

## A REPLY TO PEREBOOM, ZIMMERMAN AND SMITH

### 1. Introduction

When I was Director of the University Honors Program at the University of California, Riverside, I was taught that, when managing staff, one should always give them a ‘feedback sandwich’. That is, it is wise in such circumstances to start and end with something nice, even if delivering some tough news in between the slices of niceness. In philosophical contexts such as this one, it is often the case that those who prepare the sandwiches appear to adhere to the Atkins Diet. But at least as a consumer of such sandwiches, I’m really into the carbs! I am truly grateful to my commentators, Angela Smith, Derk Pereboom, and David Zimmerman, for their thoughtful comments, their generosity, and for not skimping on that wonderful bread.

### 2. Reply to Pereboom

1. Suppose that Brown acts in the presence of an incentive to do otherwise, considers that incentive, but does not act in accordance with it (does not choose and act in the way commended by the incentive). Imagine further that Brown meets the Fischer/Ravizza criteria for guidance control: he acts as a result of his own, moderately reasons-responsive mechanism. If causal determinism obtains, I am willing to grant that Brown could not have acted in accordance with the incentive in question. But Ravizza and I suggest that all that is necessary is that Brown act from a mechanism that is, as it were, ‘keyed to’ the actual incentive in the sense that it (the mechanism) has the power or capacity to act in accordance with the incentive. So if Brown complains that it is unfair to hold him responsible (say) for not acting in accordance with an incentive in accordance with which he *could not* have acted, the response would be as follows. It is not simply that you made a choice and acted while possessing a certain general capacity to respond to reasons. Your choice actually issued from such a mechanism—a mechanism suitably keyed to the actual incentive (that is, a mechanism that had the power to respond differently to that very incentive). What matters here is not the agent’s ‘all-in’ power to choose and do otherwise; what matters is that the agent actually displays a certain general power suitably linked to the actual incentive.

I don’t believe this is an ‘ersatz’ alternative possibility approach, in the relevant sense, although I agree that it invokes the general powers of the mechanism that actually issues in the behaviour under evaluation. The Fischer/Ravizza approach then is ‘in between’ a libertarian view or any view (even a compatibilist view) that accepts the Principle of Alternative Possibilities (PAP) and a view such as that of Jay Wallace, who simply requires that an agent act while possessing a general capacity of the relevant sort.<sup>1</sup>

1. R. Jay Wallace, *Responsibility and the Moral Sentiments* (Harvard University Press, 1994).

I believe that I am in agreement with Pereboom about how to respond to Widerker's 'W-Defence'. I would say something like this. Joe is morally responsible for what he does insofar as he exhibits guidance control of his behaviour. More specifically, he acts from his own, moderately reasons-responsive mechanism. In evaluating his action, we should not ask whether he could have done otherwise, but whether he acted in accordance with suitable principles or standards of conduct. If he exhibits guidance control and meets certain epistemic constraints, then Joe can be blameworthy by virtue of failing to meet some relevant normative standard.<sup>2</sup>

Although I certainly appreciate Pereboom's kind offer to help, I am not sure that I wish to accept it. As Pereboom points out, "Brown's challenge was: 'the mechanism on which I acted did not—and *could not*—have responded to the *actual* incentive to do otherwise'" (p. 201 above). Pereboom says that Ravizza and I should have supplied the following sort of response: 'That is correct. But by virtue of the mechanism on which you acted, you did in actual fact respond to the incentive to do as you did.' I fear however that this does not really address the fundamental worry, which is not about whether one "responded [in some way or another] to the [actual] incentive to do as you did", but about whether one could have acted in accordance with what the incentive commends. I prefer an approach that is not 'ersatz' but is admittedly 'hybrid'. That is to say, I prefer to insist that the behaviour under evaluation be the actual result of a certain sort of mechanism—a mechanism with the general capacity to respond differently to the actual incentive.

2. Pereboom zeroes in on what he takes to be a problem with my reply to Stewart Goetz's critique of the Frankfurt-type cases (or their employment as part of a semi-compatibilist strategy). I am grateful for Pereboom's incisive and penetrating critique. David Goldman has independently convinced me that there are problems with my strategy of reply to Goetz (and thus my defence of my analysis of the Frankfurt-type cases). Here I shall simply suggest that this is an excellent challenge for future work. My preliminary and tentative prediction is that I will not remain entirely silent in seeking to address these (legitimate) worries.

