Replies

I wish to begin by thanking Carolina Sartorio for the idea to do this volume and for all her extraordinary work in putting it together. I am also deeply grateful to all the contributors for their thoughtful engagement with my work and for their graciousness. Of necessity, I have limited my replies to selected issues; I regret that I have not been able to address many important points raised by the contributors to this volume. I have learned a great deal by thinking about all of the contributions, and I hope to make more adequate progress in addressing the issues in the future.

Reply to Keith Lehrer

I thank Keith Lehrer for his extremely generous, constructive, and suggestive comments. As Lehrer points out, I have claimed that our lives are stories, and I have said, rather too briskly, that when we act freely, we write a sentence in the book of our lives. In later work, I qualified this somewhat, saying that when we act freely, we constrain the plausible stories of our lives. Lehrer correctly points out that at least sometimes I elide the distinction between our lives—the actual, concrete set of events that might be thought to “constitute” our lives (or, perhaps, is the physical basis on which our lives supervene)—and the stories of our lives. As Lehrer puts it:

Simply put, [the problem] is that the story of my life is one thing, and my life is another. My life may consist of my actions. There may be a meaningful narrative constructed out of descriptions and interpretations of those actions by myself or another. On the other hand, those actions may be full of sound and fury signifying nothing as the poet wrote. Whether there is a narrative constructed or not from the actions, there is the life of the person composed of those actions, meaningful or not. That life, consisting of those actions, is distinct from any description, any narrative string of sentences. (1)

---


2Page numbers in parentheses throughout the text refer to pages in this issue.

© Copyright 2011 by Social Theory and Practice, Vol. 37, No. 1 (January 2011)
Lehrer also thinks that the (admittedly important) distinction between our lives (interpreted concretely) and our stories leads to a deep problem for me:

Fischer is left with a chasm between the story of a life and the life the story is about. He switches in his discourse between the story the life is about and the life the story is about. How can the story be the life? The purpose of this paper is to articulate and solve the problem. (1)

Lehrer offers a fascinating and provocative solution, according to which our lives are identified with representations of our actions in terms of our experience of them. These representations become part of our stories, told from the first-person perspective, via a process Lehrer calls “exemplarizing”; this process involves interpretation of our actions through the lens of experience, and a certain sort of storage of the representations in memory. Lehrer argues that insofar as our lives are identified with these exemplars, they can further be identified with our stories—the gap disappears. Also, Lehrer states:

[My] idea is that my experiences can become samples or exemplars that become representational, that become vehicles of content, at the same time that they are part of the content represented. Using what is represented as the vehicle of representation closes the gap between vehicle and content. (6)

In identifying our lives with our stories, and in saying that when we act freely, we are writing sentences in the books of our lives, I was admittedly not being sufficiently careful. I am grateful to Lehrer for emphasizing the distinction between our lives (interpreted as I was interpreting them—concretely, as it were) and our stories. But whereas I agree with Lehrer that there is a gap between our lives (so understood) and the stories associated with them, I do not see exactly why this is a problem. Granted, it is a problem insofar as one wishes to claim (as I sometimes have) that our lives literally are stories and that our free actions literally are sentences in the books of our lives. But I would rather take Lehrer’s advice to be more precise and say that our lives correspond in important ways to stories, and that when we act freely, we are thereby doing something that corresponds to the addition of a sentence to the books of our lives. Although I perhaps enjoyed the somewhat flamboyant mode of expression (which, I hope the reader will agree, is a bit out of character for me), what I really intended was the point expressed in the more careful way. And I do not see why it is a problem for me—apart from my slightly imprecise formulations—that there is a gap between our lives (concretely construed) and the associated stories.

The way I look at it, we move our bodies (and fail to do so) in certain ways, and we thereby interact with the natural world and with other per-
sons. We tell ourselves stories about how this is going; our interpretations are our stories, told from our perspective. But others can also tell our stories from their perspectives. I do not think it is natural to suppose that there is just one "true" or "objectively valid" story of our lives: even from a given perspective (including our own, first-person perspective), there may be many equally valid or legitimate stories. (Later, in my reply to Ben Bradley, I explain how my general framework could accommodate the view that there is indeed a dominant perspective that yields a single "true" story of our lives.) In my view, we invoke stories to explain various phenomena in different contexts, and explanation is a pragmatic, purpose-sensitive activity. For some purposes, a story from the first-person perspective is most useful and illuminating; for other purposes, a story told from a more objective viewpoint is most helpful. So I am not sure why we would wish to privilege, as Lehrer appears to do, the story of my life as filtered through my own experiences.

One could perhaps get rid of the gap between my concrete life and the story of my life on Lehrer's approach; but this would appear to presuppose that the first-person perspective is hegemonic. But it is not obvious to me that there is any hegemonic perspective; further, it is unclear to me that if there is to be a hegemonic perspective from which our stories get told, this will be the first-person perspective. The gap, then, is not a problem for me, and it readily accommodates what I take to be a plausible flexibility and context-dependence in our storytelling.

Lehrer's comments are rich, and there is simply too much in them for me to address here. I wish, however, to highlight a difference between my approach to acting freely and his. Lehrer states:

Fischer and Ravizza\(^3\) have insisted on the importance of being open to the guidance of reasons. They should recognize that whether something is a reason that will guide my actions, that is, a reason for me, depends on my preference for being guided by such reasons. Being guided by reasons in a way that renders me free and not manipulated by a line of reasoning requires that I choose, or prefer to choose, to be guided in my choices and actions by such reasons. It must be up to me, not only that I do what I do, but that I am guided by the reasons I am. Manipulation by reasons will not make me free. More obviously, if the free action is to be an action of self-expression, it must be up to me what reasons guide my actions. My account of that self-expression is that the guidance of reasons results from an ultrapreference. (15)

Lehrer understands an "ultrapreference" (roughly speaking) as a preference that certain reasons drive the hierarchy of preferences that issue in one's actions.\(^4\)

---


\(^4\)For a systematic development of his important views on free will, see Keith Lehrer, "Freedom and the Power of Preference," in Joseph Keim Campbell, Michael O'Rourke,
Although there is much to admire in Lehrer’s theory—and Lehrer has been an important and gracious influence on me throughout my career—I do not see how adding a preference to be moved by certain reasons helps to secure a lack of manipulation. After all, one’s preferences themselves, including this sort of preference, can be subject to all sorts of responsibility-undermining factors, including manipulation. Manipulation by preferences will not make me free any more than manipulation by reasons.

This is a significant problem for Lehrer’s account, and he is not unaware of it. His reply is essentially to invoke a loop similar to the loop he identified in the exemplars that constitute our stories (on his account). Recall that these exemplars are “vehicles of content at the same time that they are part of the content represented.” This sort of loop-like structure is, in Lehrer’s view, a general phenomenon in philosophy; to help us to understand it, he has invoked an analogy with a keystone in a bridge, which both supports other stones and is itself supported by them.\(^5\) Similarly, he supposes here that one’s ultrapreference is a preference that a certain system of reasons drive one’s action and at the same time is supported or underwritten by that very system of reasons.

It is not clear to me that Lehrer can avoid the problem of manipulation by invoking this complex loop-like structure. Indeed, whenever I think of these sorts of self-referential loop-like phenomena, my head begins to spin (appropriately enough). I worry that Lehrer’s move here will not adequately answer the worries. For example, if acting from reasons in itself does not confer autonomy insofar as mere acting from reasons is consistent with being manipulated by those reasons, how exactly does pointing to loops in which certain reasons ground preferences help to rebut the charge that those preferences can also be manipulatively induced? That is, why isn’t the problem just pushed back to the issue of how reasons can underwrite the autonomy of the preferences for acting on certain reasons—the preferences in virtue of which Lehrer supposes that our acting on reasons is autonomous? Perhaps I am trapped in an unduly linear mode of thinking here, but it is just not clear how constructing a more complex self-referential structure solves the problem. At best it seems simply to make it much more difficult decisively to refute the theory.

In contrast, as Lehrer points out, I invoke the notion of “ownership” (in part) to address the problem of manipulation. Although an external agent could manipulate me in a clandestine fashion and thereby induce a reasons-responsive mechanism, it would not be mine. On the Fischer/

---

Replies

Ravizza view, guidance control requires that the relevant behavior issue from an agent's own, suitably reasons-responsive mechanism. I prefer my own, more linear, model, but I fully concede that there are challenges for my approach, one of which is that I rely on an analyzed notion of mechanism-individuation. Both Lehrer and I have sought to address the problem of manipulation. Lehrer deserves credit for honestly confronting this problem, rather than sweeping it under the rug. Perhaps both of our theories have challenges—or, as I prefer to put it, elements that are incomplete. David Lewis once said that we should not expect to eliminate the fog that can obscure our view of the philosophical landscape; sometimes the best we can do is to move the fog from one place to another. I must leave it to the reader to evaluate the usefulness of the different ways in which Lehrer and I have diverted the fog in this area of philosophy.

Reply to Meghan Griffith

I am particularly grateful to Meghan Griffith for her subtle, insightful, and extremely generous contribution. Her explanation of my views about the value of acting freely is beautifully clear and, in many ways, it improves on my own presentation. Having laid out my account of the value of acting freely as a distinctive kind of self-expression, Griffith wonders whether my characterization exhausts what we value about acting freely (22).

Griffith points out that when we consider the meaning of a narrative, we care about the beliefs of the characters, and not just their actions. Further, it makes an important difference to the meaning of the narrative whether the beliefs of the characters are true. If we model the meaning of our lives on such features of stories, Griffith asks,

[s]houldn't we, then, care deeply about our own beliefs when evaluating the meaning of our lives? And won't some of the relevant beliefs specifically concern our own agency? In other words, part of what my story means turns on what I take my role in it to be. For example, my story has a different meaning if I see myself as a passive bystander than if I see myself as fully in control of my own destiny ... If the meaning of my story depends in part on my beliefs about my agency, then it seems that my beliefs about my agency should factor into an account of what it is I value. (23)

We typically believe that we have regulative control—a control that involves genuine freedom to choose and do otherwise—at least at some crucial points in our lives. But I take seriously the possibility that the Consequence Argument is sound, and thus that causal determinism would rule out such control. Thus, if it turns out that causal determinism

---

6Meghan Griffith, "Based on a True Story: Narrative and the Value of Acting Freely," Social Theory and Practice, this issue, pp. 19-34.
obtains, and on the assumption that the Consequence Argument is sound, then a central and significant set of beliefs—core beliefs about our agency—would turn out to be false. This would clearly have important implications for the meaning of our narratives. As Griffith puts it, “a narrative with an accurate self-conception has a different meaning and value from a narrative in which there’s a discrepancy between the character’s self-conception and what is actually true within the world of the story” (24). And, of course, a similar point would apply to our lives. Griffith concludes: “My overall suggestion, then, is that even if we understand self-expression (in its narrative sense) to be the value of acting freely, the appropriate kind of self-expression might not sit comfortably with determinism” (26). Indeed, as she later puts it, “[i]f determinism is true, I never have been and never fully will be who I think I am” (29).

