Free Will, Narrative, and Retroactive Self-Constitution

Roman Altshuler

[Forthcoming in *Phenomenology and the Cognitive Sciences*. Please cite the published version. The final publication is available at Springer via http://dx.doi.org/10.1007/s11097-014-9365-z]

Abstract: John Fischer has recently argued that the value of acting freely is the value of self-expression. Drawing on David Velleman's earlier work, Fischer holds that the value of a life is a narrative value and free will is valuable insofar as it allows us to shape the narrative structure of our lives. This account rests on Fischer's distinction between regulative control and guidance control. While we lack the former kind of control, on Fischer's view, the latter is all that is needed for self-expression. I first develop Fischer's narrative account, focusing on his reliance on temporal loops as giving us control over the value of our lives. Second, I argue that the narrative account grants us greater power over the past than Fischer would allow because narrative is essential, rather than supplementary to, practical rationality. Narrative thus allows not only for changes in how we feel about episodes in our past but what those episodes in fact were and it thus allows for a kind of retroactive self-constitution. Finally, I suggest that this modification of the narrative view opens the possibility of a conception of freedom stronger than guidance control. It provides us with a limited power of self-origination as well as the ability to choose between alternative possibilities by tweaking the motivational states that lead to that choice.

John Martin Fischer has recently argued that the value of acting freely is primarily a species of the value of self-expression. Fischer holds that the value of a life cannot consist simply of the value of the goods in that life added together, but of the ordering of those goods—the value of a life, in other words, is a narrative value. Free will is, on this view, good insofar as it allows us to shape the narrative structure of our lives, thus enabling us to endow those lives with meaning. This account, furthermore, rests on Fischer's distinction between the kind of freedom involved in regulative control and guidance control. While we lack the former kind of control, on Fischer's view, the latter is all that is needed for self-expression. In this paper, I will first develop Fischer's narrative account, focusing on his reliance on temporal loops as giving us a special kind of control over the value of our lives, especially the value of past events. Second, I will argue that narrative is essential to practical rationality, and that narrative loops allow us not only to change how we feel about episodes in our past, but what motives entered into our past reasoning. If this is right, the narrative account grants us far greater power over the past than Fischer would allow—it allows for a kind of retroactive self-constitution. I will conclude with a tentative proposal suggested by this expansion of Fischer's view; I suggest that this transformation of the narrative view opens the possibility of a conception of freedom far stronger than guidance control and closer to libertarian origination.

1 Narratives, Lives, and Temporal Loops

Since Fischer draws extensively on David Velleman's conception of narrative, we can begin with a brief overview of that position. Velleman sets out to find the feature that is distinctive of the "narrative connection" or that constitutes "narrative coherence," that is, the feature that gives proper narratives their

peculiar explanatory force over and above that possessed by non-narrative explanations, such as those given in the sciences. After canvassing several competing proposals, Velleman argues that what makes something into a narrative, rather than a simple chronicle of events or a causal sequence, is the emotional arc traced in the telling of those events (Velleman 2003; 2009). A narrative, unlike other genres, is ordered in such a way as to arouse a series of emotions in order to ultimately resolve them. The audience of a narrative can understand its events "first, because it *knows how they feel*, in the sense that it experiences them as leading it through a natural emotional sequence; and second, because it *knows how it feels about them*, in the sense that it arrives at a stable attitude toward them overall" (Velleman 2003, 19). Narrative understanding, thus, involves undergoing a familiar sequence of emotions culminating in an overall sense of how the story as a whole feels.

This will have consequences for how we evaluate literature (a story in which characters develop, for example, is likely to be more satisfying than one where they change rapidly from moment to moment, making their actions seem essentially inscrutable) as well as for how we evaluate lives. As an example of the former, consider Velleman's borrowing of Aristotle's tale in which the killer of Mitys is subsequently crushed to death by a statue of his victim. This sequence of events forms a narrative in a way that a random pairing of events would not. And, Velleman argues, what makes this into a narrative is not a causal connection any more than it is a simple chronological listing of events, because the story would work even if we do not imagine something like the hand of Fate as an invisible force pushing the statue. Rather, the events form a story because they complete an emotional cadence in the audience. This works in life as in literature.

Thus, for example, a life in which a person, after years of failure, learns from her mistakes and achieves greatness will seem like a better life than one in which she first achieves greatness and then fails, or one in which

she succeeds not as a result of struggling through her failures, but through a random stroke of luck. And the reason the first sort of life is a better one than the last two is that it makes for a better story. This narrative conception of the value of a life, on Velleman's view, explains a number of our intuitions, which would otherwise seem puzzling: why, for example, do we think learning from one's mistakes is better than simply ignoring them and learning something else instead, even something that could be more useful in attaining future benefits? And if a person actually benefits from learning from her mistakes, that life will be better than one where she benefits after those mistakes but in a way unrelated to them because "an edifying misfortune is not just offset but redeemed, by being given a meaningful place in one's progress through life" (Velleman 2000a, 65). The initial application of this view is to the debate about well-being, especially in opposition to any view on which the quality of a life is a function of the sum of momentary well-being in that life. But the narrative view also has a role to play in explaining the value of free will and moral responsibility.

On Fischer's account narrativity helps make sense of the value of acting freely for the same reason it can explain the asymmetry between lives that involve improvement and those that involve decline: because narrative can retroactively affect the meaning of events. Because lives have narrative value, their meaning, like the meaning of a narrative, can "loop"; that is, it is open to revision on the basis of later events. Just as the meaning of the round table is shaped by Arthur's ultimate failure (but also by his subsequent legend!), so the repeated failures of our imaginary subject's early life are changed in meaning by her eventual success. Following Velleman's suggestion that events can have "retrospective significance" due to the fact that "events in a person's life can borrow significance from both preceding and succeeding events" (Velleman 2000a, 68), Fischer argues that "it is a distinctive feature of narratives that later events can alter the 'meaning' or

'significance' of earlier events. In this sense narratives can have 'loops,' as Derrida… contended. It is not that we can change the physico-causal past; but we can sometimes change its meaning and thus its contribution to the value of our lives overall" (Fischer 2009a, 147). It is worth noting that "meaning" and "significance" are introduced in scare quotes, in part due to the difficulty of specifying just what it is that narrative allows us to retroactively change. But the key point, for Fischer, is that the narrative structure of our lives makes it possible for us to retroactively improve our lives not simply by adding more good to them, but by changing the value of past misfortunes by making something good come of them, "thus vindicating our pasts" (Fischer 2009a, 152).

