JOHN MARTIN FISCHER

RESPONSIBILITY AND SELF-EXPRESSION

(Received 18 November 1998; accepted 1 December 1998)

ABSTRACT. I present two different "models" of moral responsibility – two different accounts of what we value in behavior for which the agent can legitimately be held morally responsible. On the first model, what we value is making a certain sort of difference to the world. On the second model, which I favor, we value a certain kind of self-expression. I argue that if one adopts the self-expression view, then one will be inclined to accept that moral responsibility need not require alternative possibilities.

KEY WORDS: alternative possibilities, control, free will, moral responsibility, narrative, self-expression, weakness of will

To be morally responsible for one's behavior is to be an apt target for what Peter Strawson called the "reactive attitudes" – and certain associated practices – on the basis of it.[1] The reactive attitudes include resentment, indignation, hatred, love, gratitude, and respect. The associated practices include moral praise and blame, and reward and punishment.

Moral responsibility requires (among other things) control of one's behavior. But there are different kinds of control.[2] One sort of control entails the existence of genuinely accessible alternative possibilities; I call this sort of control, "regulative control." The presence of regulative control is typically signaled by the use of the preposition, "over." So, when an individual has control over his behavior, he has more than one path available to him; he (say) performs an action, but he could have done otherwise (in the sense of "could" that expresses the distinctive sort of ability involved in free will).

I believe that an agent can control his behavior, and be in control *of* it, without having control *over* it. In such a circumstance, the agent has

---

[1] Peter Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962), pp. 187–211.

[2] For the distinction between the two kinds of control, the claim that guidance control is sufficient for moral responsibility, and an elaboration of the notion of guidance control, see John Martin Fischer, *The Metaphysics of Free Will: An Essay on Control*, Vol. 14 of the Aristotelian Society Monograph Series (Oxford: Blackwell Publishers, 1994); and John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge Series in Philosophy and Law (Cambridge: Cambridge University Press, 1998).

The Journal of Ethics **3:** 277–297, 1999.

what I call "guidance control," but not regulative control. He guides his behavior in the way characteristic of agents who act freely, and yet he does not have alternative possibilities with respect to his decision or action. Of course, an agent may have both sorts of control – regulative control and guidance control. But the fact that an agent can have guidance control without regulative control shows that they are distinct forms of control.

I contend that moral responsibility requires guidance control, but not regulative control. That is, guidance control exhausts the "freedom-relevant" (as opposed to the epistemic) component of moral responsibility. In this paper I wish to provide a measure of intuitive support for the claim that guidance control is all the control (or freedom) necessary for moral responsibility. I begin by exploring some recent attempts to defend the view that alternative possibilities are required for moral responsibility (and thus that regulative control is an essential ingredient of moral responsibility). I shall propose what I take to be the intuitive "picture" that drives the view that alternative possibilities are required for moral responsibility. I go on to offer an argument against the view that regulative control is required for moral responsibility; on my view, this argument shows that the picture behind the regulative control view of moral responsibility is not the correct one – it doesn't capture what we value about moral responsibility. Finally, I shall develop an alternative picture which I believe both explains, at an intuitive level, what is going on with behavior for which an agent is morally responsible, and also helps to explain exactly why guidance control is all the control required for moral responsibility.

1. Responsibility and Regulative Control

1.1. *Frankfurt-Type Cases*

There can be cases in which an agent deliberates, chooses, and acts freely, on whatever your favorite account of such things is, and yet because of the presence of a fail-safe device which does not play any actual role in the agent's deliberation or behavior, the agent has no alternative possibilities with respect to choice or action. The fail-safe device does not actually intervene, but would intervene under certain counterfactual circumstances to produce exactly the same sort of choice and action as actually take place. Following recent tradition, I shall call such cases "Frankfurt-type" cases.[3]

[3] The classic source is, Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *The Journal of Philosophy* 66 (1969), pp. 829–839.

Here is a particular version of a Frankfurt-type case. In this sort of case, a crucial role is played by some kind of involuntary sign or indication of the agent's future choices and behavior.[4] So suppose Jones is in a voting booth deliberating about whether to vote for Gore or Bush. (He has left this decision until the end, much as some restaurant patrons wait until the waiter asks before making a final decision about their meal.) After serious reflection, he chooses to vote for Gore, and does vote for Gore by marking his ballot in the normal way. Unbeknownst to him, Black, a liberal neurosurgeon working with the democratic party, has implanted a device in Jones' brain which monitors Jones' brain activities.[5] If he is about to choose to vote democratic, the device simply continues monitoring and does not intervene in the process in any way. If, however, Jones is about to choose to vote (say) republican, the device triggers an intervention which involves electronic stimulation of the brain sufficient to produce a choice to vote for the democrat (and a subsequent democratic vote).

How can the device tell whether Jones is about to choose to vote republican or democratic? This is where the "prior sign" comes in. If Jones is about to choose at t_2 to vote for Gore at t_3, he shows some involuntary sign – say a neurological pattern in his brain – at t_1. Detecting this, Black's device does not intervene. But if Jones is about to choose at t_2 to vote for Bush at t_3, he shows an involuntary sign – a different neurological pattern in his brain – at t_1. This brain pattern would trigger Black's device to intervene and cause Jones to choose at t_2 to vote for Gore, and to vote for Gore at t_3.

Given that the device plays no role in Jones' deliberations and act of voting, it seems to me that Jones acts freely and is morally responsible for voting for Gore. And given the presence of Black's device, it is plausible to think that Jones does not have alternative possibilities with regard to his choice and action. So it appears that Jones is morally responsible for his choice and for voting for Gore, although he lacks regulative control over his choice and action.