Griffith is willing to entertain the possibility that an agent can see herself as fully morally responsible, even if she knows that there is a “global” Frankfurt-style counterfactual intervener associated with her so that she never has any (sufficiently robust) alternative possibilities throughout her life. This individual will, however, have to change what appears to be a basic and important belief about her agency: she must give up the belief that she has genuine access to alternative possibilities. This, of course, would be the same belief that would have to go, should she be convinced of both the truth of causal determinism and also the soundness of the Consequence Argument.

But Griffith argues that an agent who merely believes that she is subject to global counterfactual intervention (and not causal determination) will still be able to maintain a core belief in her agency; in contrast, Griffith contends that an agent who comes to believe in the truth of causal determinism cannot maintain this belief. This is part of Griffith’s overall argument that my account of the value of self-expression leaves something out—it leaves out the importance of being right about certain core beliefs about our agency. Thus, Griffith contends, my account of self-expression—if self-expression is to have the value we think it must—leaves out an arguably crucial requirement of causal indeterminism.

What exactly is the core belief that an agent can maintain, even in light of the presence of a global Frankfurt-style counterfactual intervener but not causal determinism? Griffith says:

Even if the agent gives up her belief that she had robust alternatives, she most likely can maintain a strong sense that nothing directly made her act in the ways she acted. What, then, does her new self-conception look like? One reasonable suggestion (to paraphrase a
statement from Clarke is that she may see herself as having done things her own way with nothing else having actually determined her to do them this way. (31)

Indeed, according to Griffith, an incompatibilist might even contend that the agent’s former belief in alternative possibilities was “really just her prereflective way of articulating a lack of external determination” (32). Griffith explains:

So perhaps her initial concept appealed to robust alternatives as a more ordinary way of thinking about the absence of external determination. Frankfurt cases, though not as unusual as some thought-experiments, are not something an ordinary agent would typically consider. So perhaps her initial concept appealed to robust alternatives as a more ordinary way of thinking about the absence of external determination. She previously thought, “I did it my way and nothing made me do it. After all, I could have done something else instead.” When presented with a Frankfurt scenario, however, further reflection gives her reason to revise this. Thus, the kind of revision in this case goes from a somewhat imprecisely articulated belief to a belief that is more carefully refined. (32)

Although I agree with much in Griffith’s sophisticated and generous discussion, I think we have got to a point of disagreement here. That is, I agree that it is important that agents not be incorrect about fundamental and core beliefs about themselves, and, in particular, about their agency. I also agree with both Griffith and Clarke that agents tend to have a core belief that involves the notion that in order to act freely and be morally responsible, they must do it their way—and not be “made” to act in certain ways by external factors. But how exactly to articulate and interpret this inchoate notion of doing something in one’s own way is essentially contested, and I do not believe that it is obvious that Griffith and Clarke have latched onto the correct interpretation of the commonsense notion.

The Griffith/Clarke interpretation is one way of capturing and regimenting the inchoate, commonsense view that I must do it my way. But I do not think it is any more plausible than an alternative interpretation, according to which the basic idea is that I did it my way insofar as nothing outside me forced or compelled me to do it. What I care about, and in general what ordinary agents care about, is not being pushed around, coerced, compelled, secretly manipulated, and so forth; insofar as I want to do it my way, I want my behavior to issue from certain processes and not others. But it seems to me a spurious transition—a gratuitous dialectical leap—to suppose that what I care about—insofar as I care about doing it my way—is the lack of causal determination per se. That interpretation is, at best, a highly contentious way of capturing the common-

8Griffith (31 n. 32) quotes Randolph Clarke’s suggestion that “one might ... prefer to be in a position to truly say, ‘I did it my way, and nothing beyond me determined me to do it so’.” Randolph Clarke, “Determinism and Our Self-Conception,” Philosophy and Phenomenological Research 80 (2010): 242-50, p. 246.
sense idea. Thus, I am not convinced that there is a core belief about our agency that must be abandoned by someone who takes it that causal determinism is true but not by someone who is willing to concede that we might be subject to global Frankfurt-style counterfactual intervention.

Let me emphasize that I fully concede that a belief in genuine access to alternative possibilities is a deeply entrenched part of commonsense. Indeed, such a belief is also part of more reflective theorizing about both the forward-looking and backward-looking aspects of human agency. I do not suppose that the deracination of such a belief would be straightforward or easy; it would constitute a major revision in our belief system about our agency. But I think that (perhaps over time) we could come to think of ourselves as living in a causally deterministic world and thus only having one genuinely available path into the future. That is, we could reconfigure our views about both moral responsibility and practical reasoning to allow for robust moral responsibility and also genuine practical reasoning, even in a causally deterministic world (in which it is supposed that such a world would not contain genuinely open alternative possibilities).

As Kant famously put it, we must think of ourselves as “acting under the idea of freedom.” But this need not entail, in my view, that we see ourselves as having genuinely open alternative possibilities. Rather, we must regard ourselves as having the capacity to act freely; that is, we must see ourselves as capable of exhibiting guidance control. In seeing ourselves in this way, we presuppose that we are not caused in certain ways to act as we do. But it is a spurious leap to the conclusion that we must regard ourselves as not being caused (in a deterministic fashion) to act as we do. After all, if you suggest that we go to the movies tonight, and I agree to go as a result of considering your suggestion, and it turns out that the world is causally deterministic, then my behavior would be causally determined by something external to me; but it is odd to suggest that this is an objectionable scenario, or that I care deeply about not being influenced by you in this sort of way.

Reply to Michael Nelson

I have contended that our freedom and status as morally responsible agents—as deeply different from mere animals—should not “hang on a thread.” That is, I have argued that these deeply important features of us should not depend on whether the laws of nature have associated with them objective probabilities of 100 percent or (say) 99 percent. That kind

---

of difference should not make a difference as to such important and central facts about us. I cannot imagine its being appropriate to change my fundamental view of my own and others’ agency and accountability, in the hypothetical circumstance in which a consortium of theoretical physicists fiddle with their equations (differential equations, presumably) and get the result that the laws of nature have 100 percent probabilities. As Nelson points out, I do not suppose that our fundamental natures as free and morally responsible agents cannot depend on any empirical claim; my contention here is rather that they should not depend on this sort of empirical claim.

So, for example, I am certainly willing to agree with Peter van Inwagen in saying that we would give up our view of ourselves as free and morally responsible, if we were to discover that Martians had implanted tiny chips in our brains at birth and have been continuously manipulating us through clandestine, but nevertheless potent, electronic stimulation. But it seems to me that there is a difference between this sort of empirical discovery and the possible discovery, by the scientists, that the laws of nature have 100 percent probabilities. Nelson insightfully presses for a principled way of explaining the putative difference, and I agree that a full defense of my view would require such an explanation.

Nelson goes on to discuss my critique of van Inwagen for potentially engaging in “metaphysical flip-flopping” (39). Van Inwagen accepts various metaphysical elements of the Consequence Argument, including principles that capture the fixity of the past, the fixity of the natural laws, and the “Transfer of Powerlessness” principle (van Inwagen’s Principle Beta). Indeed, he argues on a priori grounds for their plausibility. But he also says that if (per improbabile) he were convinced by the theoretical physicists of the truth of causal determinism, he would not give up his view that we are (sometimes) free and morally responsible; rather, he would give up one of the elements that go into the Consequence Argument. He identifies the Transfer of Powerlessness Principle as the least unassailable such ingredient.

I worry that one’s fundamental metaphysical assumptions should not be voted off the island on the basis of some new empirical discovery (such as that causal determinism is true). Doesn’t it just seem really weird to conclude that, say, the past is not fixed, having just discovered that the laws of nature have associated with them 100 percent probabilities rather than mere 99 percent probabilities? That is, isn’t it incredibly odd to suppose that if the laws have 99 percent probabilities, the past is fixed, but if they have 100 percent probabilities, the past is not fixed? Also, wouldn’t it be similarly weird to conclude that the fixity of the laws of nature depend similarly on whether causal determinism obtains? Finally, wouldn’t it appear to be equally jarring to conclude that the Transfer Principle is
false, having discovered that the probabilities of the laws are indeed 100 percent? What is the connection between the discovery in question and the jettisoning of independently plausible metaphysical principles, argued for on a priori grounds?

Lamentably, I must confess that things are not so simple here. Intellectual history is littered with instances in which “first principles” previously thought to be true and indeed known to be true a priori must be rejected due to the march of science. And there might be chains of interlocking claims that imply that some “first principle” must be rejected, given some new empirical data. But still, despite these complexities (and more), I cling to the admittedly inchoate idea that there is something dialectically infelicitous about jettisoning (say) the Transfer Principle, in light of the discovery of the truth of causal determinism.

To say a bit more, I think we have reflective judgments about “dependencies”—about what should depend on what. And I do not think that the acceptability of the principles of the Fixity of the Past or the Fixity of the Natural Laws should depend on whether the laws of nature have probabilities of 100 percent or 99 percent. In my view, the same is true with respect to the Transfer of Powerlessness Principle.

In a fascinating discussion, Nelson suggests a general way of understanding why the incompatibilist position appears to be problematic here. He says:

If we are justified in believing \( p \) on grounds \( G \) and we have reason to believe that \( p \) and \( q \) are incompatible, where \( q \) is some empirical claim, then there must be an account of how the empirical evidence supporting \( q \) undermines either \( G \) (and thereby the support for \( p \)) or the evidence in favor of the incompatibility of \( p \) and \( q \). (38)

He goes on to say:

Standard incompatibilist accounts lack an adequate account of the interaction between possible empirical evidence in support of the truth of determinism and our intuitive grounds in favor of our own freedom and moral responsibility. Standard incompatibilists cannot account for the way in which justified belief in our ordinary self-conception is resilient to certain possible empirical discoveries. (38)

Nelson is nicely capturing the problem with jettisoning a principle like the Fixity of the Past or the Transfer Principle on the basis of the discovery that causal determinism is true. To put the point a bit harshly: it seems like this move is rather blatantly ad hoc.