3. One powerful worry about compatibilism with respect to causal determinism and moral responsibility stems from the difficulty of distinguishing in a general, principled way between cases of indisputably responsibility-undermining manipulation and cases of 'mere causal determination' (which is putatively consistent with moral responsibility). Pereboom has presented a challenging 'four-case' argument that highlights this difficulty.

In my work, I have sought to give a general characterization of the distinction between responsibility-undermining factors and mere causal determination. Whereas others have been content to put 'stock cases' on each side of the dividing line, I have tried to present a characterization that explains and systematizes our judgements about the clear cases, and even provides some measure of guidance in 'unclear' or difficult cases. In response to Pereboom's four-case

2. I discuss the W-Defence and develop my response in 'Free Will and Moral Responsibility', *My Way*, pp. 182–216.

argument, I have said that Plum is morally responsible in all four cases (for all that has been said in the description of the cases).<sup>3</sup> I have insisted on the distinction between moral responsibility and such notions as moral praiseworthiness and blameworthiness. (Obviously, other philosophers, such as Harry Frankfurt, have made this sort of distinction.) For example, it seems obvious that someone could be morally responsible for performing a morally neutral action (and thus be neither praiseworthy nor blameworthy for so acting).

Pereboom can accommodate the possibility of moral responsibility for morally neutral behaviour. But he contends that one cannot be morally responsible for immoral or wrong behaviour without ipso facto being blameworthy, in the sense of ‘moral responsibility’ at issue in the Free Will debates. He says:

This is because for an agent to be morally responsible for an action in the sense at issue is for it to belong to him in such a way that he would deserve blame if he understood that it was morally wrong, and he would deserve credit or perhaps praise if he understood that it was morally exemplary, supposing that this desert is basic in the sense that the agent would deserve the blame or credit just because he has performed the action (given understanding of its moral status), and not by virtue of consequentialist considerations. (p. 211f above)

But I do not see why it is clear that it is this notion of moral responsibility, or only this notion, that is ‘at issue’ in the debates about causal determinism, free will, and moral responsibility. It seems to me that a Strawsonian approach to the concept of moral responsibility (in the sense at issue) is at least equally plausible. On such an approach, one is morally responsible insofar as one is an apt candidate for the reactive attitudes on the basis of one’s behaviour (to the extent that one acts freely). I have no objection to understanding this ‘aptness’ in terms of its being an open possibility that the agent deserves (in the way characterized by Pereboom) certain reactions.

In a pre-Quinean moment, I distinguish between the concept of moral responsibility and the conditions in which the concept obtains. The Fischer/Ravizza account of guidance control (in terms of ownership and moderate reasons-responsiveness) is an attempt at giving at least part of the conditions of application of the concept of moral responsibility. On this sort of approach, given that the relevant epistemic constraints are met, one is morally responsible insofar as one exhibits a certain minimal or ‘threshold’ level of a distinctive sort of control (guidance control). Partly in virtue of exhibiting this control, one is morally responsible, and thus an apt candidate for the reactive attitudes. But one could be an apt candidate for the reactive attitudes without its being appropriate or justified, all-things-considered, to apply any such attitude (in the context). I am commending a view on which there are different moments in the judgement of moral responsibility; the first moment involves ‘eligibility’ for the reactive attitudes, and a second moment involves the actual application (or withholding) of such attitudes.

3. ‘Responsibility and Manipulation’, *My Way*, pp. 223–50.