At this point it may be objected that, despite the initial appearance, Jones *does* have at least *some* alternative possibility. Although Jones cannot choose or vote differently, he can still exhibit a different neurological pattern in his brain N^* (from the one he actually exhibits, N). I have called such an alternative possibility a "flicker of freedom." The flicker theorist contends that our moral responsibility always can be traced back

---

[4] For this kind of Frankfurt-type case, see David Blumenfeld, "The Principle of Alternate Possibilities," *The Journal of Philosophy* 67 (1971), pp. 339–344.

[5] Of course, this sort of example is a highly implausible science-fiction scenario, since most neurosurgeons are certainly not liberal!

to some suitably placed flicker of freedom; our responsibility is grounded in and derives from such alternative possibilities.

I concede that one can always find a flicker of freedom in the Frankfurt-type cases insofar as they are developed as "prior-sign" cases. That is, the agent will always at least have the power to exhibit an alternative sign. But I contend that the mere involuntary display of some sign – such as a neurological pattern in the brain, a blush, or a furrowed brow – is too thin a reed on which to rest moral responsibility. The power involuntarily to exhibit a different sign seems to me to be insufficiently robust to ground our attributions of moral responsibility.

I have argued for this contention at some length elsewhere.[6] The debate here is subtle and complex; there are different versions of the flicker strategy, and various different responses. But for my purposes in this paper perhaps it will be enough to reiterate one line of argument I have developed against the flicker approach. Note that in the alternative sequence (in which Jones shows neurological pattern N^*, which is indicative of an impending decision to vote for Bush), the sign is entirely involuntary and the subsequent decision and vote are produced electronically. Thus, in the alternative sequence Jones cannot be said to be choosing and acting freely, and similarly, cannot be thought to be morally responsible for his choice and action.

Imagine, just for a moment, that there are absolutely no alternative possibilities, even the flimsy and exiguous flickers of freedom we have recently been entertaining. A regulative control theorist would say that under such circumstances the relevant agent cannot be morally responsible for his choice and action. Now add the flickers of freedom we have been considering – the power to exhibit a different neurological pattern, N^*. I find it very hard to see how adding this power can transform a situation in which there is no moral responsibility into one in which there is moral responsibility. How can adding a pathway along which Jones does *not* freely vote for Gore and is *not* morally responsible for voting for Gore make it the case that Jones actually *is* morally responsible for voting for Gore? This would seem to be alchemy, and it is just as incredible.[7]

Similarly, suppose one had a theory of knowledge according to which some individual *S* (the individual in question is always called "*S*"!) knows that *p* only if *S* can discriminate *p* from relevant alternatives. This is structurally analogous to the view that moral responsibility requires regulative control. Whereas such a view is plausible, it would certainly be absurd to suppose that what transforms some case of lack of knowledge into a case of

---

[6] Fischer, *The Metaphysics of Free Will*, pp. 131–159.

[7] Fischer, *The Metaphysics of Free Will*, p. 141.

knowledge would be the existence of some alternative scenario in which the agent makes a mistake. How can adding a scenario in which *S* lacks knowledge (in this way) make it the case that *S* actually has knowledge?[8]

Now of course it has been suggested that on the "relevant-alternatives" model of knowledge, it might paradoxically turn out that one knows less by knowing more. That is, by having more background knowledge one makes fewer alternatives relevant, and it thus becomes more difficult to rule out these alternatives. The "flip-side" of the coin is that one can know more by knowing less. But this is a matter of having less background information; it is not a matter of adding a scenario in which one makes a mistake to a situation of lack of knowledge to transform it – almost as if by magic – into a situation in which one has knowledge.

1.2. *The "New Defense" of Regulative Control*

Recently a number of philosophers have defended the regulative control model in a way that might seem to be promising, even in light of the sort of argument I have just sketched. Basically the strategy involves identifying some more robust alternative possibility which exists, even in the Frankfurt-type examples (with the prior-sign structure). The proponents of this strategy might concede that the power to exhibit a different involuntary sign is a mere flicker of freedom, but they will insist that there are deeper, more important kinds of powers possessed by agents in the examples.

Consider, for example, the following remarks of Michael McKenna:

> Here I believe that Fischer has not fully addressed what motivates the advocate of [regulative control] ... what intuitively drives [the proponent of regulative control] is the kind of control needed in order for us to avoid being the author of a *particular* act and thus avoid being responsible for the production of *that* particular action ... It is a matter of holding people accountable for what they do only if they can avoid any blame or punishment that might fall upon them for performing those very particular actions which they do perform ...[9]

McKenna elaborates as follows:

> The issue ... here is whether the will ... places *my* stamp upon the world, and whether *its is up to me ... to have that particular stamp or some other as my mark upon the world.* In the Frankfurt-type cases the alternatives are, either doing what one does of one's own intention, or being coerced into performing the same kind of action against one's will. These alternatives do seem to be quite impoverished; however, they mean all the difference between one's doing something of one's own will, and one's not doing that kind of thing

---

[8] Fischer, *The Metaphysics of Free Will*, pp. 141–142.