Nelson, however, invokes his suggestion above to defend van Inwagen (as opposed to the “standard incompatibilist”) against the charge of dialectical infelicity; nevertheless, he still thinks van Inwagen’s overall position is untenable, because of putative counterexamples to the Transfer Principle (40). Nelson concludes:
Nonskeptical incompatibilists face a problem. They must offer an argument for incompatibilism that is both independently compelling and meets the “what if we discover determinism is true?” challenge in an adequate way. (40)

Of course, I agree with Nelson about the general challenge for incompatibilists. I wish, however, to dispute his claim that van Inwagen himself is not guilty of any sort of dialectical infelicity. Nelson attributes his defense of van Inwagen (as regards dialectical felicity) in part to Keith DeRose (39 n. 8), so I’ll call it the “Nelson/DeRose” strategy. According to their strategy, we invoke Nelson’s suggestion (quoted above) about the relationship between empirical evidence and the grounds for believing various interrelated propositions. Now van Inwagen imagines that God speaks to him and reveals the truth of causal determinism. Nelson says:

Absent God’s speech, van Inwagen takes himself to be justified in his incompatibilism. Given God’s speech, van Inwagen abandons that incompatibilism. Calling this a flip-flop suggests that the switch is unprincipled and arbitrary. But, within the system of claims made in *An Essay on Free Will*, the switch is principled. Van Inwagen claims that the evidence that he sometimes acts freely—which, we can stipulate, the imagined support for the truth of determinism leaves untouched—together with this new evidence that determinism is true provides the material for compelling counterexamples to the Transfer Principle ... This principle forms a crucial premise in the Consequence Argument for the incompatibility of determinism and freedom. Absent reasons to believe that determinism is true, there are no counterexamples to this principle and so the support offered by the Consequence Argument stands. (39)

But the Nelson/DeRose strategy to defend van Inwagen against dialectical infelicity (in virtue of being subject to objectionable flip-flopping), despite its manifest ingenuity, seems to me to be so strong as to entail unacceptable results. In my view, the Consequence Argument can be formulated using various different ingredients; indeed, it can be formulated without the Transfer Principle. One such formulation uses the Basic Principle of the Fixity of the Past and Laws, according to which one’s freedom is the power to add to the given past, holding fixed the laws of nature. This is an incredibly attractive and natural principle. I don’t think one should be able to jettison it simply because one has discovered that the laws of nature have 100 percent probabilities rather than 99 percent probabilities. Indeed, there is something deeply problematic about giving up the Basic Principle on the basis of being convinced that causal determinism is true; again, that seems objectionably ad hoc. It feels like an act of philosophical desperation.

But unless I am misunderstanding it, the Nelson/DeRose strategy

---

would at least appear to allow one to do just that. One could reason that we have perfectly good evidence that we are free, which is not touched by the support for the truth of causal determinism. Thus, given the truth of causal determinism and holding fixed that we are free, we can give up a crucial premise in the argument for the incompatibility of determinism and freedom: the Basic Principle. I see no difference between the argument here and the specific implementation of the strategy Nelson attributes to van Inwagen (in which the Transfer Principle is abandoned so cruelly). But this seems obviously problematic; after all, the Basic Principle is deeply appealing, and its appeal does not in any way seem to depend on the evidence that supports causal determinism. I would actually say the same thing about jettisoning the Transfer Principle; but it is even clearer to me that pushing the Basic Principle overboard because one has discovered that causal determinism is true is dialectically worrisome. Insofar as the Nelson/DeRose strategy permits these moves, it is, in my view, problematic. Indeed, it seems to me that it is just a formal way of describing the sort of philosophical desperation I pointed to above; to put the point in an undoubtedly tendentious and unfair fashion, it is a (slightly) formal regimentation of the lengths one can be forced to go to to save one’s views!

What has gone wrong here—in my opinion—is that one is allowed to “hold fixed” the view that we are free, even when taking on board the truth of causal determinism. Then, of course, one must give up one of the principles that lead to incompatibilism; what could be more obvious? But that one has to do it doesn’t make it OK. Some rabbis seek to make any chicken kosher, but a better approach would be to note that it cannot be assumed that it is obvious that we are free, once the thesis of causal determinism is accepted as true; we do not necessarily reject our freedom in such a circumstance, but we “bracket” the assumption of our freedom, for the sake of the argument. Now we consider various principles and assess their independent plausibility. Such principles might include the Transfer of Powerlessness Principle, the Basic Principle of the Fixity of the Past and Laws, and so forth; at this point in the dialectic, one evaluates the principles, without any prior commitment to the existence of free will. An appropriate target audience for the assessment of the principles might be a group of fair-minded and reasonable agnostics about the compatibility of causal determinism and free will. This is indeed the sort of target audience van Inwagen thinks is most appropriate for philosophical argumentation. It is certainly not implausible that a group of fair-minded and reasonable agnostics about compatibilism and indeed about the existence of free will might well find such principles as the Transfer Principle and the Basic Principle highly attractive.

I have the space only briefly to address Nelson’s second main thesis.
(about my narrative conception of freedom). Nelson denies that the value of life is not merely a function of the values of its segments; rather, he contends that the nature and value of the segments need to be properly specified, and once we do this, there is no challenge to an additive conception of value. So, for example, he claims that a lesson learned from a mistake is a different lesson from a lesson learned from (say) reading a book or “by magic,” even if in some sense the “content” of the lesson is the same. Further, Nelson denies that it is a necessary condition for moral responsibility that a story of the action can be told; that is, he denies my suggestion that it must be the case that some description of one’s choice and behavior count as part of a genuine “story,” strictly speaking, or a narrative, in order for an agent to be morally responsible. He says:

Suppose that I act in a way that does not “resonate appropriately” with my past and is fairly inept storytelling. It’s a real mystery why I did it and seems to come out of the blue as one reads my life story. Nonetheless, no external forces drove my behavior and no wayward motives hijacked my deliberative processes. My choosing to so act was neither compelled nor the result of manipulation. I chose of my own accord, however arbitrarily and inexplicably ... So my choice “resonating appropriately” with my past is not plausibly a necessary condition for being morally responsible. (44)

Note that in this context my point was not so much about the ineptitude of the agent’s storytelling, but that of the third-party evaluator of the agent. Note, also, that I would insist that we are concerned that no external forces of a certain sort drove my behavior; as in my discussion of Griffith, it is crucial to observe that it would be highly contentious (and, in my view, a gratuitous and spurious leap from the uncontroversial data) to interpret our concern as requiring that nothing external to us causally determine our behavior. Indeed, I have sought to capture our concern to do it our own way by articulating the notion of guidance control; I have argued that we can systematize our reflective judgments about responsibility-undermining factors by requiring that the actual sequence issuing in behavior exhibit a distinctive kind of control: guidance control. Insofar as guidance control is present, the stock responsibility-undermining conditions are absent.

Nelson helpfully points to a certain unfortunate lack of clarity in my presentation of my overall view (which I call a “framework for moral responsibility”). The framework has different layers—contentions at different theoretical levels, as it were. I claim that the freedom-relevant condition on moral responsibility is “guidance-control”; this is part of what might be called a “first-order” account of the phenomenon of moral responsibility (according to which I offer the truth-conditions for moral responsibility attributions). Further, I offer a meta-claim about moral responsibility: a claim about the value of being such as to be morally responsible. Here my claim is that this value is identical to the value of a
certain species of artistic self-expression: when we act freely, we are artists insofar as we make it the case that some of the descriptions of our behavior are narratives, strictly speaking (and not mere chronicles), and also that our lives have an irreducible narrative dimension of value. Nelson’s discussion brings out the importance of distinguishing the different levels.

I would, however, suggest that there is a certain kind of intimate connection between the relevant ideas at the first and second levels. That is, I claim that an agent exhibits guidance control of his behavior if and only if the behavior can be described by a narrative, strictly speaking (and not a mere chronicle). On my view, there is no such thing as being morally responsible for behavior that is truly “arbitrary and inexplicable”; after all, guidance control requires suitable reasons-responsiveness, and one should be able to invoke the reasons to tell the story.

I am not sure exactly what to say about Nelson’s contention that a purely additive notion of value can be defended, once one is sufficiently careful about the individuation of the putatively valuable things (such as lessons). This deserves a more careful treatment. But I am inclined to suggest—and this is only a suggestion at this point—that if Nelson is correct here, my main point could be preserved by reformulating it as a point about the source-dependence of the nature of the valuable thing. So, for example, Nelson claims that the source of the lesson is essential to its nature (and thus value); here my contention would be that only creatures who act freely have value that gets individuated in this way (in part via source-dependence). So, for example, it can only matter whether one’s lesson comes from learning from past mistakes or from (say) magic, if one is a free creature. As Nelson points out, I contend that acting freely is what transforms us into creatures who enact stories, in the strict sense; I could equally say that acting freely is what generates a certain source-dependence of value.

Reply to Ben Bradley

Ben Bradley raises many challenging and subtle problems, some of which I simply must put aside on this occasion. Of course, a fully worked-out theory of value would need to address many of the pressing worries Bradley puts forward. Here I shall pick a few central concerns.

Bradley says:

Narrativism is an answer to the question, “What determines how well a person’s whole life goes for her?” This is a question about prudential value, or personal well-being. Here are two simple principles about well-being: (i) the total well-being contained in an indi-

individual's life is just the sum, for every moment of the life, of the individual's well-being level at that moment ("additivism"); (ii) an individual's well-being level at a moment is determined just by what happens at that moment, not at any other moments ("internalism"). The view I will call "narrativism" either rejects internalism, claiming that the value of a moment can be determined by what happens at other moments, or rejects additivism, claiming that the value of a life cannot be determined by summing the values of the moments. (47-48)

To lay my cards on the table, I reject both additivism and internalism. Further, I adopt a notion of well-being that is not exhausted by experiential features (such as pleasure, pain, and so forth).

I contend that it is in virtue of acting freely that we have an irreducibly narrative dimension of value. Along this dimension of value, it is better to learn from one's mistakes than from reading a book (or magic), it is better to flourish as a result of hard work and one's own conscientious efforts (or perhaps courage) rather than by winning the lottery (or receiving an inheritance), and so forth. Bradley correctly points out that our preferences for various lives over others—based on what might have been thought to be considerations of narrativity—are not generated by acting freely. So, for example, in Bradley's case, in which A's life improves and B's declines but neither acts freely, he says that we would still prefer A's life to B's. Thus Bradley rejects "free will improvementism."