On my approach, moral responsibility flows from a distinctive (but minimal) sort of control. We human beings can possess this sort of control, but most nonhuman animals do not. It is minimal in the sense that even someone subject to significant manipulation (such as Plum in Pereboom's Cases 1 and 2) has it, but possessing it distinguishes an individual such as Plum from a mere robot or someone acting as a result of a mere tropism or epileptic seizure or even more intrusive and thoroughgoing manipulation. It is important to mark the distinction between us and both mere animals and automata; we have a distinctive sort of control that mere animals and automata lack. In my view, 'moral responsibility' is what we have and the mere animals and automata lack. It is also important to distinguish someone such as Plum in Case 1 from a fully and robustly blameworthy agent; some conditions have to be added to the minimal epistemic and control conditions to get us to praiseworthiness/blameworthiness. Moral responsibility is more 'abstract' than (say) blameworthiness; moral responsibility is the 'gateway' to the thicker, more robust notions of praiseworthiness and blameworthiness. Whereas moral responsibility is underwritten by control of a distinctive sort (guidance control, on my view), praiseworthiness and blameworthiness flow from such considerations as the circumstances of acquisition of basic values and dispositions, the possibilities for reflective evaluation of such features of one's mental economy, and the difficulty of (or reasonableness of) resisting one's desires and other motivational states (among others).

### 3. Reply to Zimmerman

David Zimmerman is willing to accept the distinction between being morally responsible (in the sense of being an apt target of the reactive attitudes) and its being justified, all-things-considered, to have one of these attitudes (and perhaps associated activities, such as punishment or reward). He even leaves it an open question whether someone could do something all-things-considered wrong and still it not be the case that the individual is justifiably blamed. He says:

The distinction that Fischer invokes is real enough. However, it has no bearing that I can see on the issue of whether a source-historicist semi-compatibilist ought to include a substantive condition. So, I continue to worry that *if* Fischer's theory is truly source-historicist, as he continues to insist that it is (*passim* in *My Way*), then it cannot limit itself to conditions that stop short of *substantive* value formation. (p. 227 above)

Zimmerman adds that my deployment of the distinction in reply to Pereboom's four-case argument is "dialectically irrelevant":

Fischer insists that the hapless Professor in all four cases is morally responsible for killing Mrs White, even though he is massively interfered with by the neurologists, because he is not compelled, and not a robot. However, Fischer's

reason for insisting that the Professor is nonetheless not blameworthy has nothing to do with the distinction he invokes, for withholding blame here is not a matter of its not being justified to blame him. What the Professor does is not morally neutral, and under the circumstances there is no countervailing justification for it. If we must withhold blame, then the only remaining reason is that Professor Plum is somehow exempted from it. (p. 227 above—the quoted passage is in brackets in the original.)

But I believe that it is *not* justified to blame Professor Plum in (say) Case 1 precisely because Plum is not blameworthy. I am inclined to agree that what Professor Plum has done is wrong, that he ought not to have done it, and so forth; but it does not follow, as far as I can see, that he is blameworthy for so acting. (Of course, if one held as a conceptual matter that if someone does something morally wrong, he is *eo ipso* blameworthy, then one would of necessity deny the possibility of moral responsibility for a wrong action without blameworthiness; here ‘blame’ would perhaps be merely some sort of judgement of wrongness. But this is evidently not Zimmerman’s position, and it is not at all obvious or uncontroversial that the conceptual point is true.)

It would be desirable if there were fairly ordinary and uncontroversial cases of agents who do the wrong thing but nevertheless are both morally responsible and not blameworthy. Unfortunately, the examples I think of are contentious. Consider, for example, cases of significant, long-term, and even brutal spousal abuse. As a result of such abuse, and in a context of considerable coercion (short of compulsion), a woman kills her abusive spouse. I think it is at least plausible that this sort of case is one in which there is moral responsibility, the agent does the wrong thing, but the context and provenance of her motivational states are such as to render her not blameworthy.

I think it is obvious (as I stated above) that considerations such as (but not limited to) how an individual’s values and preferences were formed are relevant to whether that individual is blameworthy for his behaviour. It would seem that the provenance of Professor Plum’s motivational states—his values, preferences, and so on—argues against his blameworthiness. But moral responsibility, in my view, can still be present; this is because moral responsibility is, in my view, a matter of how we play the cards that are dealt us. We don’t need to have made the cards, or own the factory that makes the cards, or have made up the rules of the game, and so forth. Similarly, moral responsibility is a matter of exercising a distinctive kind of control—guidance control—even in the absence of a more robust kind of control of the inputs to our mechanisms of practical reasoning.<sup>4</sup>

I am not inclined, then, to add a substantive source-historicist condition to my account of guidance control (and moral responsibility). But I may be wrong about this issue; after all, it is NOT ‘My Way or the Highway’! And if so, I would simply point out that it shouldn’t be required of someone offering