[9] Michael S. McKenna, "Alternative Possibilities and the Failure of the Counterexample Strategy," *Journal of Social Philosophy* 28 (1997), pp. 71–85; the quotation is from pp. 73–74.

of one's own will ... What more fundamental kind of control can there be here other than the control for one to either have a particular will or not have it?[10]

McKenna is claiming that even in the Frankfurt-type cases, the relevant agent has a significant and robust power: the power either to be the author of his action or not, and thus the power to be morally responsible for his action or not. A similar point is made in an interesting recent article by Keith Wyma.[11] Wyma begins with an example which suggests that many of us experienced something like a Frankfurt-type example as we were growing up:

When I was four years old and learning to ride a bicycle, I reached a point where my father decided I no longer needed training wheels. But he still worried that I might fall. So on my first attempt "without a net," he ran alongside as I pedaled. His arms encircled without touching me, his hands resting lightly upon me, but not holding me upright. I rode straight ahead. My father did not push or guide me, but if I had faltered or veered suddenly to the side, he would have tightened his grip, keeping me vertical and on track. After finally braking to a stop, I was jubilant but somewhat hesitant over whether I should be. I wondered, had I really ridden my bike on my own? ... Was the triumph of riding straight down the street mine or not?[12]

Wyma goes on to argue for an intuition very similar to McKenna's. On Wyma's view, moral responsibility requires a certain kind of "leeway." And this leeway is specified by what Wyma calls the "Principle of Possibly Passing the Buck" (PPPB):

A person is morally responsible for something she has done, *A*, only if she has failed to do something she could have done, *B*, such that doing *B* would have rendered her morally non-responsible for *A*.[13]

Of course, in a Frankfurt-type case the relevant agent would not be morally responsible in the alternative sequence; Jones would not be morally responsible for voting for Gore, in the circumstance in which Black's device were triggered. Thus Wyma has apparently identified a significant sort of "leeway," even in the Frankfurt-type examples. At the end of his paper, Wyma returns to the analogy with which he started, saying:

I believe the bike riding triumph *was* mine, because even though I could not have fallen or crashed while my father hovered protectively over me, I could still have faltered enough that he would have had to steady me; and because I had leeway to falter but did not do so,

[10] McKenna, "Alternative Possibilities and the Failure of the Counterexample Strategy," pp. 74–75.

[11] Keith D. Wyma, "Moral Responsibility and Leeway for Action," *American Philosophical Quarterly* 34 (1997), pp. 57–70.

[12] Wyma, "Moral Responsibility and Leeway for Action," p. 57.

[13] Wyma, "Moral Responsibility and Leeway for Action," p. 59.

the success of riding was truly mine. *PPPB* vindicates a similar kind of leeway as being necessary for ascriptions of moral responsibility.[14]

Additionally, Michael Otsuka has recently defended a principle similar to Wyma's Principle of Possibly Passing the Buck. Otsuka calls his principle, the "Principle of Avoidable Blame":

> One is blameworthy for performing an act of a given type only if one could instead have behaved in a manner for which one would have been entirely *blameless*.[15]

Thus, all three defenders of regulative control seem to be pointing to the same sort of alternative possibility which they claim is present quite generally, and hence even in the Frankfurt-type examples. This is the freedom to "pass the buck" or "escape" or "avoid" moral responsibility. One might say that these theorists are seeking to fan the flickers of freedom.

In my view there is an intuitive picture that drives all proponents of regulative control, no matter what sort of alternative possibility they identify as grounding ascriptions of moral responsibility. The idea is that moral responsibility requires *making a difference*. Slightly more carefully, an agent is morally responsible for his behavior only if he makes a difference to the world in so behaving. But of course an agent can in some sense make a difference when performing an action under coercion, duress, or (say) direct electronic stimulation of the brain. Given that the agent acts in such cases, the world is different than it would have been, had he not so acted. Obviously, this mere counterfactual difference is *not* the sort of difference envisaged by the proponent of the regulative control model of moral responsibility.

Rather, the regulative control theorist believes that moral responsibility requires the ability to make a difference in the sense of selecting one from various paths the world could take, where these various paths are all genuinely available to the agent. The basic idea here is *selection* from among options that are really accessible to the agent. When one selects from a set of feasible options, one makes a difference: the world goes one way rather than another, or takes one path rather than another, among various paths the agent can cause the world to take. This, I believe, is the basic intuitive idea behind the regulative control model. The recent defenses of regulative control help to make the idea more compelling by identifying an *important kind* of difference. On this view, an agent is morally responsible insofar as

---

[14] Wyma, "Moral Responsibility and Leeway for Action," p. 68.

[15] Michael Otsuka, "Incompatibilism and the Avoidability of Blame," *Ethics* 108 (1998), p. 688. Otsuka qualifies the principle to apply to cases in which it is not the case that everything one is capable of doing at a given point in time is blameworthy because of some previous choice for which one is to blame.

he makes an important difference to the world: he selects a world in which he is accountable for his behavior, rather than one in which he is not.

1.3. *A Reply*

Despite the manifest appeal of the new defenses of the regulative control model, I remain unconvinced. I believe that problems similar to the problems with the earlier defenses of regulative control also plague the new approaches. Recall that the problem with saying that it is the possibility of exhibiting a different prior sign or indicator of future decision (and action) that grounds moral responsibility is that the envisaged possibility is too exiguous and flimsy. The displaying of such a sign would not even be voluntary behavior. How could moral responsibility rest on such a delicate foundation?

Now it might be thought that the possibility of avoiding authorship or the possibility of avoiding moral responsibility would be a more substantial basis for moral responsibility. But I believe there are similar problems here. Note that in the alternative sequence in a Frankfurt-type case the agent would indeed be avoiding (say) moral responsibility, but he would be doing so "accidentally." The agent would *not* be *voluntarily* avoiding responsibility. The suggestion that avoiding responsibility is a sufficiently robust basis for moral responsibility may get some of its plausibility from the fact that in a typical context in which we would say that someone has avoided (say) blameworthiness, it would be in virtue of some voluntary action. Typically, the relevant facts about the various paths available to the agent would be accessible to him, and he would voluntarily choose a right action (rather than a morally objectionable one). Here we would say that the agent avoided blameworthiness; but this is a very different sort of context from the Frankfurt-type cases. In the Frankfurt-type cases, the agent does not choose to be morally responsible rather than not – these issues play no role in his deliberations. And in the alternative scenario in a Frankfurt-type case (of the prior-sign variety), the agent does not choose to escape responsibility, or voluntarily choose anything which implies his escaping responsibility.