We could not shape the narratives of our lives—or at least could not have a role in giving them the narrative value they have—if either we lacked the capacity to engage in reasoned decision-making, or we were entirely subject to manipulation by external agents (in which case our lives might still have narrative value, but that value would be attributable to our manipulators rather than ourselves as their narrators). Thus, Fischer holds that the ability to shape the narrative value of our lives gives free will its value for us. If so, then free will does not lose its value—or so Fischer thinks—if it consists only of what he calls *guidance* control, rather than *regulative* control. Without focusing on the details of Fischer's account, regulative control is the sort of control often defended by free will libertarians, that is, a kind of control we have only if alternative possibilities are genuinely open to us, which Fischer sometimes describes as the ability "to make a relevant difference to the world" (Fischer 2006, 113).

Fischer denies that we have, or need, such control over our actions. What we do have—and need—is rather guidance control, the sort of control recognized by most compatibilists, requiring that we can *guide* our actions through the "actual sequence" that leads up to them, but not that we be able

to choose among open possibilities. On Frankfurt's well-known version of such a view, for example, such control requires that we identify with the volitions that lead to our actions (Frankfurt 1982); on Fischer's own account, such control requires that our actions proceed from our own moderately reasons-responsive mechanisms (Fischer and Ravizza 1998).[1] Roughly speaking, to guide one's actions is to act in a way that is responsive to reasons in a way that appropriately reflects the self so that, for example, some forms of coercion or hypnosis would be ruled out. Since guidance only requires that we be able to shape our lives, or at least constrain the possible narratives that can be told of them (Fischer 2009b, 172), guidance control is sufficient to satisfy the value requirement on free will; that is, it gives us the kind of "free will worth wanting," in Dennett's famous phrase. Another way of making the point is to suggest that if the value of freedom is self-expression, then this value can be satisfied without alternative possibilities, since what we need to express ourselves is an actual sequence of events that appropriately allows our actions to stem from us; whether a different sequence or conclusion was possible is irrelevant to the question of whether the action involves self-expression. It is by acting freely—in the sense of exercising guidance control—that we give our lives value along the narrative dimension.

Fischer draws a great deal more out of the idea that the value of free action is a sort of narrative value, or the value of self expression (Fischer 2009c). But like Velleman, he is careful to avoid allowing narrative loops to do too much work. They allow us to change our past, to be sure, but in only a limited and largely uncontroversial way. As noted above, Fischer does not think it is possible to change the physical past by means of narrative, but

[1] I will not dwell on the account presented in *Responsibility and Control* here, because Fischer repeatedly stresses that his views on the relation between guidance control and narrative do not depend either on the specifics of his own view or, even, on which "actual-sequence" view one holds (Fischer 2009b, 165–166).

"whereas we cannot go backward in physico-causal space-time and 'change the past,' we can readily go backward in narrative space-time. Whereas it is a constraint on our freedom that the physico-causal past be fixed, and that our actions be extensions of the given (physico-causal) past, it is precisely our capacity to act freely that provides the ingredient that allows for backward travel in narrative space-time" (Fischer 2009a, 152). The capacity to act freely is a capacity to change the past, but in a way limited by the constraint that the physical past is unalterable; we can change only the "meaning" or "significance" of that past.[2] The precise extent of our freedom, then, will depend on the scope we give to these terms. And while Fischer relies on these terms (along with some metaphorical figures of speech) to make his point while leaving his views on their scope fairly sparse, he does seem to have some stringent limits on that scope in mind.

Since what distinguishes narrative, on Velleman's and Fischer's views, is its *affective* component, what free will allows us to do is change the way we *feel* about events or actions in our past, *via* our actions in the present or future. That is, what we do can alter the emotional meaning, but only the emotional meaning, of past events. Now, this is certainly no small feat, since it suggests that virtually any event, no matter how tragic, stupid, or horrific, can still be redeemed through the telling of a sufficiently creative narrative. The worst mistakes can be salvaged, if not transmuted into fortuitous windfalls, by a life in which their correction leads to greatness. Some actions, of course, may seem to be irredeemably bad—it may be hard to imagine, for example, how a Nazi's wholehearted activities in concentration camps could be redeemed by even the most uplifting story—but whether actions can be redeemed, and to what extent, will itself be determined by the narrative

[2] A point worth considering in this regard is that, if Fischer is right, then "meaning" cannot merely supervene on physical reality, since meaning can be altered retroactively, while physical reality cannot.

conventions within which the story is told. Again, though, however unlimited this power to change the past may seem in principle, for Fischer and Velleman it is so only within a very specific domain, that of affective meaning.

2 Narrative as Incidental to Practical Rationality

For Fischer, narrative and guidance work along separate tracks, as it were. We govern ourselves through our actual sequence mechanisms. Thus, if we accept Fischer's own view for purposes of illustration, we act autonomously or responsibly when our actions are governed by our own reasons-responsive mechanisms. Narrative comes in at a different level: it is not involved in our self-governance, but in filling in the account of why that self-governance matters to us (this is, in part, why Fischer can hold that his account of the role of narrative is compatible with any actual sequence view). In contrast, Velleman holds that narrative plays a central role within self-governance itself. On his overall view, self-understanding is a constitutive aim of action, and this aim explains the exercise of practical rationality (Velleman 2000b; Velleman 1992). Briefly, the idea is this: in acting, we seek (not usually explicitly) to understand our actions. And we can succeed in doing so only if our actions make sense in context, that is, if they are explained by our motives. The aim of self-understanding works in the background of our practical reasoning to steer it in the direction of acting on motives that it makes the most sense to us to act on. Practical reasoning as such, then, is subordinate to self-understanding insofar as reasoning how to act is essentially a matter of seeing what it makes the most sense to do.