4. I have argued that what I have called Total Control is an unattainable fantasy—the product of a kind of metaphysical megalomania: John Martin Fischer, ‘The Cards that Are Dealt You’, *Journal of Ethics*, 10 (2006), pp. 107–29.

a theory of responsibility (or any philosophical theory) that he provide a reductive analysis of all key elements of the theory. Yes, that would be ideal. But a theory could still be helpful and illuminating if certain key ideas are unanalysed. So, for example, a historicist semi-compatibilist might invoke the (unanalysed) distinction between moral education and persuasion, on the one hand, and indoctrination on the other, pointing out that an ideal and completely comprehensive theory would offer some sort of reductive account of the distinction. The theory would then be like (say) causal theories of action (or frankly causal theories of just about anything) in leaving the ‘appropriate’ sort of causation (and thus the distinction between deviant and non-deviant causation) unanalysed. Some philosophers think that this sort of move is fatal to the theories. But I am inclined to think that it just shows that the theories, helpful and instructive as they may be, remain incomplete. The unanalysed notion can be considered a place-holder indicating a locus for future work.

#### 4. Reply to Smith

I have argued that the value of acting freely is the value of a certain kind of self-expression: we care about making a statement, rather than making a difference. When an agent acts freely, she writes a sentence in the book of her life; that is, unlike that of mere cows and rats, her life is a story and she has an irreducible narrative dimension of value. I believe that acting freely is the ingredient which, when added to others, renders humans beings such as to have genuine stories or narratives and also the narrative dimension of value; our lives can thus be meaningful in ways in which the lives of rats cannot be.<sup>5</sup>

Angela Smith distinguishes the point of view of the spectator from that of the agent. She suggests that it is especially unclear that we value writing narratives (in the specific sense sketched above) *as agents*. She says:

Do we, as agents, primarily value “render[ing] it true that [our] life has a certain distinctive kind of value: narrative value”? This is less clear to me. Part of what I find puzzling about this claim is that it makes no reference at all to our relations with other people. Fischer refers to the relations among different “sentences” in the book of our life, sentences whose meanings are “fixed in part by relationships to other sentences in this book, that is, by the overall narrative structure of the life”. But he never suggests that the meaning of our sentences might be fixed, in part, by our relationships with *other people*. This is surprising, because it seems to me that what we chiefly value about ‘telling our own story’ is the fact that there are others around who can appreciate, respond to, and perhaps even challenge the narratives we put forward. (p. 220 above)

5. John Martin Fischer, ‘Free Will, Death, and Immortality: The Role of Narrative’, *Philosophical Papers*, 34 (2005), pp. 379–404.

Smith helpfully suggests that what we care about might more accurately be described as ‘making a conversation’.

In the limited space I have here, I’d like to end with some sketchy and programmatic thoughts about Smith’s interesting ideas. First, I should simply mention that I am not convinced that what we care about would be different from the perspectives of the agent and the spectator; I have the hunch that there should be a unified account of what is valuable from these different perspectives—an account that exhibits a kind of resiliency across the various legitimate perspectives on our agency. This is not something I can argue for here, although I have made some preliminary attempts elsewhere.<sup>6</sup>

I believe that we do value self-expression, and that the value of acting freely (and thus being morally responsible agents) is a species of the value of self-expression. I am inclined to think of this as a sort of ‘artistic’ value, but if it is, it may be unique or *sui generis* (in a way I shall explain below). Why exactly do we value this sort of self-expression? This is a hard question. I think it is connected to questions about the meaning of our lives.

As I said above, our lives are stories, whereas the lives of rats and even cats are not. Certainly, one can tell the ‘story’ (speaking loosely) of a rock or a rat or a cat, but these accounts are not ‘stories’ in a strict sense. They are not narratives. Of course, it is an interesting and vexed question just what has to be added to a mere account or chronicle of events to get to a ‘story’ (strictly speaking) or a narrative; I am inclined to accept the bare bones at least of the suggestion of David Velleman that the necessary additional ingredient is some distinctive sort of *understanding* provided by the account.<sup>7</sup> Further, insofar as we act freely our lives have a narrative dimension of value. Along this dimension, one does not simply add together momentary levels of well-being, and the meanings or values of events depend on certain distinctive relationships with other events. Velleman calls these ‘dramatic’ or ‘narrative’ relationships.