To bring this point out a bit more clearly, note that in the alternative scenario in a Frankfurt-type case the agent does not deliberate about whether or not to embrace moral responsibility. So issues about whether or not to be morally responsible play no explicit role in his deliberations. Further, they play no "implicit" role either. They might play an implicit role in the sort of context discussed above in which an agent has internalized certain norms on the basis of which he chooses to do what he takes to be the right action. If he successfully avoids blameworthiness here, it is

partly in virtue of his having internalized norms the relevant community shares. Given these norms, the agent can reasonably expect to escape blame, if he chooses as he does. But in the alternative scenarios in the Frankfurt-type cases issues about moral responsibility obviously do not play an implicit role of this sort.

To the extent that issues pertaining to moral responsibility play neither an explicit nor an implicit role, I shall say that moral responsibility is not "internally related" to the agent's behavior in the alternative sequence of a Frankfurt-type case. And my point is that it is very plausible that moral responsibility must be so related to the agent's behavior, in order for the alternative possibility in question to be sufficiently robust to ground ascriptions of moral responsibility.

Of course, I do not accept the "alternative-possibilities" or regulative control model of moral responsibility. But my contention is that, *if* you do buy into this traditional picture, then you should *also* accept that the alternative possibilities must be *of a certain sort* – they must be sufficiently robust. This same point has been highlighted by a philosopher with a very different orientation from mine: Robert Kane.[16] (Kane is a libertarian who believes that alternative possibilities are required for moral responsibility.) Kane emphasizes what he calls the "dual" or "plural" voluntariness (and responsibility) conditions on moral responsibility: the relevant alternative possibilities – i.e., alternative possibilities sufficiently robust to ground moral responsibility – must themselves involve voluntary behavior (for which the agent is morally responsible). On Kane's picture, it is not enough that an agent have *just any sort of alternative possibility*; it must be an alternative in which the agent acts voluntarily and is morally responsible. Similarly, I would contend that the relevant alternative possibilities must contain voluntary, responsible behavior in which moral responsibility is internally related to the agent's behavior. My view, then, is that the new defenses of the regulative control model discussed above fall prey to the same sort of problem that afflicted earlier such defenses: the alternatives they postulate are not sufficiently robust.

In Frankfurt-type cases, an agent is morally responsible for his action, although he lacks the relevant kinds of alternative possibilities. He cannot, then, make a relevant difference to the world; he does *not* (in the appropriate way) select one path for the world to take, among various genuinely open paths. But the agent is nevertheless fully and robustly morally responsible for what he does.

---

[16] Robert Kane, *Free Will and Values*, SUNY Series in Philosophy (Albany: State University of New York Press, 1985), esp. p. 60; and *The Significance of Free Will* (New York and Oxford: Oxford University Press, 1996), esp. pp. 107–115.

Return to Wyma's striking claim about his early bike-riding experience, "...I believe the bike riding triumph *was* mine, because even though I could not have fallen or crashed while my father hovered protectively over me, I could still have faltered enough that he would have had to steady me; and because I had leeway to falter but did not do so, the success of riding was truly mine."[17] Whereas we could quibble endlessly about details of these sorts of examples, it seems to me that the *intuitive point* is quite clear: it is *not* the possibility of faltering slightly that makes the young Wyma's bike riding triumph truly his. This has to do *not* with whether he could have faltered slightly, but with how he rode the bike – how he moved the pedals, balanced, and so forth, and by what sort of causal process this all took place.

Wyma says, "[The Principle of Possibly Passing the Buck] begins to map out the negative space around the positive core of moral responsibility, similar to the way one might produce a silhouette by coloring in the space outside a person's profile."[18] But in focusing on the negative space, one can be distracted from what really counts; there is a danger that one will be looking at mere shadows, as with Plato's cave-dwellers. Rather than charting the negative space around moral responsibility, I have a modest suggestion for Wyma: Think positive!

La Rochefoucauld suggested that we can learn about death only by not focusing directly upon it, just as it is prudent to avert one's eyes from the sun. I am not sure that he is correct about death, but in any case I would suggest that even if so, moral responsibility is crucially different from the sun and death: in order to understand why someone is morally responsible for his behavior, we ought not to avert our eyes or focus on the "negative space;" we ought to gaze directly at the properties of the causal process that issues in the behavior in question.

2. Guidance Control and Self-Expression

Our moral responsibility, then, is not – at least in my view – based on our capacity to make a difference to the world. I grant that reasonable people can disagree with this conclusion (and with the associated claim that regulative control is not required for moral responsibility). That is, I concede that the plausibility-arguments I offered above (including the claim that responsibility must be internally related to the agent's behavior in the alternative sequence) are not decisive; they leave room for a defense

[17] Wyma, "Moral Responsibility and Leeway for Action," p. 68.

[18] Wyma, "Moral Responsibility and Leeway for Action," p. 68.

of the regulative control model. I want now to seek to sketch (in what will no doubt be a preliminary way) a different intuitive picture of moral responsibility. With this alternative picture in hand, I will return to the issue of whether regulative control is necessary for moral responsibility.

Begin with an analogy with artistic creativity. Suppose a sculptor creates a sculpture in the "normal" way – the sculptor is not being manipulated, coerced, and so forth, and is driven by his own creativity. But imagine further that, if he hadn't created this sculpture, some other artist would have created exactly the same sort of sculpture – a different particular sculpture that is nonetheless molecule-for-molecule isomorphic to the sculpture actually produced by the artist. I am not sure why exactly the production of this sort of sculpture is overdetermined in this way, but it really doesn't matter exactly why this is so – only that it is so. It may be, for example, that a friend of the artist has discussed the sculpture with him, and is bent on producing it, if the artist doesn't do so himself.