Narrative, as a mode of understanding, plays directly into this aim of our agency: our self-narration allows us to reason about how to act. In some work, Velleman gives narrative a central place in constituting our autonomous or self-governing identity. Arguing against Dennett's claim that

the self is a fictional construct created through narratives (Dennett 1992), Velleman responds that while we do construct a narrative self, this self is not fictional; it is, rather, constitutive of "agential unity, in virtue of which a person is self-governed or autonomous" (Velleman 2006, 223). In deciding how to act, an agent examines his circumstances and motives, and his self-understanding weighs the options before him by "providing a potential *rationale* for his next action—that is, an account that would make the action intelligible, a coherent development in his story" (Velleman 2006, 219). Velleman takes the relevant notion of coherence here to be the one outlined above.

Fischer and Velleman may thus seem to accord very different roles to narrative within human agency. For Fischer, narrative adds a dimension of value to agential mechanisms that otherwise operate on their own. For Velleman, narrative not only provides that dimension of value, but also the underlying coherence that allows agents to decide which motives to act on in the first place. However, the difference turns out to be minor. On both views narrative plays the role of guiding agents to act in ways that are *meaningful*, lending significance to the past. And despite the central place Velleman accords to narrative in his response to Dennett, elsewhere he dials it back by allowing narrative to vie with other modes of practical rationality for control of our agency. Recall that in his account of narrative coherence Velleman seeks to distinguish narrative explanations from causal ones; this distinction also plays a role within our agency, since "it implies that practical reasoning is fragmented into the pursuit of two asymmetrically dependent modes of self-understanding. I think that we aim to make sense of ourselves not just in the mode of causal explanation but also in the mode of storytelling. We consequently aim to do things for which we have both an explanation, revealing why we came to do them, and a narrative that helps to clarify how we feel about them or what they mean to us" (Velleman 2009, 201). So, on

one hand, we seek to understand ourselves in terms of standard practical rationality, seeing which action would best satisfy our existing motives in the circumstances; on the other, we seek to take meaningful actions, ones that will make sense in emotional terms. But, since Velleman also argues that we cannot even make sense emotionally of behavior that we cannot explain causally, it follows that "practical reasoning aimed at narrative self-understanding can supplement but not replace reasoning aimed at self-understanding in causal-psychological terms. And I suspect that this supplementary mode of practical reasoning is optional, at least to some extent" (Velleman 2009, 203).

On the view that emerges in Velleman, then, narrative understanding plays a role in practical rationality, but only a supplementary and (somewhat) disposable one. Fischer's account can borrow from this model with almost no alteration. In what follows, however, I will challenge two key features of this consensus. First, I will argue that narrative plays a much stronger role in practical rationality than the supplementary one of adding emotional meaning to our actions, as if practical reasoning could go on just as well (perhaps somewhat impoverished, but not crippled) without narrative understanding. This is because the very materials of practical rationality—centrally, the motives that explain our actions, and the way in which they explain them—are themselves narrative-dependent. Second, I will argue that, if narrative can shape practical rationality at a fundamental level and thus cannot be limited to providing a veneer of value to a pre-existing rational order, its "loops" should be able to change not merely what our past actions mean in emotional terms, but also what the motives explaining those actions were.

3 Narrative as Indispensable to Practical Rationality

The limiting of the power of narrative loops to the changing of *emotional* meaning may seem to be sanctioned by the limiting of narrative's distinctive feature to that of presenting an emotional arc. But narratives can do more than convey emotion. Velleman stresses that narrative understanding is only *one* of the ways we make sense of ourselves; we do so, also, through rational understanding along causal-psychological lines. For example, a man may sacrifice everything to save a family business, an action that makes sense in emotional terms, even though the rational thing to do in terms of what would best satisfy his economic motives would be to consider the business expenses sunk costs. "From the perspective of instrumental rationality, in which we understand our actions in terms of their motivating aims, throwing good money after bad is indeed irrational, since it foreseeably tends to frustrate our aims and therefore cannot be understood as motivated by them. Having already invested attention and effort in an unpromising endeavor lends no intelligibility to the course of investing more" (Velleman 2009, 201). Why, then, do we continue to persevere in what seem like unprofitable ventures? One reason, Velleman suggests, is that narrative provides us with reasons "to seek a future that continues the narrative arc of our past" (Velleman 2009, 201) by allowing us to understand our actions in emotional terms, so that even courses of action that almost certainly will not succeed are not entirely without merit. "We stick with a marriage or a degree program after it has stopped promising to repay our efforts partly because even the story of eventual failure provides the emotional cadence of hopes disappointed, which has a comprehensible meaning" (Velleman 2009, 202). If narrative coherence rests on tracing out an emotional cadence that allows the audience to reach a sense of how it feels about the story as a whole, narrative understanding guides our actions primarily by directing us towards actions that will yield an emotional sense of closure.

What is rational for a person to do in causal-psychological terms, however, is itself embedded within the narrative understanding of his life. Why, for example, do we think that pouring money into a sinking business is irrational? No doubt part of the explanation is that we take it that a central aim of a business is to make money; wasting money on a failing business is thus nonsensical. But imagine a culture—say, one driven by a pure Protestant ethic—in which striving is far more important than any financial reward, which is, after all, only an earthly prize. In such a culture, to sell a failing business rather than do everything in one's power to save it may be seen as an irrational action, one in which one's greed blinds one to the force of right reasons. Few actions could be called "rational" or "irrational" full stop, independently of the background self-understanding of a culture and its members. Velleman could, of course, respond that an adherent of a pure Protestant ethic of this sort would simply have different motives; continuing to plow his business into the ground would make sense to him in causal-psychological terms, and he would thus do it even if narrative understanding did not factor into guiding his agency. This is precisely my point: narrative enters into practical reasoning at two junctures. One of them may well be the one Velleman and Fischer have in mind: narrative allows us to clarify how we feel about what we do (and thus to act in ways we can have coherent feelings about). But narrative also enters at the more fundamental point of fixing what psychological motives enter into causal-psychological explanations. Our motives themselves—what we want, what we aim to accomplish—rest on our narrative self-understanding and on the narratives we adopt from our cultural milieu. Rationality (in causal-psychological terms) and narrativity are thus not distinct parallel modes of understanding; rather, narrative conditions what counts as rational in the first place. This is not to deny that they *may* run parallel to each other in the ways Velleman suggests, but this parallelism is already embedded within a wider narrative framework.

