assuaging the worries of critics who accuse Velleman's causal-psychological account of practical reasoning of an implausible intellectualism. Second, given that narrative self-understanding was initially supposed to be equivalent to causal-psychological self-understanding it could borrow its plausibility as a source of practical rationality from Velleman's earlier work on the role of causal-psychological self-understanding in practical rationality. Now that we've denied the equivalence of these modes of self-understanding, that borrowed plausibility must be returned and narrative self-understanding must pay its own way as a legitimate source of practical reason. Can it do so? Not, I will argue, unless we understand causal-psychological understanding as build upon a bedrock of narrative self-understanding. Narrative understanding, it will emerge, should be seen as essential to practical rationality, and the regular cadences that pattern the ebb and flow of our emotional lives are essential to narrative understanding. To see why this is so, however, we first need to know a little more about the motivations for Velleman's causal-psychological view.

4   Three Ways to Build an Agent

Consider (following Velleman 2001b) three ways in which we might go about building an agent. Let's give ourselves a head start by beginning with a creature that is a bundle of competing drives or motivations. How should we modify such a creature to make it capable of agency like our own? One simple strategy is suggested by Donald Davidson's (1963, 1980) influential causal theory of action—for a happening to qualify as an action, it is necessary and sufficient that it be caused in an appropriate way by an intention to act. Your reading these words is an exercise

of your agency because it is a happening caused by your intentions. Your yawning, rumbling stomach, or sagging eyelids are not exercises of agency because they are not so caused. To build an agent, then, perhaps we need only to take a creature that can have intentions and move around, and ensure it is wired up such that some of its movements will be appropriately caused by its intentions. Those movements will qualify as actions.

But note the looming difficulties of spelling out what it is for an action to be *appropriately* caused by an intention. Harry Frankfurt (1975, 1988) points out that our activities can be caused by our desires and intentions without our feeling that those activities are expressive of our will—we can be *alienated* from some of the forces that motivate us. This morning I desired to hit the snooze button on my alarm for a third time and formed a corresponding intention that caused me to do so. But my desire for a third snooze was competing with other, worthier desires: to get up, go for a run, get working and generally achieve great things. As often happens, the desire to snooze won out. But its winning out was, so it seems, an imperfect reflection of my will and agency— if *I* had been allowed to pick the winning desire, I would have gone with 'achieve great things'. Even if we don't wish to say that agency was wholly absent in my snooze-button pressing, it looks like a poor paradigm case for understanding what's distinctive about human action. In fleshing out what it is for an action to be 'appropriately caused' by an intention, our blueprint for designing an agent should allow for the distinction between akratic actions like my snooze-button pressing and those that are more authentic expressions of the agent's will. Frankfurt's hierarchical model of agency tries to do this by appealing to higher-order desires or volitions. This morning my first-order desire to snooze trumped its worthier competitors. But it was at best a partial exercise of agency, since I had a *second*-order desire that my first-order desire to get up and achieve great things should win out. On Frankfurt's hierarchical model, the happenings that are paradigmatic exercises of agency are those caused by a first-order desire that the agent *identifies with* via having a second-order desire to be actuated by it. Our second agent-building strategy, then, involves adding a layer of second-order volitional states to our creature's psychology which can harmonize

(or clash) with the first-order states that actuate its behaviour. When such harmony obtains, we have an exercise of agency.

This looks better—but if we were dissatisfied with our first strategy, we shouldn't accept our second as it is. We objected to Davidson's causal model by noting that we can feel alienated from the intentions that end up moving us—we do not recognize them as authentic expressions of our wills. Velleman (1992, 2001a, drawing on Watson [1982] and Bratman [1999]) notes that we can likewise feel alienated from our second-order desires or identifications. I might find myself—in fact, most mornings *do* find myself—giving a second-order endorsement to a first-order desire for further snoozing that I find mildly shameful. It appears that some second-order endorsements of first-order desires can reflect an agent's will better than others. A second-order endorsement might be a product of ennui, depression, inebriation or mind-control. An account of agency should try to accommodate a sense in which the behaviours flowing from such endorsements are imperfect exercises of agency. Simply appealing to higher levels of reflective endorsement won't help, unless we are given a reason why the endorsements at some particular level can't be produced in the defective ways suggested above. We need, it seems, an account of what it is for an *agent* to identify with an actuating desire, as opposed to that identification being something that simply befalls them.