There is a pretty clear sense in which the artist does not make a difference to the world in creating the sculpture. He does not make a difference defined in terms of end-states (individuated in a natural, broad way). The very same kind of sculpture would have been produced, had he not created the sculpture himself.[19] And yet there is also a clear sense in which the artist's creative activity has value. I suggest that we value the artist's activity not because he makes a certain sort of difference to the world, but because he expresses himself in a certain way. He does not make a difference; but he *does* make a statement.

My idea is that we can understand the intuitive picture behind moral responsibility in a similar way. When an agent exhibits guidance control and is thus morally responsible for his behavior, he need not be understood to be making a difference to the world; or better, it is unattractive to think that the explanation of his moral responsibility – the intuitive reason why we hold him morally responsible – is that he makes a difference to the world. Rather, the suggestion is that the individual is morally responsible, when he exhibits guidance control, insofar as he *expresses himself* in a certain way. To a first approximation, the "value" of morally responsible action is understood as analogous to the value of artistic self-expression.

[19] Of course, someone might point out that it must be a different particular sculpture in the alternative sequence, since it would have been created by a different individual from the actual artist. I do not deny that one can say this, or that one can contend that the value of the artist's creative activity then consists in making a difference – in creating the actual sculpture rather than a different particular sculpture. But I do not find this explanation as natural and compelling as the explanation sketched in the text. It seems problematic in the same way as the flicker of freedom strategy for explaining the value of acting so as to be morally responsible.

But if the value of moral responsibilly responsible action is self-expression, what exactly is expressed? This is a question that deserves an answer, and yet it is perhaps not as easy as one might have supposed to answer it. Consider, for example, the following passage from Sarah Broadie's book, *Ethics With Aristotle*:

> In voluntary action we pursue an objective which is before us and which figures as a good to us so far as we pursue it; but on another level we enact by our action, and thereby propound into public space, a conception of the kind of practical being that it is good (or at least all right) to be: a kind typified by pursuit of this kind of goal in this sort of way under such conditions.[20]

The problem with the sort of view suggested by Broadie, in my view, is the very real phenomenon of weakness of will. That is, one sometimes freely does what one does *not* believe is good, or rational, or even all right. So, in voluntarily and freely performing some act, it would *not* in general be accurate to take one to be saying that the relevant goal is good or even all right, or that it is good or all right to be the kind of practical being that typically pursues this kind of goal.[21]

I suppose it could be urged that one is at least saying that the goal in question is *to some degree* good, and so one could be taken to be expressing the idea that it is at least to some degree good or defensible to be the sort of practical being who pursues such a goal. But even this seems implausible, as one can presumably freely do something one does not find to any degree good or morally defensible. Now perhaps it will be replied that whenever one acts, one must have some sort of pro-attitude toward the behavior in question (or the goal it is taken to promote). I agree, but it is somewhat disappointing to be told that the message conveyed by the agent in acting is something we know, as a conceptual point, from the mere fact that the agent has performed an action. Whereas I do not wish to deny that one can find here part of the message of action, I believe that it will be more illuminating to seek an alternative account of what is expressed by the agent in acting.

To develop such an account, I begin by noting that various philosophers have suggested that our lives have "narrative structures" – that our lives are in some sense stories.[22] This is an intriguing and suggestive idea (even

---

[20] Sarah Broadie, *Ethics With Aristotle* (Oxford: Oxford University Press, 1991), p. 159.

[21] A similar problem afflicts the view that in voluntarily and freely performing some act, one is "standing for something."

[22] See, for example, Alasdair MacIntyre, *After Virtue* (Notre Dame: University of Notre Dame Press, 1981); Charles Taylor, *Sources of the Self: The Making of Modern Identity* (Cambridge, MA: Harvard University Press, 1989); and Alexander Nehamas, *Nietzsche: Life as Literature* (Cambridge, MA: Harvard University Press, 1985).

if it is difficult to flesh out precisely), and different philosophers have developed it in different ways (and for different purposes). Here I shall rely on David Velleman's presentation of the idea in his seminal paper, "Well-Being and Time."[23]

Velleman is concerned to argue that "well-being is not additive." This claim involves various ideas. One is that we cannot simply add up the welfare values of segments of an individual's life to get a total value that accurately reflects our judgments about the value of the individual's life as a whole. Another idea is that the welfare values of the segments depend crucially on their "narrative" or "dramatic" relationships with other parts of the life.

Velleman says:

> Consider two different lives that you might live. One life begins in the depths but takes an upward trend: a childhood of deprivation, a troubled youth, struggles and setbacks in early adulthood, followed finally by success and satisfaction in middle age and a peaceful retirement. Another life begins at the heights but slides downhill: a blissful childhood and youth, precocious triumphs and rewards in early adulthood, followed by a mid-life strewn with disasters that lead to misery in old age. Surely, we can imagine two such lives as containing equal sums of momentary well-being.
>
> ... Yet even if we were to map each moment in one life onto a moment of equal well-being in the other, we would not have shown these lives to be equally good. For ... one is a story of improvement while the other is a story of deterioration. ... the former story would seem like a better life-story – not, of course, in the sense that it makes for a better story in the telling or the hearing, but rather in the sense that it is the story of a better life.[24]

Now it might be thought that the moral of Velleman's story is that we have a general tendency to weight welfare that occurs *later* in life more heavily. But whereas this would issue in a non-additive conception of welfare, it is not the moral Velleman wishes to draw. Rather, Velleman says:

> The reason why later benefits are thought to have a greater impact on the value of one's life is not that greater weight is attached to what comes later. Rather, it is that later events are thought to alter the meaning of earlier events, thereby altering their contribution to the value of one's life. [Additionally] ... [t]he meaning of a benefit depends not only on whether it follows or precedes hardships but also on the specific narrative relation between the goods and evils involved.[25]

To illustrate this point, Velleman gives the example of the importance of drawing lessons from one's misfortunes. We typically think it important

---

[23] J. David Velleman, "Well-Being and Time," *Pacific Philosophical Quarterly* 72 (1991), pp. 48–77; this paper is reprinted in John Martin Fischer (ed.), *The Metaphysics of Death*, Stanford Series in Philosophy (Stanford: Stanford University Press, 1993), pp. 329–357 (all subsequent pages references will be to the reprinted paper).