Narratives teach us how to respond to certain motives, and what motives it is appropriate to have in response to environmental cues. They can also, as already noted, shape our behavior by shaping our self-understanding. That self-understanding, moreover, may only be accessible via narrative; our motives and reasons only make sense within a wider narrative practice of giving reasons and explanations and their content is thus shaped by the concrete particulars of this practice (MacIntyre 2007, chap. 15). In light of this, why should we think that narrative provides a mode of understanding that is only a supplement to practical reasoning? On Velleman's model, affect is the irreducible feature of narrative, distinguishing it from explanations of other kinds. If that's the case, it follows that non-affective aspects of our motivation and behavior are understandable in non-narrative terms, and so long as we are unconcerned with this dimension of value we can go about our practical reasoning without any need for narrative. We can understand our behavior even in the absence of emotional understanding. So the view that narrative understanding is inessential to practical rationality rests on the idea that affect is the glue involved in narrative coherence. There are three lines of response.

First, Fischer himself stresses that while free agency—as the narrating of a story—may be an aesthetic activity, that does not mean that the life it shapes must itself be evaluated from an aesthetic, rather than a moral or prudential, perspective; the opposite is frequently appropriate (Fischer 2009b, 169). We may normally judge an aesthetic production such as a novel by aesthetic criteria (is the writing good? are the characters believable?), but we can also judge a novel by its social impact, as in the case of *Uncle Tom's Cabin*, or the moral message it conveys (such judgments are one reason for the long history of banned books worldwide), or even its prudential qualities (is it wise to publish this in such a volatile political climate?). Conversely, Fischer sees free action as a primarily aesthetic activity that it is rarely

appropriate to judge by aesthetic criteria (how dramatically you announced your resignation, my dear!). If so, then narrative loops should affect not merely our *emotional* evaluation of past actions, but also their moral or prudential or rational meaning.[3]

Second, even if tracing an emotional arc *were* the defining feature of narrative, this would not preclude narrative's functioning to shape, as well as emotionally color, our diachronically extended agency. Emotional responses to a narrative may reveal connections between two events—a motive, for example, and an action—that would be impossible to notice or make sense of without that affective illumination and thus could not be made in a non-narrative way. Finally, if even an account of narrative in emotional terms allows us to note connections that are not themselves purely emotional in nature, we may note that other features of narrative—features that have nothing to do with affect—may make connections between various events intelligible. In fact, a number of authors have rejected Velleman's view that affect is central to narrative coherence, either replacing or supplementing it with other features.[4] One such feature, for example, is that of thematic unity,

[3] Peter Goldie (2012, chap. 1) argues that while narratives typically involve causal, emotional, and evaluative elements, they need not do so, and they can disclose a different kind of meaning: they can show, for example, how the events they describe hang together in the view of a character or the narrator. One such sense of meaning is conveyed by Thomas Scanlon, who takes "the meaning of an action for a person [to be] the significance that person has reason to assign to it, given the reasons for which it was performed and the person's relation to the agent" (2008, 54). In his example, there is a difference in meaning between a phone call to an ill relative that one makes out of genuine concern and one made out of a desire to thereby cultivate the approval of a wealthy grandfather. But even if learning the motives in such cases is likely to change how we feel about the agent in question, how we see him changes independently of how we feel.

[4] Velleman's affective view of narrative coherence is just one of many attempts to define essential features of narrative. For views that reject Velleman's proposal, see Noël Carroll (2007) and Gregory Currie (2010, chap. 2), who emphasizes that narrative requires causality in a fairly loose sense—including all sorts of dependence relations that would not pass muster with philosophical accounts of causality—and adds thematic unity as an important feature of narratives. Mark Bevir (2000) also emphasizes thematic unity, focusing not on the overall unity of the narrative but on the way narrative connections are formed between individual elements that share loose "themes." For

which allows the formation of a narrative from two events that lack causal relations or even emotional resonance. Tweaking any of these aspects of narrative coherence, then, may change the way the elements of the narrative are connected to each other; such tweaks may, furthermore, change connections that are unrelated to, or that subvene, emotional resonance. None of this is meant to deny that practical reasoning has its own, narrative-independent norms. But if those norms are to function practically—to guide us from thought to thought—they must have a content on which to operate. Narrative (via aspects such as affect or thematic unity) shapes the connections between situation, motive, and action that provide the stuff of practical reasoning. And if free action can, via narrative, allow us to "go backward in narrative space-time," it may allow us to change facets of meaning independent of affect.

This is one piece of my account: a change in narrative can reconfigure the relations between the elements narrated. If those elements are motives and actions, the change in narrative can reconfigure how we are to understand the springs of our past actions. But this is not by itself sufficient to make my second point: that this change serves to reconfigure not simply how we *evaluate* those past actions, but what motivated them in the first place. For all I have said so far, even if narrative does change how we *see* our past motives, it may make no difference to what those motives were. If narrative is indispensable to our practical reasoning, then a change in narrative can lead to changes in what sorts of actions will appear rational or even intelligible in light of an agent's given motives and circumstances. But this does not yet show that the motive-action connection can be changed retroactively. Even if the argument so far succeeds, it might still be possible that changing an agent's cultural or personal narrative can only (1)

---

a very schematic overview of some leading contenders in the quest to uncover central elements of narrative coherence, see Paisley Livingston (2013, 341–342).

reconfigure his future actions prospectively, by altering what action it makes sense for him to take given his motives on future occasions, and (2) alter his evaluation of his past actions retrospectively, by altering how he judges the rationality and justifiability of his past actions, which are no longer sanctioned by the new narrative in terms of which he understands himself. Without denying that both (1) and (2) are possible, I want to suggest a third possibility: that a change in narrative can retroactively reconfigure the connection between past actions and motives.