This leads to our third strategy for agent-building—Velleman's own. Velleman traces the failures of our first two strategies to the appearance that the psychological dynamics they suggested as constitutive of agency can obtain without the agent being suitably involved. The Davidsonian account understands agency as the government of your movements by intentions—but what if those intentions aren't the ones that *you,* the agent, want to be governed by? The Frankfurtian account understands agency as movements that are governed by intentions which have received higher-order endorsement—but what if those endorsements are not ones that *you,* the agent, recognize as expressive of your will? Velleman's strategy is to specify the psychological dynamics of a kind of motivated activity that cannot, by definition, unfold independently of the agent's involvement. Such dynamics would need to encompass whatever state, process, or organization, we thought

constituted agency. So what *does* constitute agency? Our objections to the previous agent-building strategies have relied on understanding the agent as capable of standing back and surveying possible determinants of behaviour (intentions, motives, higher-order volitions), and taking a stance on which should prevail. This surveying and stance-taking, thinks Velleman, is simply the functional role of an agent: 'that of a single party prepared to reflect on, and take sides with, potential determinants of behavior at any level in the hierarchy of attitudes' (1992, p. 19). Given this functional characterization of agency, we can't identify behaviour-governing intentions or second-order desires with the agent, since these are aspects of the agent's psychology which the agent can scrutinize and take a stance on. So, what plays the role of the agent?

> What mental event or state might play this role of always directing and never merely undergoing such scrutiny? It can only be a motive that drives practical thought itself. That is, there must be a motive that drives the agent's critical reflection on, and endorsement or rejection of, the potential determinants of his behavior, always doing so from a position of independence from the objects of review. Only such a motive would occupy the agent's functional role, and only its contribution to behavior would constitute his own contribution. (Velleman 1992, p. 119)

Velleman argues that the motivation constitutive of agency is a drive to act in accordance with reasons (to 'act reasonably' for short).[5] You can stand back from your desires and critically reflect on them; you can also stand back from this process of critical reflection and reflect on whether you agree with its procedure or results. But you, as an agent, cannot disassociate yourself from the drive to act reasonably that informs the results of any such standing back. Suppose that, in your practical deliberation, you find yourself reflecting on your drive to act reasonably, and somehow decide that it is a bad thing—starting from now, you will no longer allow the drive to act reasonably to motivate you. This resolution

---

[5]Note that an agent might be motivated by this drive without having the concept of acting for reasons. It suffices that the drive in fact propels the agent to act reasonably, whether or not the agent understands themselves in these terms (Velleman 1992, pp. 120–121; 2001b, pp. 26–32).

is impossible to put into practice. In trying not to be motivated by the drive to act reasonably you will be trying to do what you judge yourself to have best reason to do—you will be trying to act reasonably, and your attempt will have been self-undermining. Suppose instead, then, that you arrange to have yourself undergo targeted brain damage, removing only your drive to act reasonably. Would you have succeeded in your resolution? It seems not—*you*, the agent that made the resolution, are no longer around to have your actions be motivated, or fail to be motivated, by any drive. In your place is a creature who can still step back, survey the various potential determinants of its behaviour, and ponder what it would be most appropriate to do in light of this survey. But the results of this deliberation need bear no systematic relation to what the creature goes on to do, lacking as it does the drive to act reasonably. What this creature goes on to do is, in the sense relevant for agency, no longer under its control—however its competing drives and impulses are channelled into a single coherent course of action, we have no reason to think that this channelling is a product of an aspect of its psychology with which we can identify the creature qua agent. In giving up your drive to act in accordance with reasons you gave up the motive in virtue of which your activity could intelligibly be traced back to you as an agent.

On Velleman's model, then, adding the motivational force of the drive to act reasonably to one among the agent's competing motives ensures that the resulting behaviour constitutes an exercise of agency:

> [W]hen a desire appears to provide the strongest reason for acting, then the desire to act in accordance with reasons becomes a motive to act on that desire, and the desire's motivational influence is consequently reinforced… This latter contribution to the agent's behavior is the contribution of an attitude that performs the functions definitive of agency; it is therefore, functionally speaking, the agent's contribution to the causal order. (1992, p. 121)

In place of a Frankfurtian higher-order endorsement, it is the addition of the motivational force of the drive to act reasonably to one of your other motivations that constitutes your 'identifying' with it—in

the sense of its belonging to you as an agent. We now have our third and final blueprint for building an agent. We began with a bundle of competing drives and motives and, following Davidson, hooked things up such that our creature's intentions can function as the causes of its behaviour. Following Frankfurt, we then added a capacity for the creature to step back and survey the potential psychological determinants of its behaviour, and give or withhold its endorsement of them. Finally, following Velleman, we added the background drive to act reasonably. This drive can now determine how exercises of the capacity to step back and survey are translated into behaviour, by joining its motivational force with that of the course of action that best accords with its aims. My evening decision to eschew the pub and go for a run, and Gilbert's decision to head to the closet for fresh batteries are, for Velleman, exercises of agency because our prior resolutions are modulating the competition between our present motives in a way that is guided by a drive to act reasonably that is a constitutive requirement of agency.