When I act freely, I write a sentence in the story of my life; that is, the account of my life is strictly speaking a story (rather than a mere chronicle of events), and my life has a narrative dimension of value. Insofar as a story or narrative is by definition a work of art, I am an artist, when I act freely, and I am inclined to say that the value of my free action is the value of artistic self-expression. Of course, a work of art can be evaluated along various dimensions, including both aesthetic and moral. (One should distinguish the value of the activity from the value of the product, although they will of course be related.)<sup>8</sup> There is some sense in which the ‘primary’ dimension of assessment of a work of art is aesthetic (rather than moral). That is, given the intrinsic nature of the activity—artistic self-expression—there is a kind of hegemony of aesthetic evaluation (although, of course, different modes of evaluation will be appropriate in different contexts and in light of different purposes). Similarly, a life can be evaluated along different dimensions, including aesthetic, moral, prudential, and so forth. There is a sense in which

6. Fischer, ‘Free Will and Moral Responsibility’.

7. David Velleman, ‘Narrative Explanation’, *Philosophical Review*, 112 (2003), pp. 1–26.

8. For a discussion, see ‘Responsibility and Self-Expression’, *The Journal of Ethics*, 3 (1999), pp. 277–97; reprinted as Chapter 5 of *My Way*, pp. 106–23.

the moral and prudential modes of evaluation have hegemony when evaluating a human life, given the intrinsic nature of human agents (and the typical range of purposes of evaluation). Thus I suggest that if the value of acting freely is indeed a species of the value of artistic self-expression, it is *sui generis*: it is the value of artistic self-expression whose product is not typically or primarily evaluated aesthetically. That is, when we act freely, we tell a story that is most naturally—given the intrinsic nature of human activity and a broad range of human purposes—evaluated in terms of moral and prudential considerations.

It is perhaps not surprising that the two elements of this idea about the value of acting freely correspond in some rough way to the two components of the analysis of guidance control. The value of acting freely, on the account I have sketched, is the value of artistic self-expression whose product is primarily evaluated in terms of prudence and morality. Guidance control is analysed in terms of two components: ownership and reasons-responsiveness. Ownership posits a special relationship to the self, and this corresponds to the value of artistic self-expression. And reasons-responsiveness pertains to the salient dimensions of assessment of human lives in terms of morality and prudence. Although the relationships here are perhaps less tight than one might like, it is appropriate that there be echoes of the elements of the account of the *value* of acting freely in the *analysis* of acting freely. This is one way in which parts of the overall framework for moral responsibility form a unified, cohesive whole.

To take stock, I suggested above that the value we place on acting freely is connected in certain ways to the meaning of our lives. One way in which this is true is that acting freely renders us artists in the sense adumbrated above; in virtue of acting freely, our lives can be stories and can have narrative value. We care about acting freely, then, to the extent that we value engaging in artistic self-expression. It may be that we especially care about the sort of artistic self-expression whose product is typically evaluated in moral and prudential terms. But I think there is an additional feature of this sort of self-expression (connected to the meaning of our lives) in virtue of which we deem it valuable.

In Richard Taylor's fascinating essay, 'The Meaning of Human Existence', he argues that a crucial element in our lives' having the distinctive sort of meaning they possess is our power of *creativity*.<sup>9</sup> He says:

If you were to learn that the rest of your life would be spent digging an enormous hole, then it would perhaps be a reassurance of sorts to be told that you were actually going to enjoy doing it. If, further, you were born with, or at any early age conditioned to, a strong desire to do this, then you would not need to have such a task assigned to you—you would go to great lengths to gain the opportunity and consider yourself lucky if you got it. And you would someday view the great hole you had dug with a deep sense of fulfillment. And therein does each of us find, in varying degrees, the very picture of his or her own life. (pp. 23–4)

9. Richard Taylor, 'The Meaning of Human Existence', in Burton M. Leiser (ed.), *Values in Conflict: Life, Liberty, and the Rule of Law* (Macmillan, 1981), pp. 3–27.