4 Narrative and Retroactive Constitution

The reasons we give for our actions can serve to explain, but also to rationalize them. They do so not merely by explaining what motives brought the action about, but also by showing that under the circumstances it is reasonable that those motives (or associated beliefs) bring about the action they do. The reasonability of a motive, however, will depend on a wider narrative framework—not just any motive-action connection will make sense, but only some. So, for example, depending on the cultural narrative that shapes explanations and contexts of anger, getting angry at an insult may be reasonable, getting angry at an accidental shove may be merely foolish, but getting angry at a gift that one both desires and appreciates will be unintelligible.[5] But one may come to accept a narrative framework in which unintentional slights and favors are to be regarded with anger,

[5] More accurately, the point is that such anger in response to the gift will have to be explained through psychological, rather than rational, means. Contrast this with the explanation of getting angry at an accidental shove: it may be misguided to get angry in such a circumstance (unless, of course, the accident was itself due to carelessness on the part of the person doing the shoving), but we can still understand the anger as a rational response. I am leaving the distinctions here somewhat vague because I think they really are vague: we might better say that anger at an insult is appropriate in most senses, anger at the shove is inappropriate in some senses but appropriate in others, and anger at the gift is inappropriate in most senses.

whereas insults instead call for deep contemplation and self-abasement. The new narrative can thus rationalize the second and third cases of anger, but will make the first one foolish or even unintelligible. If I come to identify with the new narrative, reflecting on my past gratitude for gifts may elicit feelings of shame; I will see my earlier feelings and actions as irrational and perhaps as indicative of my immaturity. Of course I might think something like the following: I had a different way of seeing the world back then, and while I now think that way of seeing the world was mistaken, it did rationalize my earlier behavior. But on the other hand, I might simply feel shame or confusion at my (to me now) irrational earlier actions and view my earlier way of seeing the world as mistaken precisely because it could justify such ways of behaving. The change in narrative here does not simply change how the earlier action *feels* to me; it changes whether the action makes sense as a response to the circumstances in the first place.

Here I am stressing two ways in which narrative can make sense of the relation between past actions and motives. First, it determines whether the motives justify the consequent actions. But it also shows whether the motive renders the action intelligible. In radical enough shifts of narrative, we may see a past action not merely as unjustified, but as unintelligible in light of the motive we took ourselves to have. In attempting to make sense of why we acted in this way, then, we may have to reinterpret the motive or risk failing to understand the action at all. Another example: I steal pears for fun. Years later, I undergo a spiritual conversion. I come to see my seeking fun not merely as that, but as a turning away from the eternal to the temporal, as evil in the purest sense. My motive was not fun at all, or not really fun as such, but a flight from the goodness of God.[6] I reinterpret the motive I took myself to have at the time. And my narrative self-understanding loops back to reconfigure the cause of my action.

[6] The example, of course, is adapted from Augustine's *Confessions* (2002).

These examples invite the objection that they need not involve anything like a narrative reconfiguration of the original motive. The case of anger may be interpreted as follows: in acquiring a new narrative, I simply acquire a new understanding of what actions are appropriate in response to particular motives. Using this new understanding to evaluate my past actions, I find them shameful, irrational, or incomprehensible. In the second example, what we see are just cases of self-deception, or at least mistaken interpretation. The objector can claim that stealing the pears was just fun; my later understanding simply overdramatizes that fun, attributing a deeper meaning to it that my youthful indiscretions never possessed. Alternatively, one can take the opposite line, arguing that my stealing the pears really was a flight from God, which I foolishly disguised from myself by pretending it was merely fun. There are ways of explaining each such example without assuming that narrative can retroactively change the motives for our past actions. This objection thus maintains the view that motives, like other things in the past—explosions, conquests, amusing haircuts—cannot be changed. At the most, it seems, a new narrative, or a change in an existing narrative, can show that we were mistaken about what we initially took our motives to be, revealing either motives previously overlooked or misinterpreted; in some cases, the most such a change can do is alter our evaluation of motives already identified.

The objection, however, assumes that there is a way of fixing the "real" motive for my actions apart from some narrative in which motives serve to make sense of actions, and this is precisely what we don't have. Narratives shape what those motives were. We often speak as if mental states such as feelings, wantings, intentions, or hopes are metaphysically determinate entities—like bell-bottoms or furry animals. But the mind may be far more ambiguous than that. We are filled with stirrings, urgings, and sensations, but it is hard to believe that outfitting them with identities—with

names, functions, conditions, appropriate contexts for having them or ways of expressing them—is a process completely independent of narrative forms of explaining, individuating, and combining the various contents of our minds, especially given that we learn what attitudes are appropriate to what situations at least partially *through* cultural narratives. If the identity of our mental states is not entirely fixed independently of narratives, we lose the grounds for insisting that there must be one determinate answer about what I felt, wanted, or believed at any particular time, and that all narratives can change is how we evaluate, or perhaps how we perceive, something definitive that was already there.

Consider two narratives of the same sequence of events, one in which an agent is motivated by malice, another in which he is motivated by revenge. Either—or both—may be perfectly reasonable interpretations of the facts; that is, it is possible that what makes the narratives plausible is not, or not only, the physical evidence, but a host of considerations relating to the conventions governing narratives, including the sense that *subsequent* parts of the narrative make of the sequence in question. If our understanding of a motive within a narrative can change as the narrative changes, and the motives for our actions are fixed by means of narratives, then changes in narratives can reconfigure our motives. A conversion to Christianity may change the meaning of my pear-stealing by changing what motivated me to steal them; a commitment to a cause may change one's early dabbling in it into an earnest attempt to learn more.