5   Two Sources of Practical Reason

The above sketch of Velleman's interrelated conceptions of agency and practical reason brings some problems for his account of the relationship between causal-psychological and narrative self-understanding into view. We have already noted that Velleman revised his earlier view that narrative understanding was a homely redescription of causal-psychological understanding, now holding that 'causal explanation and storytelling convey fundamentally different modes of understanding' (2009, pp. 185–186), and thus that practical rationality is driven by the pursuit of two potentially competing drives towards distinct ways of making our lives intelligible to ourselves. Velleman is thus unable to appeal to narrative self-understanding to allay worries about the intellectualist, or cognitively demanding character of his causal-psychological account of practical rationality. Is this a problem? That depends, I think, on our explanatory interests. Velleman's chief concern is with the logical structure of mature human practical rationality. Maximizing one's causal-psychological intelligibility to oneself is posited as the ideal at which

such practical rationality aims. The fact that this ideal is one that few people would consciously recognize as animating their behaviour, and is difficult or impossible to obtain, need not count against Velleman's characterization. I, for example, could not tell you the statistical rules and regularities which govern optimal betting patterns in poker, and am confident that no game of poker I ever play will approximate those patterns—but it might nonetheless be true that my poker playing activity qualifies as an attempt to play poker insofar as it attempts to track those rules and regularities, and can be assessed as going better or worse insofar as it succeeds or fails at this.

However, if our goal is to naturalistically explain the capacity for practical rationality as it is realized in and exercised by humans, we should be less sanguine about the idealizing and intellectualist features of Velleman's view. Consider how we might attempt to put the previous section's blueprints for building an agent into practice. Given current cognitive-scientific knowledge, it's plausible that we could indeed engineer systems with competing sets of motives, and even with the higher-order meta-motives to which Frankfurt appeals. But Velleman's suggested addition of a global background motivation to act reasonably that modulates the interactions of a system's existing motives and behaviours is more mysterious—this is not a component that engineers of real or artificial systems have lying around in their workshops. As a piece of construction advice it is little more helpful than being told to add an immaterial soul, or a perpetual motion machine. We might thus agree that Velleman is right about the psychological structure that our completed agent will instantiate, while objecting that his account does nothing to help us with the naturalistic goal of understanding how agents can be constructed out of simple worldly materials.[6] Velleman's shift from describing his view of practical reason in terms of mature rational agents and idealized psychological theories to simple robots and

---

[6]In the interests of simplicity I'm writing as if we can legitimately pull apart the explanatory projects of characterizing the constitutive structure of some phenomenon and spelling out how that phenomenon does or could arise in nature. But note that many will be inclined to resist this separation, holding that empirical inquiry and data are our best guides to metaphysical structure. Such folks should be especially sceptical of Velleman's intellectualism about practical reason.

viscerally understood stories looked like a promising step towards this naturalistic goal. But the admission that this shift represents a change in topic rather than a redescription shows that no progress has been made. If we care about understanding how agency emerges in the natural world then while Velleman might show us where our account must end up, he can't tell us how to get there.

Intellectualist worries about Velleman's view thus resurface once we sever the link (or come to believe that there was no link to be severed) between causal-psychological and narrative self-understanding. Severing this link also appears to undermine the claim that a drive towards narrative self-understanding can make a constitutive contribution to agency. We saw in the last section that Velleman posits the drive towards causal-psychological self-understanding as the crucial ingredient of agency because he thinks it can be plausibly identified with the functional role of the agent—it is this drive that animates the activity of weighing up and selectively modifying the force of the motives vying for control of behaviour. And this activity can be plausibly understood as the essential form of agency. What would happen if this drive towards causal-psychological self-understanding were replaced with a drive towards narrative self-understanding? The latter drive is equally capable of adjudicating between, and modulating the influence of, competing motives on behaviour. The only apparent difference is that this adjudication and modulation is driven by standards of narrative, rather than causal-psychological coherence. Perhaps, then, we simply have two candidate occupants of the functional role of the agent instead of one—and this dual occupancy is what accounts for the fractured nature of practical rationality. But this is too quick. Consider why the drive towards causal-psychological self-understanding seemed like a plausible candidate for the role of the agent in the first place. Velleman clarifies:

> Why would behaviour produced by such a mechanism be any more attributable to the person than that produced by other causes? The answer is that a person is somehow identified with his own rationality. As Aristotle put it, 'Each person seems to be his understanding.' Hence causation via a person's rational faculties qualifies as causation by the person himself. (2001b, p. 17)

This quote reveals the rationalist intuitions at work in Velleman's account of agency. The fact that adjudicating between, and modulating the motivational force of, competing determinants of behaviour is governed by rational standards is what makes it an intuitively plausible candidate to be cast in the role of the agent, qua rational intellect, thinks Velleman. In defence of this intuition we might note, as we did above, that an account of agency should explain the difference between behaviour governed by mere causal forces and behaviour guided by reasons. Velleman's account of agency straightforwardly accomplishes this by understanding the mechanisms that determine an agent's behaviour in terms of rational standards applied by rational faculties. This link does not appear to obtain if we try to construe a drive towards narrative coherence as a motivating force of agency. After all, our initial characterization of the distinction between narrative and causal-psychological coherence relied on stressing that a sequence of events can make narrative sense—via establishing and completing an emotional cadence—without being fully rationally intelligible. Think, for example, of the way in which myths and fantasies with supernatural elements, or narratives consisting of series of causally and spatiotemporally dislocated vignettes, can convey understanding. It thus appears on Velleman's view that we have no reason to attribute behaviour stemming from the motivational force of a drive towards narrative understanding to the agent qua rational intellect, or to view this behaviour as guided by reason in the way characteristic of agency. A drive towards narrative understanding no longer looks like a plausible source of practical rationality.

Velleman (2009) appears to modify his earlier view of the relationship of narrative and causal-psychological intelligibility in ways that bear on these complications. At the end of §3 we saw Velleman (2007) claim that lacking a drive towards narrative self-understanding (as might be manifested by managing one's life like a stock portfolio) amounts to lacking a virtue of practical rationality. But Velleman (2009, p. 204) describes narrative self-understanding as a supplement to causal-psychological self-understanding that is 'to some extent' optional, citing self-descriptions such as Strawson's (2004, 2017; §3, above) as evidence in favour of such a view. And he takes pains to emphasize that he now views narrative understanding as based on causal-psychological understanding:

> I assume that narrative intelligibility requires psychological intelligibility, but not vice versa. That is a story requires action, and action has to be intelligible as caused by the attitudes and attributes of a character, lest it count as no more than mere behavior. (2009, p. 185)

I will try to motivate an alternative conception of the relation between narrative and psychological intelligibility in the chapter's final section, below. For now, note that these qualifications to Velleman's view further undermine his claim that narrative intelligibility can function as a constitutive motive of practical rationality. Velleman now views the drive towards causal-psychological self-intelligibility as sufficient for practical rationality, with a drive towards narrative self-intelligibility as an optional supplement. Now that Velleman insists that narrative intelligibility presupposes psychological intelligibility, we can specify why there should be a link between behaviours motivated by a drive towards narrative self-intelligibility and agency: the ability to make narrative sense to oneself presupposes the ability to make psychological sense to oneself which Velleman identifies with the agent, qua rational intellect, and thereby qualifies as an exercise of practical rationality. But this is not because the drive towards narrative self-intelligibility is an independent source of practical rationality—it is because it is an optional way of manifesting the psychological drive with which Velleman originally identified the agent.

Here, then, is the situation. Velleman began with a plausible but demanding intellectualist account of the logical structure of practical rationality. He then added reference to a drive towards narrative intelligibility to his account in a way that appeared to make it less intellectualist, more naturalistically tractable, and better able to accommodate the apparent psychological diversity of ways in which we experience our activities as making sense to us. The cost of this was an apparent fracture in the structure of practical rationality. Upon the closer inspection we have given Velleman's views in this section, however, all these appearances are illusory. Narrative intelligibility, according to the most recent incarnation of Velleman's views, depends on the same capacities of the rational intellect he has always emphasized. Appealing to it thus moves us no closer to the goal of a naturalistically tractable account of

the constitution of agency. And when we act in ways that make more narrative than causal-psychological sense (as when we persist in the face of adversity) we are failing to properly realize our status as practical reasoners and thus as agents, since our actions fall short of the only standards that really matter for agency—those of causal-psychological consistency. The silver lining is that practical rationality no longer appears fractured—it has the same monolithic, intellectualist structure as in Velleman's earlier work.