Bradley's critique is useful in part because it encourages me to clarify my view, which I have sometimes stated less carefully or precisely than I should have. I do not intend to defend a general thesis such as free will improvementism. That is, I do not think that whenever we have a preference for a life whose well-being profile exhibits a certain shape—say, an upward tilt—this can be explained by reference to free will (more specifically, acting freely—or, in my terminology, exhibiting guidance control). My thesis is considerably more restricted: I wish to say that we prefer certain lives (or stretches of lives) to others, where the only plausible explanation for such a preference seems to me to be free will (acting freely). More specifically, I want to say that we value the obtaining of certain kinds of relationships between different parts of our lives; for example, we value flourishing as a result of fulfilling our ambition to pursue a certain career more than flourishing as a result simply of winning the lottery, and so forth. These particular preferences (and certain relevantly similar preferences) are captured in the irreducibly narrative dimension, and it is (arguably) in virtue of acting freely that we have this dimension of value. But I do not wish to contend that all shape-based life preferences are based on the view that we act freely in the relevant contexts. So I agree with Bradley's statement that "free action does not seem quite so vital to intuitions concerning narrative value" (50), but I would
still maintain that free will is vital to an important subset of such intuitions. I thank Bradley for forcing me to clarify my position here, even if the view is now a bit less grand.

In Bradley’s terminology, I defend a kind of “interpreter-relative narrativism,” according to which “(i) irreducibly global features of a life ... partly determine its value, and (ii) in at least some cases, whether a life has those features, and which of those features it has, is relative to an interpreter” (49). Bradley finds this view problematic; indeed, he says, “[t]he crucial thing to notice is that while some believe that how well off X is depends on nobody’s attitudes, and others believe how well off X is depends on X’s own attitudes, nobody believes that how well off X is depends on the attitudes of some person other than X” (52). I’m not sure how to interpret Bradley’s claim here, and thus I’m not sure I am disagreeing with him. In any case, I do think that it makes sense to suppose that how well off a particular individual X is might be a matter of disagreement between reasonable persons (perhaps adopting different perspectives, standards of assessment, weights for various dimensions, and so forth). I do not think it is so implausible to suppose that the question of how well off someone is—the degree to which he or she is flourishing or whether the individual is flourishing at all—may have different answers depending on the perspective from which one asks the question and seeks to answer it.

But Bradley disagrees, saying:

[On Fischer’s view] [t]here is no objective fact of the matter about whether my killing someone is wrong, or whether the tax code is good for society. Relative to one interpretation of the victim’s life, killing him is wrong, but relative to another, it isn’t. Relative to one interpretation of how well off the members of society are, the tax code is good for society, but relative to another, it is bad for society. And so on. These conclusions are intolerable. (53)

I’m not sure how any “wide-reflective equilibrium” methodology, to borrow Rawls’s term, or, more broadly, any “coherence approach,” can avoid the sort of problems to which Bradley points—at least at some point or other, as one extends the relevant population outward to include different groups. After all, even in our society, people do disagree about the justice of the tax code (and various other central ethical matters).

I do not, however, think that I need to cling to this interpreter-relative approach. I think that it is plausible, and I don’t know how one would argue (decisively) for an objective or interpreter-invariant approach. But if one could successfully motivate such an approach, so be it: one could then simply “designate” some perspective as the hegemonic or dominant one—the perspective that gives the one “true” interpreter-invariant answer to the question of how one is faring at a given time. Nothing in
what I say in developing my account of the value of acting freely as self-expression, or any of the associated contentions, is inconsistent with taking an objective (nonrelativistic) approach here.

Bradley ends with a worry that narrativism implies what he calls an “implausible dualism of prudence” (62). He says:

If the value of a life is not determined by the sum of the values of its moments, then one might find oneself in a situation in which there is a conflict between two courses of action, A and B, such that A leads to a better future (perhaps because it contains more momentary happiness than the future promised by B), while B results in a better overall life, due to the superior relations between future and past events. In such a situation, which course of action should one choose? (59)

Bradley argues that in such scenarios, the proponent of narrativism is committed to an irresolvable conflict; he finds this kind of “dualism” implausible (as it pertains to prudence).

My own view is that we should accept the possibility of such conflicts, and that although it is indeed an implication of narrativism, it is not a decisive problem. After all, it is plausible that there are many conflicts in practical reasoning—between duty and prudence, and perhaps between different duties, and so forth. Presumably, there are conflicts between long-term and short-term prudence; it is perhaps not so surprising that there can also be conflicts between different dimensions of prudence—the additive and narrative dimensions.

I believe that it is actually an advantage of narrativism that it highlights a certain sort of mistake by proponents of what I take to be an unduly narrow picture of rationality. Economists and various theorists of practical reasoning propose that we ignore “sunk costs”; after all, there’s no point in crying over spilt milk, and what’s over is over. More specifically (and less metaphorically), the view is that rationality requires us just to focus on costs and benefits going forward. The idea behind the exhortation to ignore sunk costs is precisely the idea that as between lives A and B, one should choose A: in A one’s life goes better from here on out, as it were. But the notion that rationality unequivocally mandates A, and that there is nothing to be said from the perspective of rationality on behalf of B, is problematic. It is a virtue of narrativism that it brings out the poverty of an unduly restrictive approach to rationality—the economist’s conception, according to which we ignore sunk costs.

This is an important point. I believe that we care deeply not just about how our lives go from here, but also about our overall life stories (our narratives). Sometimes we can change the “meaning” or value of previous segments of our lives in virtue of what we do now (or in the future), and sometimes, at least, we will care more about our overall narrative than about how our lives go from here on out. Of course, there is no...
algorithm that decides the issue; as above, I am willing to concede that there can be irresolvable conflicts (irresolvable by an algorithm, that is) between the various dimensions of prudence. But it is important to keep in mind that the economist's picture of rationality is not the only viable model: it corresponds to only one of the various irreducible dimensions of rationality. Narrativism also helps us to see that death can rob an individual not just of experiences he would have had in the future, but of the possibility to change the meaning of past segments of his life—to rewrite his story. Death thus can rob an individual of his past, as well as his future.

**Reply to Peter A. Graham**

I have argued that the intuitive reasons why the Frankfurt cases show the falsity of the Principle of Alternate Possibilities (PAP: an individual is morally responsible for his action only if he could have done otherwise) should also incline one to reject the Ought-Implies-Can Maxim (OIC). I have proceeded first by suggesting that if one rejects PAP for actions, one ought to reject it (or a related principle) as regards omissions. Here is a case I provided, which Graham dubs BAD SALLY:

Suppose that by raising her hand, Sally can save a drowning child (by alerting the lifeguard that the child is drowning) ... Sally sees a child drowning and there is no good reason why she should not raise her hand, but she simply does not raise her hand. Imagine, further, that this is a Frankfurt-type omissions case, and unknown to Sally, she is temporarily paralyzed in such a way as to make her unable to raise her hand.

In my view she is morally responsible (and blameworthy) for not raising her hand, even though she could not have raised her hand. (66)

Graham helpfully regiments my construal of the natural motivation for OIC as follows:

i. If OIC is false, then it is possible for there to be cases in which a person ought to φ even though she doesn’t φ and can’t φ.

ii. Necessarily, if a person ought to φ and she doesn’t φ, then she is blameworthy for not-φ-ing.

iii. Therefore, if OIC is false, then it is possible for there to be cases in which a person is blameworthy for not-φ-ing even though she can’t φ.

iv. It is impossible for there to be cases in which a person is blameworthy for not-φ-ing even though she can’t φ.

v. Therefore, it is not the case that OIC is false. (67)

---

12Peter A. Graham, “Fischer on Blameworthiness and ‘Ought’ Implies ‘Can’,” *Social Theory and Practice*, this issue, pp. 63-80.

Given this understanding of the motivation for OIC, Graham regiments my argument against it:

1. Sally is blameworthy for not raising her hand in BAD SALLY.
2. Necessarily, if one is blameworthy for φ-ing, then one ought not to φ.
3. Therefore, Sally ought not to not raise her hand in BAD SALLY.
4. If OIC is true, it is not the case that Sally ought not to not raise her hand in BAD SALLY.
5. Therefore, it is not the case that OIC is true. (69)

Graham points out that my argumentation appears to make use of certain principles bridging blameworthiness. Premise ii in my construal of the natural motivation for OIC appears to make use of

(I) Necessarily, if S ought to φ and S does not φ, then S is blameworthy for not-φ-ing. (69)

And premise 2 of my argument against OIC (as regimented above) is:

(II) Necessarily, if S is blameworthy for φ-ing, then S ought not to φ. (69)

Graham rejects both of these principles. It might be best to explain and evaluate his reasons for rejecting (I) first and then turn to (II). Graham offers two examples in which he thinks there can be “blameless wrongdoing,” the first (but not the second) involving ignorance. The first case is:

COAT: When leaving a party, Lorne takes Josh’s coat, mistaking it (albeit justifiedly) for his own. (70)

And the second case is:

HORRIBLE THREAT: An evil villain tells Alex that he will gruesomely torture and kill his infant daughter if he doesn’t painlessly murder Ben. Alex knows that he can refrain from killing Ben, but, in order to save his daughter, he does so anyway. (70)

In both of these cases, according to Graham, the agents act wrongly—they do what they ought not to do—but they are not blameworthy. Thus, (I) is unacceptable, and insofar as my articulation of the motivation for OIC relies on (I), OIC is (as of now) unmotivated.

I agree with Graham that there can be cases of blameless wrongdoing, some of which do not involve ignorance; I am grateful to Graham for helping us to see the relationship between such cases and the intuitive ideas behind OIC. I think that what is going on in the (nonignorance) counterexamples to (I) is that there is considerable force placed on the agent, such that it would not be reasonable to expect the agent to resist the force under the circumstances. Thus, although I agree with Graham that (I), as it stands, is not acceptable, I would wish to suggest a modifi-
cation of (I) that captures and accommodates what is going on in the problematic cases—and still plays the required role in the explanation of the natural motivation for OIC. One might consider something like:

\[(I^*) \text{Necessarily, if } S \text{ ought to } \phi \text{ and } S \text{ does not } \phi, \text{ and it is not the case that there is considerable pressure for } S \text{ not to } \phi (\text{pressure it would not be reasonable to expect } S \text{ to resist}), \text{ then } S \text{ is blameworthy for not-} \phi\text{-ing.}\]

Graham also provides various cases in which he contends that (II) is false: TORTURE, CURE, TROLLEY, and GANGRENE (70-72). I agree with Graham that these cases provide prima facie reason to reject (II); further, I agree with him that the natural and initially appealing strategies for responding to these cases (and seeking to defend (II)) are problematic. I am uncertain as to how to evaluate the dialectic at this point: there may be ways of resisting Graham's conclusion here. But I am inclined to accept that Graham has provided strong argumentation against (II).