My suggestion, again, is that we simply lack a narrative-independent way of identifying the motives for our actions. So if I act in the belief that I am motivated by motive M_1, a change in my narrative may make me reconsider that identification of the motive, instead convincing me that I acted on motive M_2. This seems largely uncontroversial. A conventional explanation would have it either that the belief that I acted on M_1 was mistaken insofar as it

mistook my motive, or that initially the motive was epistemically indeterminate, and the further details brought out by ongoing investigation, or further connections illuminated by a change in narrative, helped to either correctly pin down the motive or to make it more determinate. My proposal here amounts to the claim that there is another possibility: our motives are metaphysically indeterminate; what gives them determinacy are the narratives we use to pick them out. How would one set out to *objectively* identify what motivated an action? It will not do to simply ask the agent, since agents can change their minds. Perhaps we might examine an MRI of the agent's brain at the time she decided to act. But this could not yield a determinate answer since *motives*, unlike patterns of neural activity, do not appear on brain scans. They are fleshed out by reference to the agent's other beliefs, the way the agent's community identifies mental states, and the extent to which they rationalize the action; a motive that utterly fails to rationalize an action, that does not succeed in making it intelligible at all, isn't a motive; and it is for this reason that if an agent's narrative cannot make her action intelligible in terms of a specific motive, searching for a different motive is a reasonable option. As Thomas Uebel (2012) suggests, narrative explanations of action can help identify singular causal claims about which of an agent's motives brought the action about. But if there *is* no way to identify the causes of action without recourse to narrative, then the details fleshed out by the narrative explanation do not merely make a pre-existing motive clear. Rather, they play a role in constituting the motive. And this means that a different narrative can, in principle, yield a different account of the motive without a change in the factual details of the case.

So my claim is not simply that agents do not always know what their past motives were and can better understand them through acquiring a more detailed narrative or recalling more data. It is rather that, to some extent, the *correct* account of a past motive depends on the narrative explanation of the

action used to flesh it out. A final consideration in favor of this view might note that narratives—bound as they are by considerations of thematic unity—come with conditions of relevance built in. A narrative about Augustine's conversion is likely to leave out details of his shaving practices, unless those details are relevant to his spiritual transformation. Similarly, a narrative explanation of an agent's action will include a subset of the available data about the circumstances, the agent's beliefs and desires, and so on. The narrative in part determines what it is about by determining what is relevant to that theme and excluding what is not. Again, this will mean that keeping all unchangeable facts about the physico-causal past constant, different accounts of the causal sequence that led up to an action in terms of mental states will be possible. Nor is it clear that *some* explanation must be *the* correct one.[7]

5 Narrative and Retroactive Self-Constitution

I have been speaking of changes in cultural narrative rather than speaking, as Fischer does, of changes in our past that occur as a result of our individual developing narratives. But if the wider cultural narrative can make a difference to the identity of a past motive, this should work on a smaller scale as well. First, it is clear that agents can alter the narrative in terms of which they interpret and identify their past motives by undertaking certain actions. One may join a cult or move to a different country and make every effort to

[7] This does not mean, however, that *any* interpretation can be correct. I've argued that the identity of motives is partially, but only partially, constituted by narratives. Thus, motives can only be correctly identified by narratives that are consistent with the existing facts. I am not here taking a position on just how much of a motive's identity should be left to narrative and how much to fixed facts, but we can suggest the following. If the (otherwise) most coherent explanation of an action does not require reference to emotion, neural activity in the limbic system need not concern us. Similarly, if a narrative calls for an emotional explanation in the absence of activity in the limbic system, we should seek a better narrative.

assimilate. And one may do so even with some awareness of the changes this will bring about in one's self-interpretation. Augustine did not, after all, simply find himself a Christian one fine morning; he undertook actions that made his conversion possible and, ultimately, likely. But it is not necessary to drastically alter one's narrative in order to alter one's past motives; the same effect can be accomplished by means of advancing one's existing narrative.

Consider, for example, Ivan, who defends his honor at every opportunity by challenging anyone he perceives as having slighted him to a duel. Ivan was brought up to believe that a man has nothing without his honor, and that disposing of anyone who offends it in armed combat is the proper way of preserving it. Moreover, as an honorable man, Ivan objects to viciousness and brutality in any form. But one day Ivan is struck by the crying widow of his latest adversary, and inquires into her well-being. Pursuing this inquiry, he comes to see himself through her eyes—as a vicious and brutal monster. He actively sets out to learn more about his culture of honor, and he begins to suspect that the honor code he has embraced may be merely a cover for the brutality of the men who espouse it. He comes to see his own quest for honor as an excuse to kill without guilt, and he now turns against the stirrings that drive him to anger at every perceived slight—he no longer sees them as consisting of a desire to preserve his honor, but merely as brute tendencies toward harm. Here Ivan's past motives have been reconfigured in light of his new narrative self-understanding; and it is because they have been reconfigured that he can now struggle against what he sees as his lifelong character.

Or consider Nina, whose upper class upbringing leaves her with a strong dislike of the poor and a sense that they deserve little compassion for bringing their poverty upon themselves. When she hears pleas for social welfare on television, or finds herself accosted in the street by someone asking for change, she is revolted. But she is also fascinated: why, she

wonders, would people go through such effort to make themselves dependent on others? To satisfy this curiosity, she occasionally reads about people in poverty; sometimes she might engage some poor unfortunate in conversation. Her curiosity, she believes, is driven by a desire to understand corrupt human nature; lacking any charitable motive, she can only express a puzzled disdain at every revelation. But in time she decides to continue her experiments: she finds individuals to whom she offers assistance, checking up on them and ensuring (through intermediaries) that it is not wasted. Her motive, again, is experimental: she wants only to know if people accustomed to ruining their lives are capable of improving it as well. With more revelations, and with several success stories, she becomes genuinely interested in the lives of her test subjects; she reaches out to them with more assistance, and she finds other candidates. She engages wholeheartedly in charity work, seeing her earlier dispassionate curiosity as a longing to help others. Having transformed her scientific motives into charitable ones, she engages in a life of altruism convinced that her character has undergone no change; she has always been concerned with helping others, but only recently has taken an active role in this direction.