6   A Strange Inversion of Practical Reasoning

The appropriate response to these problems, I think, is to reverse the relations of dependence Velleman sees between capacities for rendering events intelligible in causal-psychological and narrative terms. Velleman (2009) argues that understanding a narrative presupposes the capacity to understand events in terms of the inferentially related psychological states that caused them:

> Narrative understanding is not a substitute for causal-psychological self-understanding. The tale of Oedipus makes for a good story even if we don't understand why the prophecy of Tiresias came true; but the tale would make no sense of any kind if we didn't understand, from one scene to the next, why Oedipus was behaving as he did… Hence practical reasoning aimed at narrative self-understanding can supplement but not replace reasoning aimed at self-understanding in causal-psychological terms. (p. 203)

But this line of thinking undersells the scope of Velleman's original account of narrative understanding. That account specified a possible way of understanding the relationship between *any* sequence of events, in terms of the way in which its constituents hang together according to a familiar emotional cadence. Velleman rightly notes above that, in some cases, we must exercise relatively sophisticated folk-psychological capacities in order to grasp the events we aim to make intelligible before we can perceive any emotional cadence that unites them. Fully grasping

the emotional beats of a complex espionage thriller or Shakespearean farce requires some prior understanding of the psychological states of the characters and the ways in which they are revealed or hidden by their activities. But relations between non-psychological states can also be grasped according to their place in a familiar emotional cadence. The very locution of 'cadence' draws attention to the fact that we can exercise narrative understanding in experiencing the affective structure of non-psychological states or events such as a series of musical notes. In such a case we, the subjects of understanding, are ushered through a series of affective psychological states that determine our perspective on the object of understanding—but this process need not involve imputing psychological states to ourselves, the object of understanding, or an author behind it. Even when the object of our narrative understanding is ourselves or other agents, some behaviour can be rendered intelligible via its fit with an emotional cadence without deploying the sophisticated folk-psychological capacities emphasized by Velleman. Think of an infant learning to enjoy the activity of peek-a-boo, or simple turn-taking games. Children who are too young to reliably parse sequences of behaviour into intentional actions, or to manifest sensitivity to the potential divergence between behaviours and their psychological causes, can nonetheless come to grasp ritualized play events and the expressive behaviours they involve in terms of their place in a familiar emotional cadence. This kind of understanding is manifested by the way in which infants anticipate the next stage in the play, and their reactions when such anticipations are confirmed or frustrated. To understand events in this way is not yet to understand a narrative, or the psychological states of any of the parties involved—but it nonetheless requires exercising the mode of understanding relations between events which Velleman argues we employ when grasping the events in a narrative. The positive suggestion with which I close this chapter is that such exercises of the capacity for narrative understanding, when suitably encultured and scaffolded, form the bedrock for the mature folk-psychological capacities which Velleman makes central to practical rationality.

Luckily, this is not a case I need to make from scratch. The past 20 years or so have seen a host of important work that aims to question

nativist, modular and cognitivist accounts of capacities for folk-psychological understanding by emphasizing the role of embodied, emotional and situational factors, and by drawing attention to various social scaffolds and dynamics that support and constrain our understanding of ourselves and others (e.g. Hurley 1998; McGeer 1996, 2007; Mameli 2001; Sterelny 2003, 2012; Zawidzki 2008, 2013). A key strand of that literature argues that our folk-psychological apparatus for interpreting behaviour serves not only descriptive and explanatory purposes, but also plays a *regulative* role. This kind of regulation is at work when Gilbert's avowal that he plans to go to the closet, or my avowal that I intend to go for a run this evening, shape our subsequent deliberation. Our respective bits of self-directed folk-psychology function not only as descriptions, but as motivations to conform to those descriptions in our future behaviour. The same regulative pressure is at work, these theorists argue, when I interpret others or when they interpret me—a community of sociable, eager-to-please, but quick-to-censure intentional agents will give rise to an emergent pressure to make one's own activities and those of one's peers conform to shared standards of intelligibility. To see why, and the role that narrative understanding might play here, let's return to our attempt to specify a blueprint for building an agent.