Graham begins his piece by saying: "John Martin Fischer has argued that Frankfurt scenarios can be employed to undermine the deontic principle, 'Ought' Implies 'Can.'" (63). Graham is correct, but it is also true that I did not set out initially to disprove OIC; rather, I thought that I had to reject OIC, given my acceptance of what I take to be the moral of the Frankfurt stories. Graham correctly regiments my worries in presenting what he calls "Fischer's argument against OIC" (69); I thought it was an argument that shows that if one accepts what I want to say about the Frankfurt stories, one must give up OIC. If Graham is correct in his argument against (II), then it would appear that I would not have to give up OIC after all. That is, I could still say what I wish to say about the Frankfurt cases (including BAD SALLY) without having to give up OIC. I always thought that rejecting OIC was a significant (although, in the end, bearable) cost of my views about the Frankfurt cases; thus, I would certainly welcome the possibility of maintaining an actual-sequence approach to moral responsibility without having to give up OIC.

Reply to Pamela Hieronymi

As I said in my reply to Nelson, my overall framework for moral responsibility has different elements at different "levels of analysis." I offer truth-conditions for attributions of moral responsibility in terms of guidance control, where this distinctive kind of control involves mechanism ownership and moderate reasons-responsiveness. But I also seek to give

---

14Pamela Hieronymi, "Making a Difference," Social Theory and Practice, this issue, pp. 81-94.
an account of the value of our being such as to be morally responsible in terms of the value of self-expression.

When we act freely, we do not necessarily make a difference (according to my interpretation of “make a difference”), but we do make a statement. That is, we could be subject to Frankfurt-style counterfactual interventions throughout our lives, and thus it might never be the case that we are able to do otherwise; and yet we can still act freely by displaying guidance control. Similarly, on my view we can act freely in a causally deterministic world (in which, arguably, we do not have freedom to do otherwise). In so acting, I have suggested that we transform the chronicles of our lives into “stories, strictly speaking” or narratives. Thus, in acting freely, we become authors: every free action corresponds to a sentence in the narrative of our lives—we can be said to make a statement, even if we do not make a difference. In acting freely, we sometimes exhibit weakness of the will; in such cases we do not do what we think we ought to do, all things considered, and so we do not “endorse” or “stand behind” our actions. In such cases, we do not, in Sarah Brodie’s nice phrase, “propound into public space” a value we take to be defensible. But we do—to speak metaphorically (and with the caveats in my reply to Lehrer above)—write a sentence in the book of our lives, and we thereby make a statement.

Hieronymi takes issue with my interpretation of “making a difference,” and she believes that, properly interpreted, we can indeed be conceptualized as making a difference through our free actions. She says:

Hieronymi takes issue with my interpretation of “making a difference,” and she believes that, properly interpreted, we can indeed be conceptualized as making a difference through our free actions. She says:

Here I would like to suggest that, despite all the important, above-noted agreement with Fischer, he cedes too much to the Consequence Argument when he grants that we may not make a difference. It seems to me that we plainly do—at least often enough—make a difference and that the arcane ruminations and deliverances of the theoretical physicists will not show otherwise, even if they show that the past together with the laws of nature fix the future. (83)

In a footnote attached to this passage, Hieronymi says:

The general form of my response will be to suggest that nothing could count as exercising regulative control [the kind of control that involves freedom to do otherwise]—I suspect it is a kind of illusion—and that, in any case, it is not required for making a meaningful difference. (83 n. 11)

In setting up the dialectic, Hieronymi expresses considerable puzzlement about the Consequence Argument; she is especially perplexed at why many philosophers (including me) take it seriously at all. She says:

In setting up the dialectic, Hieronymi expresses considerable puzzlement about the Consequence Argument; she is especially perplexed at why many philosophers (including me) take it seriously at all. She says:

The typical incompatibilist … [contends that] [i]f determinism is true, then when I strike down my chosen path, I will not have made a genuine difference, not a real one—indeed, I won’t have really selected an outcome or chosen, at all. (84)
Hieronymi goes on to say that this conclusion—that we have no real control over our actions—"is, right on its face, absurd" (84). She further thinks that a focus on the Consequence Argument can obscure the real nature of the threat to our control, which, she thinks, comes from the "simple claim that each event can be adequately explained by events that precede it" (85). She states, about the threat to our control seemingly posed by causal determinism:

It is worth noting that this basic threat can be generated by weaker and surer claims than the claim that the future is fixed by the past together with deterministic laws of nature. For example, the threat can be generated by what T.M. Scanlon calls "the Causal Thesis," the thesis that "all of our actions have antecedent causes to which they are linked by causal laws of the kind that govern other events in the universe, whether these laws are deterministic or merely probabilistic." (85)

In further explaining the threat in question, it is clear that Hieronymi takes it that the basic threat to our control is mechanism, rather than causal determinism.

Hieronymi presents an account of autonomy, by reference to which she contends that we often do indeed "make a difference" (in what she takes to be the relevant sense). She says: "By autonomy I will mean the ability or capacity to effectively bring about that which you authentically want" (90). She goes on to sketch accounts of both crucial elements: the notion of "having the capacity to effectively bring about" and "authenticity." About the first notion, she says:

To have [the ability to effectively bring things about] is to possess certain ordinary skills, dispositions, and resources. Among these are the basic use of one's limbs (or of some substitute), some amount of knowledge or information about the world, the ability to engage in means-end reasoning, and a capacity for planning. (90)

Hieronymi goes on further to develop this initial account by adding more refined capacities, and so forth (90).

Hieronymi says that the second element—authenticity—is a bit more difficult to articulate, but she says:

Authenticity is, approximately, the ability to be oneself among others, open to their influence, without defining oneself either by or against those others or their expectations, as such ... My characterization of authenticity is here very rough, and it leaves much to be desired. Even so, I think it clear enough that authenticity—the kind of sensitivity to, openness to, care for, and yet independence from one’s social environment that is both its substance and its goal—is not the kind of thing that we have simply in virtue of being a human being, or possessing rational nature or a reflective self-consciousness, or any other progeny of the immaterial soul. Nor, I think, would it be secured by "genuine metaphysical access to alternative possibilities." To be sure, there is some resonance between authenticity and regulative control: the first is the ability to be oneself among others, open to their influence, without defining oneself either by or against those others or their expectations, as such; the second is the ability to bring about changes, where those
changes are intended by you but are not (also) determined by the past and the laws of nature. (91-92)

There is much to admire in Hieronymi’s subtle and provocative piece. She makes a particularly important point when she notes that in Harry Frankfurt’s work on agency, his notion of doing what one “really wants” admits of two interpretations. More specifically, “really wanting” can be interpreted as the notion relevant to giving an account of moral responsibility or the notion relevant to the “thicker” or more substantive notions of authenticity and autonomy. I have independently noted a kind of “mission creep” in the literature on Frankfurt; at first, his notion of “identification” (what one “really wants”) is clearly a minimal notion relevant to the analysis of moral responsibility, but later it shifts subtly to a more robust notion related to autonomy. I agree with Hieronymi that considerable confusion can result from not distinguishing these notions:

It is easy to confuse the two, because we would readily say, of someone who is authentic in her desires and has the skills and resources to effectively act upon them, that she is able to do what she really wants to do. And so she enjoys freedom of a very important sort: she is her own person, self-possessed and effective. And yet, crucially, the freedom and identity gained by growing in authenticity and effectiveness is not the freedom required for moral responsibility. (93)

In addition to the significant areas of agreement (including, in the end, our defenses of broadly similar kinds of compatibilistic approaches), there are some points of disagreement between Hieronymi and me. First, I claim that regulative control involves a dual power, and thus it involves, at least in part, a certain distinctive kind of freedom to do otherwise. Hieronymi appears to deny that she has a grasp on this notion (the freedom that is putatively ruled out by the Consequence Argument), but it seems to me that we can at least point in a helpful direction here. I take it that we all intuitively and pre-theoretically have an inclination to believe that we sometimes have the relevant dual power; further, we also naturally believe that, at least sometimes, when we behave in a certain way, we nevertheless could have done otherwise. These are strong and resilient inclinations; I think we have a grasp on the concept of “power” or “freedom” involved in these common-sense views, even if we cannot fully articulate the concept.

The view I have of the matter here is similar to that of Peter van Inwagen:

I use the term “free will” out of respect for tradition. My use of the term is not meant to imply that I think there is such a “faculty” as “the will.” When I say of a man that he “has free will” I mean that very often, if not always, when he has to choose between two or more mutually incompatible courses of action ... each of these courses of action is such that he can, or is able to, or has it within his power to carry it out.16

Thus, an agent has regulative control if and only if she has free will (on van Inwagen’s interpretation of “free will”). Van Inwagen goes on to say:

“Free will,” then, is to be defined in terms of “can.” But how is “can” to be defined? I am afraid I do not know how to define “can,” any more than I know how to define “law of nature.” Nevertheless, I think that the concept expressed by “can” ... —the concept of the power or ability of an agent to act—is as clear as any philosophically interesting concept is likely to be. In fact, I doubt very much whether there are any simpler or better understood concepts in terms of which this concept be might be explained.17

Van Inwagen goes on to distinguish the relevant notion from other concepts, with which the concept of human power or ability might be confused: moral or legal permissibility, physical possibility, epistemic possibility, causal power or capacity.18 I am in agreement with van Inwagen in distinguishing the concept of human power or freedom—what might be called “free-will possibility” or the “free-will can”—from the other concepts mentioned above. Further, I agree with his view that we have a tolerably clear grasp of the concept of free-will possibility. For example, I think I can grasp the concepts involved in the contentions that I am now free to stop working on this paper and instead go refill my cup of coffee, that you are now free to stop reading and take a break, and so forth.

I believe that we can invoke this inchoate grasp of the concept of free-will possibility (even in the absence of our ability fully to articulate it) to explain the distinction between what are sometimes called “metaphysical” and “normative” approaches to moral responsibility. This distinction is notoriously difficult to characterize; elsewhere, I have explored in a preliminary way various attempts to characterize it.19 Here I simply suggest that what is definitive of a (broadly speaking) metaphysical approach to moral responsibility is the claim that in order to be morally responsible, an agent must possess whatever corresponds to or is “picked out” by the concept of free-will possibility. In contrast, a proponent of a (broadly speaking) normative approach to moral responsibility will deny this claim; he or she will suppose that it is not a necessary condition of

17Ibid., pp. 8-9.
18Ibid., pp. 9-11.
true attributions of moral responsibility that the agent in question have the power or freedom corresponding to the concept of free will-potentiality. Rather, on the normative approach, we can proceed in developing an account of moral responsibility without any sort of inquiry into whether the agent in question has this sort of freedom or power.