[24] Velleman, p. 331.

[25] Velleman, pp. 334–335, and p. 336.

to learn from life's tragedies; the fact that we have been improved *as a result of going through a tragic experience* adds to the total value of our lives in a distinctive way. As Velleman puts it:

> If a life's value were a sum of momentary well-being, learning from a misfortune would be no more important than learning from other sources, since every lesson learned would add so much value and no more to the sum of one's well-being. On being invited to learn from a personal tragedy, one would therefore be entitled to reply, "No, I think I'll read a book instead." Edification would offset the losses incurred in the tragedy, but its having been derived from the tragedy would not render edification more valuable ... [26]

Velleman similarly asks us to consider two lives. In the first life you have ten years of unhappiness and trouble in a marriage followed by divorce, after which you remarry happily. In the second life the ten years of unhappiness in marriage lead to eventual happiness as the relationship matures.[27] About this example, Velleman says:

> Both lives contain ten years of marital strife followed by contentment; but let us suppose that in the former, you regard your first ten years of marriage as a dead loss, whereas in the latter you regard them as the foundation of your happiness. The bad times are just as bad in both lives, but in one they are cast off and in the other they are redeemed. Surely, these two decades can affect the value of your life differently, even if you are equally well off at each moment of their duration.[28]

I shall follow Velleman in contending that life has a narrative structure in the specific sense that the meanings and values of the parts of our lives are affected by their narrative relationships with other parts of our lives, and the welfare value of our lives as a whole are not simple additive functions of the values of the parts. In this sense, then, our lives are stories.[29] And in performing an action at a given time, we can be understood as writing a sentence in the book of our lives.

I suggested above that the distinctive value in acting in such a way as to be held morally responsible lies in a certain sort of self-expression. But the question then arose as to what precisely is expressed by ordinary actions. My answer is that it is not most fruitful to look for a "message" of action of the sort suggested by Broadie – that one believes that the action promotes a defensible goal. Rather, what is expressed by an agent in acting is the meaning of the sentence of the book of his life. And this meaning is fixed in part by relationships to other sentences in this book,

---

[26] Velleman, p. 336.

[27] Velleman, p. 337.

[28] Velleman, p. 337.

[29] As Velleman points out, this view should not be confused with the view sometimes attributed to Nietzsche that literary or aesthetic considerations determine the value of a life (Nehamas, *Nietzsche: Life as Literature*).

i.e., by the overall narrative structure of the life. In acting, an individual need not be "propounding into public space" *any* sort of vision of the good or defensible life. Rather, his action writes part of the book of his life, and gets its meaning from its place in this story. This suggests a more "holistic" picture of what gets expressed by an agent in acting so as to be morally responsible, and one which is more illuminating than the mere fact that the agent had a pro-attitude toward moving his body in a certain way in the context.

I have sketched an analogy between action for which an agent can be held morally responsible and artistic self-expression. Further, I have claimed that when one exhibits guidance control one can be understood to be engaging in a specific kind of self-expression: one is writing part of the book of one's life. It does not follow however that the self-expression involved in action is a kind of *artistic* self-expression. Obviously one can write narratives that are not most appropriately categorized as works of art; for example, one can simply write a history of a region or family. This sort of narrative can have precisely the characteristics identified by Velleman without being properly considered a work of art.[30] The dimensions of assessment of this kind of narrative are not primarily aesthetic. Similarly, the dimensions of assessment of a human life are not primarily aesthetic, but moral and prudential.

I have tried to give an account of what might be called the value we place on acting so as to be morally responsible. In so doing, I have been seeking to sketch what I have suggested is the "picture" which grounds the guidance-control model of moral responsibility – the view that guidance control, and not regulative control, is the freedom required for moral responsibility. Whereas I believe I have put some of the elements of this picture in place, I still don't think I have fully captured the value of acting so as to be morally responsible.

To explain. Consider someone who is, intuitively, not morally responsible for what he does because he is to a significant degree subject to coercion, manipulation, and pressures that render his behavior not suitably responsive to reasons. Such an individual may nevertheless express himself in the relevant way: he may write the story of his life, a story to which certain moral and prudential judgments can attach. That is, presumably we can evaluate this individual's behavior in such a way as to judge it as good or bad, prudent or imprudent, and so forth. Now of course we need to *distinguish* these normative judgments from the further normative judgments and attitudes constitutive of *moral responsibility*: the reactive attitudes (such as indignation, resentment, gratitude, respect, and so forth).

[30] For this point I am indebted to Eric Schwitzgebel.

The sort of individual in question can live a life that is legitimately judged in terms of the first kind of normative considerations, but not the second.

Why not simply specify that the picture that grounds moral responsibility requires that the individual's life be subject to normative judgments of the second kind, i.e., those involved in the reactive attitudes? Perhaps one could do this, but I feel uncomfortable doing so because it seems to introduce a troubling circularity. My project is to identify what I have called the "picture" that supports the claim that guidance control, and not regulative control, is required for moral responsibility. Alternatively, I have characterized my project as seeking to identify the value we place on acting so as to be morally responsible. Ideally, it seems to me, we should be able to specify this value without importing the notion of moral responsibility. And yet to require that one's life-story be accessible to normative evaluation in the sense of the appropriate application of the reactive attitudes would do precisely this, for moral responsibility just is rational accessibility to the reactive attitudes.