Suppose once again that we begin with a creature that is a bundle of competing drives and motivations. Add the capacity for that creature to occupy emotional states, understood as valenced bodily states that preferentially direct attention to specific aspects of the creature's situation. Now add the capacity for those states to link together in recognizable sequences or cadences which the creature can grasp via anticipation or retrospection (the kind of primitive capacity for narrative understanding which I suggested above may be at work when young children grasp the structure of events in simple play activities). Suppose also that the range of emotional cadences via which we can understand events is developmentally plastic—while the emotional cadences that structure understanding of some kinds of events (such as those involved in a game of peek-a-boo) might be innate, exposure to the right events and associated bodily states can inculcate a grasp of new cadences. The emotional beats that structure our understanding of complex murder mysteries or espionage thrillers, for example, are unlikely to be innately understood.

But a subject with a developmental trajectory that includes repeated exposure to the right books or films can come to anticipate or retrospect the relationships between events in such plots as unthinkingly as a child anticipates the next event in a game of peek-a-boo. Two final suppositions about our creature—first, it has an innate appetite for intelligibility, and distaste for unintelligibility. In this context, this simply means that our creature likes it when events fit familiar emotional cadences—it's usually happier when its anticipations pan out, or when its retrospection fits recent events into a familiar emotional structure (babies, for example, prefer it when games of peek-a-boo follow established protocol). Second, our creatures are sociable—they like to please, don't like to displease, and are sensitive to indications of pleasure or displeasure from their fellow creatures. Such a creature will be driven to understand its world via the emotional cadences with which it is familiar, and will acquire familiarity with new emotional cadences via repeated exposure over the course of its development.

Now, suppose further that we scaffold our creature's development in a particular way—we surround it with other sociable creatures who likewise crave intelligibility and approval, and who exhibit patterns of behaviour that is intelligible in folk-psychological ways. Suppose also that our creature's developmental trajectory involves regular interaction with caregivers who repeatedly expose it to affectively engaging narratives conveying stereotypical folk-psychological wisdom via the emotional beats of their story structures (Sterelny 2003, 2012; Hutto 2008). When our creature's development is scaffolded in this way, it will acquire a drive to make its own behaviour intelligible in terms of whatever emotional cadences structure the understanding of its fellow creatures. Perceptible regularities in the behaviours of its fellow creatures will eventually be narratively understood in terms of the reliably similar structured patterns of embodied affective reactions—emotional cadences—they produce for our creature. A helping hand in learning the most important of such cadences is provided by the creature's caregivers, who supply a regular and easily digestible diet of tried-and-tested narratives and interactive practices that highlight patterns which are particularly important for understanding and predicting behaviour in their community. Crucially, because our creatures are sociable and

intelligibility-craving, they will acquire a drive to render *themselves* intelligible in terms of the structure and patterns that characterize their understanding of their peers. They themselves are objects whose behaviour their peers are attempting to understand; as intelligibility-craving creatures, their peers will be happier when the creature behaves intelligibly by their lights; as a sociable creature, our creature has a drive to make its peers happy, and will thus acquire a motivation to behave intelligibly by the lights of their peers. For the kinds of creatures we have described, intersubjective understanding thus acquires a *regulative* as well as a *predictive* role (McGeer, Mameli, op. cit.). Given this, creatures with the features and social scaffolding described above will eventually acquire a drive to act in ways that conform to whatever standards of intelligibility are prevalent in their community.[7]

This is how a drive towards folk-psychological intelligibility could be built out of a suitably scaffolded drive towards narrative intelligibility. We have reversed the relations of dependence that Velleman (2009) argues obtain between these drives. In doing so, we have resolved the problems I raised for his account in §4. While the blueprint for creature-design just presented is more complex than the Vellemanian blueprint of §3, it is also more naturalistically tractable—it is easier to see how we could realize that blueprint with cognitive structures and processes whose operations we broadly understand. Where Velleman speaks of bolting a drive to act reasonably on to a bundle of competing motivations, I have suggested how such a drive could gradually emerge from a suitably scaffolded, developmentally plastic, motivated creature that can make sense of events via emotional cadences. While defending the

---

[7]One question that a fully developed version of the present proposal should address concerns the origins of the practices of intersubjective understanding that scaffold development. Given the design specifications above, a suitably developmentally plastic creature will acquire a drive to act intelligibly by the standards of whatever practices of intersubjective understanding are present in its community—hence sophisticated folk-psychological understanding can emerge from primitive narrative understanding. But how do sophisticated practices of intersubjective understanding come to exist in the scaffolding environment in the first place? A naturalistic story here must show how folk-psychological practices like our own can bootstrap themselves into existence from humbler beginnings given the existence of a community of creatures with the psychological structure described above. Sterelny (2003, 2012) and Zawidzki (2013) in particular have provided plausible, empirically informed suggestions about how such bootstrapping could occur.