Now this suggestion about how to draw the distinction between metaphysical and normative approaches would need to be developed considerably. Note that although my suggestion as an account of the normative approach resonates with Peter Strawson’s famous claim that an agent need not meet any metaphysical condition in order to be legitimately deemed morally responsible, it is a bit more careful in specifying precisely the metaphysical condition in question: the possession of a power corresponding to our (tolerably clear) concept of human power or freedom. Additionally, this way of explaining the distinction between metaphysical and normative approaches to moral responsibility appears to place my semicompatibilism on the normative side, since (for reasons different from those of Peter Strawson) I contend that moral responsibility does not require the possession of the kind of power in question. On the other hand, I do believe that some “metaphysical proposition” must be true of an agent in order for him to be morally responsible: he must exhibit guidance control. The question of how to distinguish “normative” from “metaphysical” conceptions of moral responsibility is indeed vexed.

The assumption that we share a common, inchoate concept of free-will possibility also helps to illuminate some dark issues in the literature on free will. For example, one often reads about the “compatibilist sense of freedom” and the “incompatibilist sense of freedom”; but it is more fruitful and accurate, I would suggest, to suppose that we all share an inchoate notion of free-will possibility, and the disagreements between compatibilists and incompatibilists are about how to articulate this concept more explicitly. (Otherwise, we wouldn’t be disagreeing with each other.) Also, I agree with van Inwagen’s curmudgeonly attitude toward those who use the term “libertarian freedom”; as with van Inwagen, a reader will (I hope!) search in vain for any use (as opposed to mention) of this term in my work. I do not think it is fruitful to conceptualize the debates in terms of different concepts of freedom, one of which is the libertarian concept of freedom; rather, we all grasp (in an admittedly inchoate fashion) the concept of free-will possibility, and the libertarians have a distinctive (and, of course, contentious) view of how to articulate this concept more precisely.

As I pointed out above, Hieronymi states that she will “suggest” that

---

nothing could count as possessing regulative control, and that regulative control is a “kind of illusion.” Given that regulative control simply involves (as a component) the free-will possibility picked out by our ordinary concept, it should at least be possible to conceive of agents having regulative control. Surely, it would be possible for them to have such control (and also guidance control) in a causally indeterministic universe (of a certain kind). But it should also be possible at least to conceive of regulative control being present in a causally deterministic universe; that is, it should be possible to understand what it would be for an agent to have regulative control, even if the universe turned out to be causally deterministic. The Consequence Argument then seeks to show that this possibility that we can understand and entertain as a real possibility is not in fact a real possibility, if causal determinism obtains.

Hieronymi finds it “absurd” that we lack control over our behavior, although she offers some reflections on why many philosophers have been gripped by this absurd conclusion. I will simply point out that the Consequence Argument is a **skeptical argument**. That is, it challenges a central part of our ordinary, commonsense picture of the world, just as Cartesian skepticism challenges our commonsense view that we know certain propositions about external reality. I suppose it is “absurd” to think that I don’t know that there is a computer in front of me right now as I’m typing. But what is perhaps interesting about Cartesian skepticism is that it invokes certain other parts of commonsense to challenge this ordinary view that I do indeed know that there is a computer in front of me. Further, it can be interesting and illuminating to seek to show exactly why such a skeptical argument fails. Some philosophers just dismiss epistemic skepticism, perhaps along Moorean lines; similarly, one could simply dismiss the Consequence Argument and its manifestly “absurd” conclusion. It is striking that indeed many philosophers do just this—they either dismiss it entirely or give it short shrift. The ranks of CADs (Consequence Argument Deniers) include many distinguished and influential philosophers. But I have always thought that it can be helpful to take the Consequence Argument seriously and to seek to defend compatibilism, even giving the Consequence Argument its due.

One reason—but surely not the only reason—why I have taken this approach is that something like the Consequence Argument—perhaps in a different version or guise—has worried philosophers for millennia. To be a bit more specific, philosophers (and others) have considered with great care skeptical arguments about our freedom mounted from the possibility of the prior truth of statements about the future and also from the possibility of an essentially omniscient God’s foreknowledge of the future. Incompatibilists have been animated by worries (of various sorts) about the fixity of the past; and the notion of freedom presupposed by all
of these various arguments is the notion involved in regulative control. And many contemporary philosophers—incompatibilists and compatibilists—continue to debate these issues. I have thus sought to take seriously the threat posed by the Consequence Argument—as well as the parallel argument (say) for the incompatibility of God’s foreknowledge and human freedom to do otherwise. I do not suppose that these arguments are obviously problematic. Further, I do not think the conclusions are uninteresting or unimportant. For instance, if the Consequence Argument is sound, we would need substantially to revise or reconfigure our conceptualization of both the forward-looking (practical reasoning) and backward-looking (moral responsibility) aspects of agency. This is, I suppose, the kernel of truth in Hieronymi’s claim that the conclusion of the Consequence Argument is “absurd.” In the end, I have come to the conclusion that the sort of freedom in question—human freedom to do otherwise—is not indeed required for practical reasoning and moral responsibility. But I wish to emphasize that this conclusion would lead to significant revisions in our ordinary picture of ourselves.

Hieronymi contends that it is not causal determinism, but mechanism, that poses the relevant threat to our agency. One finds similar claims in the work of Hilary Bok21 and (as quoted by Hieronymi (85)) Tim Scanlon. But I have for a while found this view puzzling. Note that one cannot get the Consequence Argument (in any of its myriad formulations) off the ground at all, given mere causation rather than deterministic causation. Intuitively, if we simply assume causation (rather than deterministic causation), an agent could do otherwise without so acting that either the past or the natural laws would have been different: the agent’s doing otherwise is compatible with the actual past and natural laws. So if one’s worries stem from the Consequence Argument, it is clear that the threat comes specifically from the assumption of causal determinism. And it seems to me that the Consequence Argument best captures what should worry anyone who is concerned about agency and its relationship with causation. Note, again, that philosophers have focused specifically on what are essentially versions or perhaps siblings of the Consequence Argument for millennia. That is, they have worried about the fixity of the past and how this might rule out our freedom to do otherwise, given minimal facts about “logic” or facts about God. If our worries are conceptualized as continuous with the millennia of debates about such issues, then it might be best to think of the threat to our agency as coming from the Consequence Argument.

Think of the dialectic this way. Suppose someone comes to our town

---

and announces: “Causal determinism is incompatible with freedom to do otherwise.” Given the highly contentious nature of this claim, we townspeople are likely to ask why. The visitor could simply say: “You should just be able to see that it is,” but this would be unsatisfying. This is where the Consequence Argument is helpful: it points to the fixity of the past and natural laws, which can independently be seen to hold. The visitor thus has an explanation for his assertion in terms of more basic, less contentious claims. In contrast, if another visitor were to arrive and proclaim: “By the way, universal causation itself is incompatible with freedom to do otherwise [apart from the assumption that such causation is deterministic],” we would also be inclined to ask why. But this time I don’t know what the new visitor would say. He cannot advert to the considerations that drive the Consequence Argument, and, obviously, he cannot simply say: “You just should be able to see that mere causation rules out freedom to do otherwise.”

The new stranger can certainly try various strategies here, but analogies with mere machines and the other stock considerations don’t seem to me to go very far. If we are machines, we are very complex machines—with minds. Even if one could explain a bit of behavior by invoking prior physical causes, we would need an argument that this would necessarily crowd out other explanations, including explanations that invoke mental states. In any case, it seems to me that the first visitor is on much firmer ground than the second; this is a reason to interpret the incompatibilist as arguing from causal determination rather than mere causation, especially if one wishes to be charitable to one’s opponent. Although one certainly can find different ideas driving different developments of worries about free will, and, in particular, incompatibilism, it has always seemed to me that the strongest defense of compatibilism will be maximally charitable to the incompatibilist, and I also think that the Consequence Argument provides the most challenging version of incompatibilism.

I do, however, fully agree with Hieronymi that it would be good to identify and give an account of a kind of freedom that need not involve the possession of the distinctive human power or freedom to do otherwise. Whereas I have offered the notion of guidance control, she presents an account of autonomy in terms of the ability or capacity to effectively bring about what one authentically desires. I commend Hieronymi’s suggestive and nuanced account, and I look forward to its further development. She further offers her notion of autonomy as a promising account of what she takes to be the important notion of “making a difference.” Here I am inclined to let a thousand flowers bloom—or, perhaps in a better metaphor, not make too much of a fuss. I defined “making a difference” in terms of having the dual power corresponding to our ordinary
concept of the freedom linked to moral responsibility. According to my view, one makes a difference insofar as one freely selects one from various paths genuinely available to one—paths that one has the free-will power to select. It is clear that Hieronymi’s account of autonomy does not entail that the agent be able to make a difference, in my sense. But I fully concede that there are various legitimate interpretations of the somewhat protean and, in any case, imprecise ordinary notion of “making a difference.” Hieronymi’s suggestion is a fascinating way of capturing this notion of making a difference.

Reply to Neal A. Tognazzini

I am extremely grateful to Tognazzini’s extraordinarily generous and constructive contribution. With the pun completely intended, I consider myself very lucky to have had the opportunity to work with Neal. Since it would seem rather odd to subject his very friendly paper to extensive critical analysis, I’ll just very briefly explore a few small points.

Tognazzini regiments the argument from luck as follows (107). We assume that the worlds in question are indeterministic. I’ll add the numbers for the propositions and put it in a slightly different format; note that each proposition (after 1) is supposed to follow from the previous one:

1. Nothing accounts for the difference between the two worlds in question (i.e., the world in which the relevant agent decides and behaves as he actually does and the world in which he decides and behaves differently).
2. Nothing about the agent accounts for the difference.
3. Nothing about the agent accounts for the decision that is actually made.
4. The decision is just a matter of luck for the agent.
5. The agent is not morally responsible for his decision (and behavior).

Tognazzini employs resources from the Fischer/Ravizza approach to moral responsibility, specifically our account of “taking responsibility” and mechanism ownership, to argue that either (3) does not follow from (2) or (4) does not follow from (3). He says:

If we emphasize the analogy with the drunk driver, then the right response is that something about the agent—namely, the operation of his mechanism—does after all account for the decision that is actually made, even though the very same thing might have accounted for a different outcome. This would allow us to reject the third step of the above argument.

---

argument. If, on the other hand, we emphasize the analogy with the two assassins, then the right response is that although nothing about the agent accounts for the decision that is actually made, that decision can nevertheless be attributed to the agent (and hence it isn’t a matter of luck for him) thanks to his taking responsibility for the mechanism from which it issued. This would allow us to reject the fourth step of the above argument. (107)

As I said above, I greatly appreciate this thoughtful suggestion. I wish to note that, at least in my reading of the literature by proponents of the luck argument, I do not typically find something corresponding to Tognazzini’s (3); rather, the argument seems to proceed from (2) directly to (4) and (5). For example, having described an indeterministic scenario created by his infamous (at least in philosophy) goddess Diana, Alfred Mele says:

Diana suspects that his making that alternative decision rather than deciding in accordance with his best judgment—that is, that difference between $W$ and the actual world—is just a matter of bad luck or, more precisely, of worse luck in $W$ for the agent than in the actual world. After all, because the worlds do not diverge before the agent decides, there is no difference in them to account for the difference in decisions. This suspicion leads Diana to suspect that, in $W$, the agent should not be blamed for the decision he makes there.23

Here Mele appears to go directly from something like (2) to something like (4) and (5). That is, he goes from the point that nothing (about the agent) explains the relevant cross-world difference to a point about moral responsibility. This suggests that Tognazzini’s way of blocking the luck argument (rejecting the move from (2) to (3)) might not work, at least for all proponents of it.