I suppose that some might think that in writing *My Way*, I have already dug an enormous hole! (You might then think of these current efforts as (perhaps futile) attempts to dig my way out.)

Taylor goes on to say that what transforms our lives into *meaningful* lives is *creativity*. Whereas Smith laments the loss of the ‘God-like’ ambition to make a difference to the world, Taylor offers some consolation:

That one word [‘creative’] sums it up, and, if really understood, discloses entirely what is missing, not only in all the animate and inanimate existence that surrounds us but in the lives of the vast majority of human beings. It is also what philosophers have always sought as godlike or what makes man, in the ancient metaphor, the image of God. For what is godlike is not blind power, or aimless knowledge, or unguided reason, but simply creative power. It is the primary attribute in the very conception of God. (p. 24)

Chisholm thought of the agent as a kind of Godlike first cause. Perhaps we are indeed the images of God, but not in virtue of having the power to create *ex nihilo* or even to make a certain kind of difference to the world. Maybe it is simply in virtue of our power of artistic self-expression that we are creative in a way that renders us images of God, even if pale images.<sup>10</sup>

I turn briefly to Smith’s worry that my approach is too narcissistic or perhaps individualistic.<sup>11</sup> I believe that Smith and I are probably very close to agreement here. I suppose I would want to say that an isolated hermit could

10. Taylor says:

... creation is not just the creation of things. Creativity is a state of mind, which sometimes expresses itself in small and otherwise insignificant ways. Great or small, it is precious, and it is the only thing that finally converts life to meaning. No animal, for example, can look at nature, or at any object of nature, creatively, but a person can. A person can contemplate the simplest, and otherwise least significant thing creatively—can thus consider a blade of grass, a hill, a thunderstorm, a snowflake, virtually anything. . . . Meaningful thought need not be thought that has . . . any result beyond itself. One’s very thoughts can be poems, even if unuttered, in contrast to being trivial or banal or imitative, as most thought is most of the time. Consider two persons looking, say, at a meadow. One sees it for its size, its possible value, the use to which it might be put. He sees it, in short, only in terms of his own conditioned desires, rather as an animal would see it. The other, we can suppose, considers none of these things but is instead drawn to a tiny and insignificant flower at her feet and looks at it in a way that the other person is incapable of viewing it, in a way that no animal can view it. She looks at it creatively, not merely *finding* it meaningful, but investing it with meanings, by her own creative power. This is not the creation of an object, but it is creation just the same.

11. The author of an entry on the ‘EGR Weblog’, ([http://www.rageboy.com/2003\\_05\\_25\\_blogger-archive.html](http://www.rageboy.com/2003_05_25_blogger-archive.html)) [accessed April 17, 2006], compares the lyrics of ‘My Way’ to the definition of ‘narcissistic personality disorder’. Here is a brief quotation from the (lengthy) diatribe:

I did it my way, the guy says. Your sleek self-confident blue-eyed boy, and underneath, an arrogant urbane brutality. Swingin. Cool. As cold as they come. Where’s the love in Sinatra’s love songs, I’m wondering. . . . I wonder how he treated his women in Vegas. Like a gentleman, no doubt. Like one-a the guys. Like Luciano and Giancana. Smooth operators all. Made in the shade.

Regrets? Not me. F—you, Frank. (“Have It Your Way: DSM IV Meets Sid Vicious,” May 27, 2003).

still act freely and be morally responsible; he could still be an apt candidate for the reactive attitudes, even if there were no one (else) around to apply them, and it might conceivably be helpful to mark the distinction between his behaviour and that of his pet cat (or the animals he hunts and devours)—even though there would be no one (else) around actually to mark the distinction. But I certainly grant that the statements we care about (in acting freely) are *typically* parts of ‘conversations’.<sup>12</sup> In making such statements, we make connections—connections with other people and even causes ‘larger than ourselves’ that are valuable parts of meaningful lives. I am inclined to accept the idea that we care about writing sentences in the books (stories) of our lives, where these sentences are typically parts of conversations with others we care about. In thus making a statement, I make a connection. In writing my story, I help to write our story. My way becomes a part of our way.

12. I discuss the conversation model of moral responsibility in ‘Introduction: A Framework for Moral Responsibility’, *My Way*, pp. 1–37, especially pp. 26–7.