In both cases, agents change the role past motives play in their actions—and thus reconfigure the motives themselves—by changing the ongoing narrative of their lives. In speaking loosely of "changing" past motives, however, I do not mean anything spooky—like reaching back in time and rewiring one's own neural synapses. I mean only that by changing the narrative of our lives, we can change the meaning of items within that narrative. And if the identity of a motive depends in part on the narrative explanation given of the action, then it follows that some of the items constituted by a narrative may appear differently within a different narrative. No changes are made to the neural synapses; but which of those synapses—and the circumstances in which the agent finds herself—are

relevant to making the action intelligible (to making it appear as an action, for that matter) will depend on which of them are picked out by the narrative used.

6 Temporal Loops and Free Will

If narrative loops can change our motives *post facto* in this way, then these loops give us a freedom that appears to be stronger than guidance control. What is thought to rule out regulative control is the causal link between motive and action, and if loops can allow actions to alter the motives for which they are taken (or, if future actions can alter the motives for past actions), then narrative loops can restore something akin to regulative control. As I mentioned above, Fischer limns regulative control as the ability to make a relevant difference to the world.[8] On this view, at time *t* it is possible for me to perform either action A or action B, such that there are at least two possible worlds that are exactly identical up to time *t*, and yet in one of them I do A and in the other I do B at *t*.[9] The "relevant difference" I make to the world, on this conception, is one of making the world different in its physico-causal constitution from what it could have been, given the state of the world prior to time *t*. Fischer denies that we can do this, and for the sake of the present argument I agree with him.[10] This is why I said that narrative loops can restore something *akin* to regulative control.

[8] This is the kind of free will libertarians like Ginet (1990) and Kane (1996) have in mind.

[9] This gloss will work for my purposes here, but it is an oversimplification insofar as it assumes that actions occur at definable times. On a more complex account, on which actions are processes, themselves linked to other processes either by being embedded in them or being rationally or physically linked to them, the picture becomes significantly more complicated.

[10] I am ignoring important nuances here, since Fischer's brand of semi-compatibilism can be taken up by either determinists or indeterminists. The latter may grant that it is possible for there to be two possible worlds in which I perform different actions, A and B, but deny that it is within *my*

Compatibilist and semi-compatibilist lack of regulative control seems to rule out free will in three ways: (1) Because the agent's action stems entirely from the actual sequence, the agent seems to lack any ability to originate the action. What originates the action is the sequence, and the sequences that preceded it, not the agent. (2) Faced with two possible future actions, A and B, the agent can actually only choose one of them, the one that follows from the actual sequence. (3) Since the physical events of the actual sequence follow in accordance with the laws of nature from previous states of the universe about which the agent has no choice, the agent has no choice about which sequence she acts on.[11] On this view, we may seem to lack free will altogether, since each time we are faced with an apparent choice, the outcome is neither something we originate nor something we have a genuine choice about given our motivational make-up. Considerations of this sort drive many free will theorists toward libertarianism and regulative control. The modification I've made to Fischer and Velleman's view of narrative can resolve the apparent difficulties with (1) and (2), although it cannot respond to (3) without further metaphysics.[12] I want to suggest that the view I've outlined, while insufficient to allay all libertarian concerns, may go some way toward meeting them.

A short detour into action theory can help clarify my point. On virtually *every* view of action, an action is not merely a physical bodily movement; it is a physico-psychical movement. On the causal view, what distinguishes actions from other events is their psychological etiology: actions are movements that are appropriately caused by a combination of

---

power to decide which of these worlds will come about. The point, then, is not that I cannot do something that will make a difference to the world, but that it is not up to me whether I do so.

[11] This is meant only to be a quick sketch of how guidance control without regulative control might be thought to exclude free will. It is not my goal here to defend any of these claims.

[12] Whether (1) and (2) can ultimately be resolved without resolving (3) is a separate concern I will not address here.

beliefs and desires and, perhaps, intentions (Davidson 1980). More recent versions adopt Searle's (1983; 2001) modification, on which actions are not simply the subsequent effects of an intention, but are caused by an intention-in-action that accompanies them. Another recent development, harking back to Anscombe (2000), attempts to understand action as incorporating intention as an intrinsic feature, such that the unfolding of the action just is the unfolding of the intention (Thompson 2008; McDowell 2011). And some hold that actions are embodied reasons, containing an act (that is, the bodily movement) for the sake of an end (Korsgaard 2008). On each of these views actions are not simply bodily movements; what makes them actions in the first place is a psychic element. But the psychic elements, as I have been arguing, depend for their identity, in part, on the narrative explanation of the action in question. As such they are metaphysically indeterminate and can potentially take on different identities through different narratives. So while narrative trips back in space-time cannot change the physico-causal past, which includes the physical movements our bodies perform, they can reconfigure the accompanying components that make the movement an action, or an action of a certain kind (a selfless action, for example, or a jealous one).

On Fischer's view, the actual sequence leading up to any given past action will be unaffected by changes in narrative. But on the view I am proposing, the actual sequence leading up to a past action remains constant only insofar as it consists of physico-causal events. Insofar as it contains mental events, they are open to reidentification by changes in narrative. We cannot make a difference to the physico-causal constitution of the past. But we can make a difference to our interpretation of that constitution, and so to the way in which we pick out psychological states like motives (and the way in which we pick out *which* of those motives were causally active in the production of our actions) from the sequence of physical events in the past.

Even if we do not, in the present, have a power to choose between alternatives because our actions follow from our existing motivational make-up, we have a power to undertake courses of action that lead to changes in the narratives by which we make sense of our pasts, and in doing so we can reconfigure what motives our pasts contained. And since our present is the upshot of the past, we thereby have a power to make a difference to the world in the present and future.