empirical credentials of this picture is a task for another time, I think the prospects are good. In addition to the work by proponents of the regulative dimension of folk-psychology cited above, work on embodied resonance and interpersonal coordination (as suggested by e.g. Hurley 2008) and constructionist approaches to emotion (e.g. Barrett 2017) yields many suggestions about how we humans might instantiate the blueprint sketched in this section. Viewing the drive towards folk-psychological intelligibility as a culturally scaffolded drive towards narrative intelligibility helps restore the credentials of the latter drive as a legitimate source of practical rationality. On Velleman's (2009) view, the drive towards narrative intelligibility can qualify as a source of practical rationality only because it is an optional way of manifesting a drive towards causal-psychological intelligibility. But the blueprint above shows how we can agree with the rationalist intuitions motivating Velleman's account of practical rationality (whereby the drive to act reasonably is plausibly identified with the agent because there is a privileged link between personhood and rational faculties) while still holding that the drive toward narrative intelligibility is essential to agency. This is because the latter drive makes an essential contribution to the motive force of the drive to act reasonably—creatures that meet the design specifications above acquire a motive to act reasonably only because their drive towards narrative understanding takes on a particular, culturally inculcated form whereby their sociable and intelligibility-craving nature motivates them to self-regulate their behaviour according to whatever folk-psychological principles structure the intersubjective understanding of their peers. More tentatively, the blueprint above might suggest how particular instances of a drive towards narrative intelligibility can qualify as a source of practical rationality in its own right. On certain permissive social practice accounts of reasons (e.g. Rorty 1979, 1999), a good reason is simply one that most of our peers are happy to accept. Such views would allow the norms of practical rationality to be fixed by the practices of intersubjective understanding that a given community actually employs, even when those standards fall short of strict standards of calculative rationality. As we saw above (§2), an appealing feature of a narrative conception of practical rationality is that it makes sense of the appearance that we do indeed make

sense of our own activity and that of others in diverse ways—persisting in the face of adversity can appear a reasonable course of action to us even when it is likely to fail to maximize the satisfaction of an agent's projects and preferences. If the norms of practical rationality are fixed via reference to actual practices of intersubjective understanding then the blueprint above suggests how any such norms can acquire motive force for a suitable enculturated creature, regardless of how closely they align with the norm of maximizing causal-psychological intelligibility.

I have argued, then, that we can save the best features of Velleman's view of practical rationality by inverting it—rather than founding the capacity for narrative understanding on sophisticated capacities for causal-psychological understanding of ourselves and our peers, I have sketched how the latter capacities might be built out of socially scaffolded capacities for narrative understanding, and how such scaffolding can imbue the principles that govern the frameworks we use to make sense of each other with the motive force that Velleman argues is constitutive of practical rationality. On this picture, practical rationality is essential to agency, narrative understanding is essential to practical rationality, and the regular cadences that pattern the ebb and flow of our emotional lives are essential to narrative understanding. It is in virtue of narrative cadences that move us emotionally that we can be moved agentially. That's my story, and I'm sticking to it.

References

Anscombe, G.E.M. 1957. *Intention*. Oxford: Blackwell.

Barrett, L.F. 2017. The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization. *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Bratman, M. 1999. Identification, Decision, and Treating as a Reason. In *Faces of Intention: Selected Essays on Intention and Agency*, 185–206. Cambridge: Cambridge University Press.

Davenport, J. J. (2012). *Narrative Identity, Autonomy, and Mortality: From Frankfurt and MacIntyre to Kierkegaard*. New York: Routledge.

Davidson, D. 1963. Actions, Reasons and Causes. *Journal of Philosophy* 60 (23): 685–700.

———. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Dennett, D., and N. Humphrey. 1998. Speaking for Our Selves. In *Brainchildren: Essays on Designing Minds*, 185–206. Cambridge, MA: MIT Press.

Frankfurt, H. 1975. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68 (1): 5–20.

———. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

Goldie, P. 2000. *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.

———. 2012. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.

Hurley, S. 1998. Vehicles, Contents, Conceptual Structure and Externalism. *Analysis* 58 (1): 1–6.