Further, it seems to me that a proponent of the luck argument might concede Tognazzini’s point that even in an indeterministic scenario, there is a sense in which the relevant agent can “own” his decisions and their consequences in virtue of the agent’s having “taken responsibility” (in the Fischer/Ravizza sense) for the mechanism from which they issue. Nevertheless, the proponent of the luck argument might insist that “ownership” in this sense is not enough to rebut or outweigh the worries stemming from (2). If nothing accounts for the cross-world differences, then it might seem that the agent cannot be blamed, even though he has taken responsibility and thus taken ownership of his decision and its sequelae.

Perhaps we have a kind of stalemate here. Unfortunately, I do not have the space to explore these issues in more detail in this venue, but I would point interested readers to my attempt to respond precisely to the worry just adumbrated—the worry that stems from unexplained cross-world differences. In “Indeterminism and Control: The Problem of Luck,”24 I develop a strategy for responding to various versions of the

24 John Martin Fischer, “Indeterminism and Control: The Problem of Luck,” in Mi-
argument from luck, including those that proceed from unexplained cross-world differences. I should also point out that my suggested strategy for assuaging such worries could then supplement Tognazzini’s suggestion. That is, we could employ Tognazzini’s suggestion as a “positive” element that explains how an agent could “own” his decisions (and their consequences), even in a certain sort of indeterministic world; then we could reply to the worries stemming from cross-world differences employing the strategy in “Indeterminism and Control.” Thus, Tognazzini’s suggestion could be part of an overall argument that has some hope of breaking the apparent stalemate sketched above.

Reply to Frederik Kaufman

There are essentially two parts to the Brueckner/Fischer approach to answering Lucretius’s famous Mirror Image Argument (according to which he claims that our ordinary asymmetry in attitudes toward prenatal nonexistence and posthumous nonexistence cannot be sustained). First we seek to cast doubts on other approaches to responding to the argument, including the approach (favored by Kaufman) that contends that the intuitive asymmetry in attitudes can be linked to an asymmetry in possibility: whereas we could not have been born considerably earlier, we can die later than we actually die. That is, Kaufman, following but considerably expanding a suggestion by Thomas Nagel, contends that the ordinary asymmetry identified by Lucretius can indeed be defended by pointing to a metaphysical asymmetry—an asymmetry of possibility. Second, we offer our own “Parfitian” approach to explaining and defending the asymmetry, according to which it is seen to be a specific instance of a more general “bias toward the future”; holding other things fixed, we care especially about the future, and are (relatively) indifferent to the past. (It is important that in evaluating this view, one keep in mind that the preference in question pertains to scenarios in which the past and future experiences are evaluated on their own, as it were, and not insofar as they bring along or are correlated with other differences, such as memories or overall quality of life.)

I should note that it is not essential to the Brueckner/Fischer approach decisively to refute other “contenders”; that is, if other approaches are


also promising as potential defenses of the ordinary asymmetry in attitudes, then so be it. Let another thousand flowers bloom! Our main goal is to offer what we take to be a plausible defense of the asymmetry in question; if there are other reasonable contenders, this is not particularly troubling.

I shall begin my reply to Kaufman’s thoughtful and probing contribution by attempting to explain and defend the Brueckner/Fischer strategy against Kaufman’s worries. As I said above, our main goal has been to provide a promising strategy for resolving the Lucretian puzzle. We have, however, expressed some doubts about the approach that seeks to defend the ordinary asymmetry in terms of an asymmetry of possibility—in any of its specific instances. I shall thus also address some recalcitrant issues pertaining to Kaufman’s strategy of response to the puzzle.

Kaufman correctly points out that sometimes Brueckner and I say that the Parfitian considerations show that the ordinary asymmetry in attitudes toward prenatal and posthumous nonexistence is “rational,” and sometimes we (or perhaps I) say that it is “not irrational.” Kaufman is right here, and I want to try to rectify the situation by stating what I take to be the position. The point is that the asymmetry in question can be shown to have a point—it is not entirely arbitrary. More concretely, the asymmetry can be shown to have a point that connects the asymmetry to long-term survival advantage. There is, then, a clear sense in which the asymmetry can be defended from a rational point of view. Given that people care deeply about their survival—this is not a mere whim or adventitious preference—the asymmetry can be seen to be rationally defensible. I agree with Kaufman that it would not be enough simply to show that this particular asymmetry in attitudes follows from or is a specific instance of something more general; after, all the more general asymmetry might itself be problematic (such as a racial or sexual bias). But the general concern for survival is not similarly problematic—it is considerably more firmly grounded. Some, of course, would even go so far as to say that it is the essence of rationality (although nothing in the Brueckner/Fischer approach requires such a claim).

But Kaufman responds (essentially) as follows:

So Fischer gives an argument ... for the rationality of our asymmetric attitudes: since it is rational to care about our happiness, and creatures with asymmetrical attitudes (thanks to evolution) will be happier than not, the general symmetry is rational.

Lucretius might agree that it is rational to care about our happiness, but he would certainly reject Fischer’s (suppressed) premise that an attitudinal asymmetry promotes our happiness. For Lucretius, the fear of death is not some isolatable aspect of our lives: his point is that fear of death ruins our lives. It causes untold misery and anxiety and foolishness; it is all so unnecessary; since, he argues, it is irrational to fear death ... Even if a specific attitude asymmetry about birth and death is situated in a more general attitudinal asymmetry about time, according to Lucretius our fear of death is such a powerful con-
Replies

sideration that it completely overwhelms whatever happiness a general attitudinal asymmetry might foster. Because of our dread of death, we lead miserable lives. (119-20)

Note, first, that in the passage of Our Stories to which Kaufman is referring, I say:

I hope it will emerge, at least in a preliminary way, that there can be a rationale or non-arbitrary point to the Parfitian general asymmetry of which our asymmetric attitudes toward early death and late birth are a specific instance. Insofar as it is rational to care about pleasure, happiness, and even survival, the general asymmetry is, arguably, rational.27

Again, I wasn’t ideally careful or precise here, but I intended to be making a point primarily about survival advantage, and not so much simply about happiness. In any case, neither I nor Anthony Brueckner would wish to defend an excessive fear of death or a preoccupation with death, as is suggested by Kaufman’s phrase, “our dread of death.” A dread of death can certainly be pathological; and even if the pathology is widespread, it needn’t follow from a mere—and, in our view, appropriate—recognition that one’s death can be a bad thing for one. If someone “leads a miserable life” because of attitudes toward death, this is obviously a pathology; the individual should seek psychotherapy or other sources of consolation. Whereas indeed Epicurus and Lucretius thought of philosophy as providing just such therapy, I would insist on a division of labor here. Philosophy aims at the truth; psychotherapy aims at happiness. The truth is that our own deaths can be a bad thing for us. But, as above, this need not entail that we focus unduly or excessively on death; dread is a pathological add-on.

Further, my suggestion is that there is a clear survival advantage to being focused primarily on the future. Creatures who cared equally about past and future experiences would surely not be sufficiently attentive to looming dangers or prospective opportunities to thrive in the long term. My suggestion is not specifically that having a fear of death, or regarding death as a bad thing, provides a survival advantage; rather, I claim that having a bias toward the future does indeed carry with it a significant survival advantage. Again, insofar as what enhances one’s prospects for survival over time is to that extent rational, the ordinary asymmetry in our attitudes can be defended as rational.

Kaufman states:

Most importantly, Lucretius would maintain against Fischer that asymmetric attitudes about the times of our birth and death are more than merely not irrational—they are positively irrational, irrespective of whether they make us happy. Here’s why: the time before our birth is similar to the time after our death. Since rationality requires us to treat likes

alike, we should have symmetrical attitudes about these two periods of nonexistence. Since it is preposterous to become horrified at the thought of being born when we were, we should instead give up our horror of dying when we will ... Fischer has not shown that our asymmetrical attitudes are rational ... nor has Fischer secured his more modest claim that our asymmetrical attitudes are at least not irrational, since it is irrational to treat equivalent things differently, as Lucretius claims about the two periods of nonexistence before and after our lives. (120)

Kaufman announces that the time before our birth and the time after our death are “similar,” that we should “treat likes alike,” and that “it is irrational to treat equivalent things differently.” But of course the whole question is whether prenatal and posthumous nonexistence are indeed relevantly similar or “equivalent.” Perhaps they are relevantly similar or equivalent from a God’s-eye perspective—not situated at any particular point in time. But that’s not the question; the question is whether, from a situated temporal perspective, the two periods are relevantly similar or equivalent. Manifestly they are not: prenatal nonexistence is a deprivation of past life and experiences, whereas posthumous nonexistence is a deprivation of future life and experiences.

Additionally, Kaufman points out that even if the ordinary asymmetry in our attitudes toward prenatal and posthumous nonexistence can in general be defended, it still would be irrational to have such an asymmetry in particular cases in which having the asymmetry in question would not in fact promote survival. He says:

So when it [the asymmetry in question] does not do that [promote happiness or survival], and since it is rational to be concerned about one’s happiness, it would be irrational to hold the asymmetry, even if we are unable to give it up. The difficult of fine-tuning a mechanism that (we assume) ordinarily does promote our happiness simply means that we will sometimes find ourselves holding our attitudinal asymmetry even when we should not—and that looks irrational. Maybe with genetic engineering we could fine-tune a mechanism bequeathed to us by evolution, but maybe not. And, in any case, our inability to abandon an attitude is not a reason to think it is rational to hold it. (121)

But if there is a point—a rational justification—for the general asymmetry, and one can’t fine-tune it in the way envisaged, it would seem that having it in a particular instance is justifiable. As Bernard Williams pointed out many years ago, it might be puzzling to get a bill from the gas company for some tiny amount—an amount less than the cost of the postage for the bill. But the puzzle dissipates when one considers that there is a clear point to sending the bills in general, and it would be excessively expensive to hire someone (this was before computers) to monitor all the bills to make sure that they are for amounts that make sense. (We could imagine many up-to-date versions of this sort of example.) Just as the gas company could defend itself for sending such a bill, an asymmetry of attitudes in a particular case in which it doesn’t have the purpose that is the
I now wish to turn to Kaufman’s defense of his alternative approach to resolving the Lucretian puzzle. He had previously distinguished between “thick” and “thin” selves and contended that it is only the thick self that is relevant to the Lucretian puzzle (and our ordinary asymmetry in attitudes). Brueckner and I have replied that people do have preferences with regard to their thin selves, and that there is no relevant asymmetry of possibility with regard to the thick self. (Dan Speak and I conceded the point that the thick self could not be deprived by the actual time of birth; we meant it as a concession for the sake of argument, as we wanted to focus on the relevance of preferences about thin selves. Here I wish to defend the Brueckner/Fischer line that it is coherent to suppose that the thick self came into existence considerably earlier.)