On the view I am proposing, we do not have *direct* control over the motives for our past actions, because we do not typically have direct volitional control over the narratives by means of which we understand ourselves; as I've mentioned, at best we can undertake courses of action that may lead to changes in our narratives. This distinguishes these narratives, which provide the context for self-understanding, from the sorts of literary and conversational narratives we ordinarily compose by exercising our free will. Although I do not want to rule out the possibility that agents may be able to undertake the task of *intentionally* changing their past motives, I doubt that they can do so directly. If they were to try, they may well run into the problem of self-deception. It is one thing to change one's life and self-understanding in such a way that one acquires a different view of a past motive. But it is hard to see how an agent could intentionally set out to rewrite their narrative so that their past malice appears as charity without thereby simultaneously affirming that one's past motive was not entirely benevolent.

But we do gain a power to change the motives behind our actions, to undergo a retroactive self-constitution, which additionally provides us a way of opening up our future, since our past plays a central role in determining the sort of person we are, the sorts of motives we are subject to, and the way we respond to those motives in action. If our practical identity is constituted by our practical commitments, which are in turn open to retroactive

modification, narrative gives us a power approaching self-creation; our actions do not simply originate in the past, since the past to some extent originates in the consequences of our present and future agency. This is not a full-blown libertarian kind of freedom; it does not give us alternative possibilities in the sense of ensuring that we could choose to act differently at the moment of decision, and it does not give us any power to make a "relevant difference" to the physico-causal world. It does not, that is, give us the power to freely choose between doing A and B at the moment the choice is made. But it does provide us with alternative actions insofar as it provides us, in a sense, with alternative pasts through a back-door, allowing us to choose between doing A and B in the future by retroactively constituting ourselves as the sorts of people—with the motives and reasons—who *would* do A or B. But rather than focusing on introducing indeterministic breaks into our choices in the present, as libertarians often do, it allows our actions to retain continuity with our past selves by allowing us some control over what our past selves have done and why, and thus over what our future selves will do. Expanding the role of narrative thus allows us to go beyond standard accounts of compatibilism and even semi-compatibilism, and towards something more akin to a semi-libertarianism.

Acknowledgements I would like to thank Luca Ferrero, Robb Eason, and audience members at the 2013 meeting of the Pacific Division of the American Philosophical Association (including my commentator, Tina Talsma) for helpful discussion of earlier versions of this paper. I am also grateful to Maureen Sie and several anonymous referees for their thorough feedback.

References

Anscombe, Elizabeth. 2000. *Intention*. Cambridge, MA: Harvard University Press.

Augustine. 2002. *The Confessions of St. Augustine*. Mineola, NY: Dover Publications.

Bevir, Mark. 2000. "Historical Explanation, Folk Psychology, and Narrative." *Philosophical Explorations* 3 (2): 152–68.

Carroll, Noël. 2007. "Narrative Closure." *Philosophical Studies* 135 (1): 1–15.

Currie, Gregory. 2010. *Narratives and Narrators: A Philosophy of Stories*. Oxford University Press.

Davidson, Donald. 1980. "Actions, Reasons, and Causes." In *Essays on Actions and Events*. New York: Oxford University Press.

Dennett, Daniel C. 1992. "The Self as a Center of Narrative Gravity." In *Self and Consciousness: Multiple Perspectives*, edited by Frank S. Kessel, Pamela M. Cole, and Dale L. Johnson. Hillsdale, NJ: Lawrence Erlbaum.

Fischer, John Martin. 2006. "Responsibility and Self-Expression." In *My Way: Essays on Moral Responsibility*, 106–29. New York: Oxford University Press.

———. 2009a. "Free Will, Death and Immortality: The Role of Narrative." In *Our Stories: Essays on Life, Death, and Free Will*. New York: Oxford University Press.

———. 2009b. "Stories and the Meaning of Life." In *Our Stories: Essays on Life, Death, and Free Will*. New York: Oxford University Press.

———. 2009c. "Introduction: Meaning in Life and Death." In *Our Stories: Essays on Life, Death, and Free Will*, 3–25. New York: Oxford University Press.

Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control*. New York: Cambridge University Press.

Frankfurt, Harry. 1982. "Freedom of the Will and the Concept of A Person." In *Free Will*. New York: Oxford University Press.

Ginet, Carl. 1990. *On Action*. New York: Cambridge University Press.

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

Kane, Robert. 1996. *The Significance of Free Will*. New York: Oxford University Press.

Korsgaard, Christine. 2008. *The Constitution of Agency*. New York: Oxford University Press.

Livingston, Paisley. 2013. "Narrative." In *The Routledge Companion to Aesthetics Third Edition*, edited by Berys Gaut and Dominic Lopes, 340–50. London: Routledge.

MacIntyre, Alasdair C. 2007. *After Virtue: a Study in Moral Theory*. Notre Dame, Ind.: University of Notre Dame Press.

McDowell, John. 2011. “Some Remarks on Intention in Action.” *The Amherst Lecture in Philosophy* 6: 1–18.
Scanlon, Thomas. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.
Searle, John. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press.
———. 2001. *Rationality in Action*. Cambridge, MA: MIT Press.
Thompson, Michael. 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*. Harvard University Press.
Uebel, Thomas. 2012. “Narratives and Action Explanation.” *Philosophy of the Social Sciences* 42 (1): 31–67.
Velleman, J. David. 1992. “What Happens When Someone Acts?” *Mind* 101: 461–81.
———. 2000a. “Well-Being and Time.” In *The Possibility of Practical Reason*, 56–84. New York: Oxford University Press.
———. 2000b. “Introduction.” In *The Possibility of Practical Reason*, 1–31. New York: Oxford University Press.
———. 2003. “Narrative Explanation.” *The Philosophical Review* 112 (1): 1–25.
———. 2006. “The Self as Narrator.” In *Self to Self*, 203–23. Cambridge: Cambridge University Press.
———. 2009. *How We Get Along*. New York: Cambridge University Press.