———. 2008. The Shared Circuits Model (SCM): How Control, Mirroring, and Simulation Can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences* 31 (1): 1–22.

Hutto, D. 2008. *Folk-Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.

Jongepier, F. (2016). Towards a Constitutive Account of Implicit Narrativity. *Phenomenology and the Cognitive Sciences* 15 (1): 51–66.

Korsgaard, C. 2009. *Self-Constitution: Agency, Identity and Integrity*. Oxford: Oxford University Press.

Lamarque, Peter. 2007. On the Distance Between Literary Narratives and Real-Life Narratives. *Royal Institute of Philosophy Supplement* 60: 117–132.

Lazarus, R.S. 1991. *Emotion and Adaptation*. New York: Oxford University Press.

Lewis, M.D. 2005. Bridging Emotion Theory and Neurobiology Through Dynamic Systems Modeling. *Behavioral and Brain Sciences* 28 (2): 169–193.

McGeer, V. 1996. Is 'Self-knowledge' an Empirical Problem? Renegotiating the Space of Philosophical Explanation. *Journal of Philosophy* 93 (10): 483–515.

———. 2007. The Regulative Dimension of Folk Psychology. In *Folk Psychology Reassessed*, ed. D.D. Hutto and M. Ratcliffe, 137–156. Dordrecht: Springer.

MacIntyre, A. 1981. *After Virtue: A Study in Moral Theory*. London: Duckworth.

Mameli, M. 2001. Mindreading, Mindshaping, and Evolution. *Biology and Philosophy* 16: 597–628.

Prinz, J. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press.

Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.

———. 1999. Ethics Without Principles. In *Philosophy and Social Hope*, 185–206. London, UK: Penguin Books.

Rudd, A. 2012. *Self, Value and Narrative: A Kierkegaardian Approach*. Oxford: Oxford University Press.

Schechtman, M. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.

———. (2014). *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford University Press.

Slaby, J. 2008. Affective Intentionality and the Feeling Body. *Phenomenology and the Cognitive Sciences* 7 (4): 429–444.

Sterelny, K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.

———. 2012. *The Evolved Apprentice*. Cambridge, MA: MIT Press.

Stokes, P. (2010). What's Missing in Episodic Self-Experience? *Journal of Consciousness Studies* 17(1–1): 119–143.

Strawson, G. 1999. The Self and the SESMET. *Journal of Consciousness Studies* 6 (4): 483–518.

———. 2004. Against Narrativity. *Ratio* 17 (4): 428–452.

———. 2017. *The Subject of Experience*. Oxford: Oxford University Press.

Taylor, C. 1985. Self-Interpreting Animals. In *Philosophical Papers I: Human Agency and Language*. Cambridge: Cambridge University Press.

Velleman, J. David. 1989. *Practical Reflection*. Princeton: Princeton University Press.

———. 1992. What Happens When Someone Acts? *Mind* 101 (403): 461–481; Reprinted in Velleman (2014), pp. 100–124.

———. 2001a. Identification and Identity. In *The Contours of Agency: Themes from the Philosophy of Harry Frankfurt*, ed. S. Buss and L. Overton. Cambridge, MA: MIT Press; Reprinted in Velleman (2006a), pp. 330–360.

———. 2001b. Introduction. In *The Possibility of Practical Reason*; Reprinted in Velleman (2014), pp. 11–46.

———. 2003. Narrative Explanation. *Philosophical Review* 112 (1): 1–25.

———. 2005. The Self as Narrator. In *Autonomy and the Challenges to Liberalism: New Essays*, ed. J. Anderson and J. Christman. Cambridge: Cambridge University Press; Reprinted in Velleman (2006a), pp. 203–223.

———. 2006a. *Self to Self: Selected Essays*. New York, NY: Cambridge University Press.

———. 2006b. The Centered Self; Velleman (2006a), pp. 253–283.

———. 2007. Reply to Catriona MacKenzie. *Philosophical Explorations* 10 (3): 283–290.

———. 2009. *How We Get Along*. New York, NY: Cambridge University Press.

———. 2014. *The Possibility of Practical Reason*, 2nd ed. Ann Arbor, MI: Maize Books.

Watson, G. 1982. Free Agency. In *Free Will*, ed. G. Watson, 205–220. Oxford: Oxford University Press.

Zawidzki, T. 2008. The Function of Folk Psychology: Mind Reading or Mind Shaping? *Philosophical Explorations* 11 (3): 193–210.

———. 2013. *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.