Individuals sometimes express preferences that can be interpreted as involving a desire that their thin selves have been associated with a different thick self. The minimal point here is that we often do care about such issues—issues involving a given thin self under different hypothetical circumstances. That is, it is just not true that we never care about our thin selves. Relatedly, one can wonder about whether it would be desirable for one’s thin self to have existed considerably earlier; this would entail that one’s “biographical life” have started earlier. It is at least coherent to suppose that one might prefer that, even if such a possibility is not widely attractive.

As Kaufman points out, Dostoyevsky’s Underground Man said he preferred that two plus two not equal four; it clearly does not follow from this sort of profession of a desire that the putatively desired state of affairs is coherent. But we do have a clear, independent way of showing that it is impossible that two plus two does not equal four. It is much more contentious whether it is coherent that a given thin self could have come into existence much earlier; similarly, it is contentious whether such a thin self could have associated the same thick self. It is going to be much more difficult to show that it is incoherent to suppose that the same thick self came into being much earlier than it is to show that it is incoherent to suppose that two plus two equals four!

And I simply do not see how it is incoherent to suppose that the thick self could have come into existence considerably earlier than it actually does; that is, I do not see how the state of affairs is metaphysically impossible. I concede that envisaged, its occurrence would be extremely unlikely. Thus, the possible worlds in which such a state of affairs obtained would be remote in logical space; to accommodate everything that

---

would be needed for any particular thick self to have come into existence significantly earlier than it actually does, significant departures from the reality of the actual world would be required. But that a possible world is remote in logical space does not necessarily denigrate its status: metaphysical possibility is metaphysical possibility.

Of course, there are some ways of interpreting the notion of a thick self such that it would indeed be metaphysically impossible for a thick self to have existed earlier. Kaufman supposes that the possibility of a thick self coming into being considerably earlier would “involve things like my meeting my wife before she was born” (124). But of course the possibility of a thick self coming into being considerably earlier need involve no such thing; it might involve all events (understood perhaps as “event-types”) being “pushed earlier,” as it were.

Similarly, Kaufman might be thinking that the truth of certain counterfactuals entails the impossibility of a thick self coming into existence considerably earlier than it actually does. For example, he says: “Were my metaphysical essence (if I have one) to exist substantially earlier, the thick biographical self associated with that entity would not be thick me” (125). But the truth of this sort of counterfactual is entirely consistent with the metaphysical possibility that thick Kaufman could have come into existence considerably earlier. The counterfactual is a different modality from the possibility claim; whereas the counterfactual points us to the closest worlds in logical space, the possibility claim can be underwritten even by remote worlds. The distinction between the two sorts of modality plays a very important role in certain debates about free will.\footnote{See Fischer, The Metaphysics of Free Will, pp. 87-110.} I will have occasion to come back to these issues below in my reply to Elizabeth Harman.

Finally, I note that we often focus our attention, when we have aspirations, hopes, dreams, and fears about the future, on relatively remote possibilities. Similarly, we often regret that even very unlikely possibilities did not materialize. Our hopes and fears do not seek out the local. So, for example, I can hope that my lottery ticket wins, or that I get an offer of an endowed chair at Harvard; and I can regret that my favorite college football team did not win the national championship last year, or that my daughter did not earn a perfect score on her SAT last week, and so forth. Given that it is metaphysically possible that my thick self could have come into existence considerably earlier, and given that we often focus our hopes, regrets, and fears on remote possibilities, it seems to me that the Lucretian puzzle may well not be adequately answered by Kaufman. Of course, as above, Brueckner and I had as our primary goal putting forward a promising alternative strategy for resolving the puzzle; if some other approaches are also suggestive, this is not in the end objectionable to us.
Reply to Elizabeth Harman

Harman brings up many issues discussed above in Kaufman’s paper and my reply. I will thus not spend as much time here on the arguments in common with those of Kaufman. But she raises some related issues, and I will also explore some of the issues I touched on above in a bit more detail.

Harman says:

The Puzzle is this:

Lucretius’s Puzzle (Central Version): Why is it that we typically wish our deaths would be later rather than earlier (later than they actually will be), but typically we lack a wish that our creations be earlier rather than later (earlier than they actually were)?

I will argue that there is no puzzle here at all. While our wish that our deaths be later rather than earlier is straightforward, easily understandable, and reasonable, it would be very odd for us to wish that our creations were earlier rather than later; this explains why we have the first wish but lack the second. (132)

Why exactly would it be “very odd” for us to wish that our creations were earlier rather than later, whereas it would not be similarly odd for us to wish for a later death? Harman here invokes an asymmetry in counterfactuals (just as Kaufman did):

A person typically wishes that her death was later rather than earlier because if her death was later, she would live a longer life.

What would things have been like if one had been created earlier? If we imagine a significantly earlier creation (at least one year earlier, say), then this is a scenario in which one would have had a completely different life. All of one’s life circumstances would have been different, all of one’s experiences would have been different, and all of one’s relationships would have been different in character (and most with different people). (132)

Not surprisingly, I reply in exactly the way I replied to Kaufman. I can concede the truth of the various counterfactuals. But the counterfactuals do not entail the relevant metaphysical impossibility claims. The counterfactuals are underwritten by features of the closest possible worlds (possible worlds with minimal departures from the actual world to accommodate the truth of their antecedents). In contrast, the metaphysical impossibility claim can be falsified by relatively remote possible worlds.

Harman goes on to say:

[N]ote that it is not the case that had one been created earlier, one would have lived a longer life. It is not the case that had one been created earlier, one would still have died at the same time. (Indeed, it would be very weirdly coincidental if this counterfactual were

---

true of a particular person. So the motivation that appears to underlie the wish to die later—that one would thereby have a longer life—could not motivate a wish to be created earlier—because it is not typically true that if one had been created earlier, one would have had a longer life. (132-33)

I agree with Harman here, and she correctly attributes to Brueckner and me an interest in a "narrowed" version of the puzzle:

Lucretius's Puzzle (Central Version, Narrowed): Why is it that we typically wish to have died later than when we will actually die (holding fixed the times of our creations), but typically we lack a wish to have been created earlier rather than when we actually were created (holding fixed the times of our deaths)? (133)

About this version, Harman says: "I think the suggestion that this is an interesting puzzle is odd (though many besides Fischer have taken it seriously)" (133). She explains:

Consideration of the counterfactual possibility in which a person dies later just is consideration of a counterfactual possibility in which she is nevertheless created at the same time—holding fixed the time of her creation requires no further stipulation, it follows straightforwardly from imagining the closest possibility in which she dies later. But consideration of the counterfactual possibility in which a person is created earlier is not consideration of a counterfactual possibility in which she nevertheless dies at the time of her actual death. As I have commented, it would be very coincidental if a person was such that, had she been created earlier, she would nevertheless have died at the same time. (Such a case is possible but unusual.) So there is nothing parallel in the two wishes being considered by the "puzzle." The first wish is a wish that death be later, with no further stipulations; that creation be held fixed is not stipulated but simply falls out of consideration of that possibility. The second wish is a wish that creation be earlier, with a big—and odd—stipulation, that death nevertheless be at its actual time. (133-34)

Harman goes on to say:

[It is no mystery why people in fact have not typically formed the wish to have been created earlier, but to have died at their actual time of death, because it is such an odd scenario to contemplate. Why would people consider the possibility in which they are created earlier but nevertheless die at the times of their actual deaths? This is not a possibility that people would naturally consider, so it is not at all surprising that they have not formed a wish for this scenario.

It might be responded that what I have said so far leaves open that once people contemplate this scenario, they ought to wish for it. (134)

Actually, I think that most ordinary people go through their lives never even entertaining the thought that they might have been born earlier. So the Lucretian puzzle cannot be about what attitudes people actually have (in an occurrent sense), prior to reflection; the point is that, upon prompting and suitable reflection, most people would report having the "ordinary" asymmetry: most people would be indifferent to prenatal nonexistence, but care deeply about posthumous nonexistence.

Further, I think I can agree with Harman's point about the "oddness"
of the question asked in the Narrowed Version of the Central Puzzle; indeed, reflection on the puzzles and associated ideas can make one feel like Alice, who, having fallen down the rabbit-hole, exclaimed, “Curiouser and curiouser ...” But oddness does not entail impossibility (as Harman herself notes). The Brueckner/Fischer view is that there is nothing incoherent about the scenarios one is asked to envision in the Narrowed Version of the puzzle; thus, presumably, the scenarios in question are metaphysically possible. Further, our point is that human hopes, dreams, and fears often seek out remote possibilities; our mental states are not required to shop local, as it were. We contend that it is at least open to an individual to wish to have been born considerably earlier (holding fixed one’s actual time of death); that is, the content of this sort of wish is not somehow contradictory. Note that we do not say that an agent ought to have such a wish, or that he ought to if he has another suitably related wish (as in the principle discussed by Harman); we simply say that such a wish is open to an agent.

As above, I wish to emphasize that the main goal of the Brueckner/Fischer strategy is to propose a solution to the Lucretian Puzzle. We have also cast some doubts on alternative proposals, but it is not so important to us to refute other purported solutions. After all, it may be that there is more than one way to skin the cat. Further, we appreciate the contributions by both Kaufman and Harman; they both helpfully indicate the extreme “oddness” of the supposition that we—with our thick selves intact—could have come into being considerably earlier. Inspired by the insights of Kaufman and Harman, perhaps it could be argued that although our hopes, aspirations, regrets, and fears are not restricted to the closest worlds, there are constraints on just how remote their targets can be.31

John Martin Fischer
Department of Philosophy
University of California, Riverside
john.fischer@ucr.edu

31There are additional ruminations on the appropriate interpretation of the Lucretian Puzzle, as well as strategies of response, in J. David Velleman, “Comments on John Martin Fischer’s Our Stories,” Philosophical Studies, forthcoming; and John Martin Fischer, “Reply to David Velleman,” Philosophical Studies, forthcoming.