In other words, I am trying to identify what exactly we value in cases in which we behave so as to be morally responsible. (Having done this, I want to employ the result – the value of acting so as to be morally responsible – to suggest that guidance control, and not regulative control, exhausts the freedom-relevant component of moral responsibility.) If I were to say, "Well, what we value is acting responsibly," this would obviously be circular and uninteresting – although no doubt true! I believe that it is similarly circular and unhelpful to say that the value of acting so as to be responsible is cashed out in part in terms of acting so as to be accessible to the reactive attitudes.

A more promising approach is to note that when one is subject to coercive pressures, manipulation, and so forth, one's self-expression is hindered in certain ways. What one wants to say, I believe, is that the value of acting so as to be morally responsible consists in *unhindered* or *unimpaired* self-expression of the relevant sort. Perhaps another way of saying the same thing is to note that when one engages in unhindered or unimpaired self-expression of the relevant kind, one is *freely* expressing oneself. I will then suggest that the value of acting so as to be morally responsible consists in one's freely expressing oneself. We value freely – in the sense of not being hindered or impaired in certain ways – writing the book of our life.

Now someone will say that I have introduced a problematic and contested notion here – the notion of "freely" expressing oneself. I have avoided the circularity mentioned above only by introducing an essentially contested notion: out of the frying pan and into the fire! After all, the

proponent of the regulative control model will insist that when one freely does anything, one must have genuinely accessible alternative possibilities. Now I certainly cannot present any decisive arguments that the ordinary notion of "acting freely" does not require alternative possibilities; indeed, this debate will presumably simply re-inscribe the debate about whether moral responsibility requires regulative control.

So I will simply stipulate a special notion of "acting freely," call it "acting freely*." When one "acts freely*" one need not have any alternative possibilities. The intuitive idea of acting freely* is that in the actual sequence that leads to one's behavior, no freedom-undermining factors operate or play a role. So, in the Frankfurt-type cases, one uncontroversially is acting freely*, even if it is controversial whether one is acting freely. Put in other words, in the Frankfurt-type cases a proponent of alternative possibilities as a condition for moral responsibility may say that insofar as acting freely is sufficient for moral responsibility, one of course needs alternative possibilities to act freely. But he should be willing to concede that there is some "actual-sequence" notion of freedom which the agent possesses, acting freely*; the agent possesses this freedom insofar as no freedom-undermining factor operates in the actual sequence that issues in the behavior. Of course, the proponent of alternative possibilities will go on to insist that *such* freedom is not sufficient for moral responsibility. Indeed, he will contend that the agents in the Frankfurt-type cases *also* possess a more robust (from his point of view) kind of freedom – one involving alternative possibilities. For my purposes, I simply want to crystallize out the "actual-sequence" notion of freedom, acting freely*.

My contention then is that the value of acting in such a way as to be morally responsible consists in freely* expressing oneself. Although this account needs to be filled in various ways, I think that it helps to capture something simple and important: the value of moral responsibility, on the guidance control model, consists in a distinctive kind of self-expression.[31]

With this sketch of an account of the value we place in acting so as to be responsible in hand, let us return, finally, to the issue of whether regulative control is required for moral responsibility. If one is in the grip

[31] I do not have an account of what precisely self-expression consists in, nor do I have a good explanation of *why* we value it. It does seem to me that we *do* in fact value something we conceive of as self-expression – something analogous to artistic self-expression. But a full defense of the view I have sketched in the text would say more about what self-expression is, and exactly why we value it (I am indebted to Paul Hoffman for pushing me on these points). Some are not inclined to find self-expression particularly valuable; these may however be precisely the same people who are not inclined to ascribe much intrinsic value to acting so as to be morally responsible. My contention is that the value, whatever one takes this to be, consists in a certain kind of self-expression.

of the picture according to which an individual must be able to make a difference, in order to he held morally responsible, one will press for the regulative control requirement. And one might not see any other plausible picture. This is part of the reason why I believe it is useful to have sketched the "self-expression" picture, which I have presented as underlying the guidance control model.

I suggest that some of the debates about whether alternative possibilities are required for moral responsibility may at some level be fueled by different intuitive pictures of moral responsibility. It may be that the proponents of the regulative control model are implicitly in the grip of the "making a difference" picture, whereas the proponents of the guidance control model are implicitly accepting the "self-expression" picture. Further, I would like to suggest that presenting the self-expression picture can be helpful for the following reason. The debates about whether alternative possibilities are required for moral responsibility have issued in what some might consider stalemates; above I conceded that I do not know of any *decisive* arguments (employing Frankfurt-type examples) for the conclusion that only guidance control, and not regulative control, is required for moral responsibility. My suggestion is that if one finds the self-expression picture of moral responsibility more compelling than the making-a-difference picture, then this should incline one toward the conclusion that guidance control exhausts the freedom-relevant component of moral responsibility.

Again, I do not suppose that this will be a knockdown argument; specifically, I do not suppose that those strongly inclined toward the regulative control model will find the self-expression picture correct. But this certainly should not be surprising; I do not think anyone should expect knockdown arguments in this realm. My point is that if direct reflection on the Frankfurt-type cases does not in itself issue in a decisive conclusion, one can perhaps be moved a bit closer to accepting the guidance control model of moral responsibility by seeing that it is supported by a natural and compelling intuitive picture – a picture one might not have seen, given the clout of the "make-a-difference" picture.

3. An Objection

The practices involved in moral responsibility have sometimes been modeled along the lines of a conversation.[32] On this view, the "reactive

[32] Gary Watson, "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in Ferdinand Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays on Moral Psychology* (Cambridge: Cambridge University Press), pp. 256–286.

attitudes" – such as resentment, indignation, love, hatred, gratitude, and respect – are responses to "statements" made by the agent in acting. One might have thought that a "self-expression" account of what we value in behavior for which the agent is morally responsible would fit naturally with a conversation model of moral responsibility. But upon reflection it can seem that if a conversation model of moral responsibility is correct, then the view I have been developing here about the content if the agent's self-expression must be is wrong. This is because the reactive attitudes are reactions to an agent's good or ill will (or indifference), as manifested in his actions. And this is quite a different matter from some sort of sentence in a book of the agent's life, the meaning and value of which is determined holistically (in terms of its dramatic relationships to what has come before and what will come after). The reactive attitudes are, after all, *direct* responses to particular bits of behavior; one obviously cannot wait to show resentment or gratitude until one has allowed the agent's entire life to play itself out (so that the appropriate "meaning" can be attached to the behavior)!

It is important, however, to distinguish different layers of meaning. Morally responsible behavior is a complex phenomenon, with various different features. It is not surprising that it may well have different layers of meaning (and different features that are relevant for different purposes). There is no doubt that the reactive attitudes are keyed to features of behavior that reflect the quality of the agent's will (his good or ill-will, or his indifference). When an agent manifests ill will through his behavior, the relevant behavior can be said to have this meaning. But this is entirely consistent with its also having a meaning that is determined by the overall narrative structure of an agent's life. And my contention is that it is this latter meaning that helps to explain the value we find in exhibiting guidance control (and thus acting so as to be morally responsible).

One question we may have is, "To what feature of behavior do we respond when we evince one of the reactive attitudes?" This feature may be the ill or good will of the agent, as manifested in the behavior. This is certainly one layer of meaning. But a different question might be, "Why exactly do we value the agent's behaving in such a way as to be morally responsible?" The answer, I have suggested, is self-expression of a different sort; more specifically, it is self-expression that depends for its meaning on a narrative structure. It is analogous to artistic self-expression, but not a species of artistic self-expression.

Consider, again, a sculptor who has created a particular sculpture. The critics may write reviews of the work in which they respond to particular aesthetic features of the sculpture. Thus the sculpture has a set of features

relevant to its aesthetic evaluation. It may be said that these features seem to be relevant to the question of what we value in the sculpture. But it is a quite different question to ask what exactly we value in the artist's creative activity. Here I have suggested that it is not necessarily that the sculptor has made a difference to the world; rather, it is that he has engaged in a certain sort of artistic self-expression. Similarly, the good or ill will of the agent as evinced in the relevant action might be the feature of the action pertinent to one's "reactive attitudes"; and yet the value of the agent's acting in such a way as to be fairly held morally responsible derives from a different feature of the action – that it is a certain sort of self-expression.

4. Conclusion

Traditionally, most philosophers have thought that moral responsibility requires alternative possibilities. That is, they have thought that moral responsibility requires a certain kind of control, which I have called, "regulative control." In my view, these philosophers are to some extent driven by an intuitive picture. On this picture, being morally responsible involves making a certain sort of difference to the world. If you make a difference, in this sense, you *select* which path the world will take, among various paths that are genuinely available. Your selection determines which way the world goes, and you thereby make a crucial difference.

But I have argued that it is at least very plausible that moral responsibility for one's behavior does *not* require that one make this sort of difference. The Frankfurt-type cases seem to me to show that one can be morally responsible for one's actions, even though one does not select the path the world will take, among various paths that are genuinely available; in these cases, suitably filled in, there is just one path the world will take. And what makes the agent morally responsible is *how he proceeds along this single path.* More specifically, the agent can exhibit a certain sort of control – guidance control – even though he lacks regulative control. Guidance control, in my view, is the "freedom-relevant" condition sufficient for moral responsibility.

There can be examples in the realm of art which are similar in structure (in certain ways) to the Frankfurt-type examples. In these cases the artist creates a work of art "on his own" and as a result of his own creative energies, and yet the very same kind of work of art would have been produced, had the artist not been inclined to do so. Typically it *is* the case that the artist has changed the world in an important way in producing a work of art; but the artistic analogues of the Frankfurt-type cases show that the artist's activity can have value *without* this being the case.

I have suggested that this value consists in a certain sort of artistic self-expression.

Similarly, I have suggested that it is natural to think of morally responsible behavior as a kind of self-expression. More carefully, what I have argued is that the self-expression picture is what intuitively drives the proponents of the view that guidance control, and not regulative control, exhausts the freedom-relevant component of moral responsibility.

I do not have a knockdown argument that the self-expression picture is superior to the make-a-difference picture, or that the self-expression picture is indeed the correct account of what we value in acting so as to be morally responsible. I hope that the self-expression picture will seem natural and compelling to many open-minded philosophers who are not sure how exactly to respond to the complicated debates concerning the Frankfurt-type cases. If one finds the self-expression model attractive, this can move one toward acceptance of the claim – suggested by the Frankfurt-type cases – that guidance control is all the freedom required for moral responsibility.[33]

Department of Philosophy
University of California
Riverside, CA 92521-0201
USA

[33] I have benefited from reading versions of this paper to the philosophy departments at the University of Rochester, the University of California, Santa Barbara, the University of California, Riverside, the University of California, Davis, Cornell University and Utah State University. Additionally, I read a version of this paper at the Southern California Philosophy Conference at the University of California, Irvine. I am especially grateful to the following for their generous and helpful comments: Ted Sider, David Braun, Richard Feldman, Christopher McMahon, Kevin Falvey, Matthew Hanser, Mark Ravizza, Michael Bratman, Paul Hoffman, Eric Schwitzgebel, and Gideon Yaffe. I am particularly honored to have the opportunity to contribute to this volume in honor of Harry Frankfurt, whose original and elegant work has stimulated and challenged me throughout my career.