

## The Trouble with Tracing

MANUEL VARGAS

In this paper I argue that the widely relied upon idea of “tracing” in the theory of moral responsibility is considerably more problematic than has been previously acknowledged. The difficulty I raise stems from requirements imposed by the knowledge condition on moral responsibility.

Suppose you believed that being a responsible agent (at least paradigmatically) involves being suitably sensitive to situation-relevant moral concerns. On this view, agents that are not suitably sensitive to the relevant moral concerns (these may include non-human animals, young children, and presumably some adults with particular conditions) are not appropriate candidates for ascriptions of moral responsibility. Though this sort of theory may have the sheen of plausibility to it, it is not difficult to generate problem cases for the theory as stated. For instance, suppose that while he is intoxicated Luis decides to drive home while his intoxication has rendered him insensitive to the relevant moral concerns. When he runs his Buick over a young mother and two kids on the way home from the bar, we do not thereby conclude that because he was in a state of intoxication sufficient to render him insensitive to the relevant moral concerns he was not morally responsible.<sup>1</sup>

The standard response to problem cases of this sort is to introduce a notion of *tracing*. Tracing is the idea that responsibility for some outcome need not be anchored in the agent or agent’s action at the moment immediately prior to outcome, but rather at some suitable time prior to the moment of deliberation or action. So one thing that we might say is that Luis’ responsibility for running over

1. Though I trust this is obvious enough for present purposes, compare Joshua Knobe, “Intentional Action in Folk Psychology: An Experimental Investigation,” *Philosophical Psychology* 16 (2003), especially the discussion of the third experiment.

the mother and two kids is anchored in his earlier decision to drink in circumstances where he might feel the temptation to drive while drunk. Unless there is some special reason for thinking otherwise, we can safely assume that at the earlier time Luis did satisfy the conditions for the appropriate degree of sensitivity to moral considerations—he simply didn't act accordingly.<sup>2</sup>

One of the nice features about tracing is that it is one of a few things to which nearly all parties in the debate about free will appeal to with equal enthusiasm. Nothing about tracing presumes the truth of one or another view about the compatibility of moral responsibility on the one hand and determinism or the natural causal order on the other.

Real-world assessments of responsibility can, of course, be considerably more complicated than the case of Luis. Depending on one's view of addiction, whether the driver was addicted to alcohol might be relevant, even if only to determine whether there are mitigating factors relevant to the assignment of blame. And, we might think it matters whether he was free and responsible for having acquired his addiction. But even these considerations seem to invite the possibility of tracing. If we learned that Luis (however improbably) was intoxicated against his will and forced behind the wheel of an already moving vehicle, these factors would, on most accounts, break the "trace" that anchors responsibility judgments.

The virtue of including a tracing component for a theory of moral responsibility is that it explains how agents can be responsible in a wide range of cases where they would otherwise seem to violate the conditions for being paradigmatically morally responsible. And, tracing approaches can be relied on in cases of both positive and negative evaluations of responsibility. Consider the case of Martin Luther declaring, "Here I stand, I can do no other."<sup>3</sup> Let us put aside issues of veracity, and let us grant that his action appropriately has our approval. Robert Kane claims that "[i]f Luther's affirmation did issue inevitably from his character and motives at the time it was made, then his moral accountability for it would depend on whether he was responsible *for being the sort of person he had become at that time.*"<sup>4</sup> Similarly, John Martin Fischer and Mark Ravizza have argued, "An individual might cultivate disposition to act virtuously in certain circumstances. It might even be the case that when he acts virtuously, his motivation to do so is so strong that the mechanism is not reasons-responsive . . . It is only when it is true that at no suitable point along the path to the action did a reasons-responsive sequence occur that an agent will not properly be held responsible for his action."<sup>5</sup>

2. I am assuming that any plausible account of the sort I have been describing here will leave room for the possibility of agents that are "suitably sensitive" but fail to act in accord with the relevant moral considerations.

3. Whatever its full history may be, Daniel Dennett's use of this case (see Daniel Dennett, *Elbow Room* [Cambridge: MIT, 1984], 133) seems to have given rise to the now standard and explicit reliance on tracing in contemporary accounts of moral responsibility.

4. Robert Kane, *The Significance of Free Will* (Oxford: Oxford, 1996), 39.

5. John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press, 1998), 50.

So, tracing can work by anchoring responsibility in either prior decisions to act (as in the case of Luis), or in the acquisition of dispositions, habits, or the self.

In this paper, I raise a difficulty for theories of moral responsibility that rely on tracing. If I am right about that there is a difficulty here, we might suppose that this shows that current accounts of tracing and/or the knowledge condition are inadequate. Or, we might suppose that these accounts are adequate, but that we are less frequently responsible than we suppose. Which conclusion we should accept is not clear. However, the deeper lesson may be this: the more theoretically attractive we make our combined accounts of tracing and the knowledge condition, the less frequently we seem to have free actions (and/or moral responsibility). Conversely, the broader range of cases we seem to insist are free (and/or responsible), the less attractive (and perhaps ad hoc) our account of tracing and the knowledge condition seems to become. So, my aim here is not to solve a problem so much as it is to illuminate a problem we have not sufficiently appreciated.

In section 1, I present a handful of cases that will figure in the discussion about the difficulties faced by standard accounts of tracing. In section 2, I introduce and discuss a knowledge condition on justified ascriptions of moral responsibility. In section 3, I show why a knowledge condition creates problems for tracing cases. In section 4, I consider the implications and significance of these difficulties.

## **1. FOUR CASES**

### **Jeff the Jerk**

Jeff is a middle aged middle manager in a mid-size company located somewhere in the Midwest. To him has fallen the task of alerting “downsized” employees of their new status as job seekers in a gloomy economy. That Jeff has the task is unfortunate for those about to be laid off, not only because they are about to lose their jobs, but—to add insult to injury—because Jeff is jerk. He is rude and inconsiderate about the feelings of others. And, he is unreflective about it. When people react poorly to his behavior (something they avoid doing because he is large, imposing, generally unsympathetic, and even a little frightening) he always writes it off as a shortcoming on the part of others. One afternoon, his superiors tell him that he needs to give notice to a group of long-time employees that they will be laid off. He does tell them, but in an altogether rude and insensitive fashion. Is Jeff responsible for the way he laid off his employees?

### **Britney the Bride**

Britney’s first experience of true love was with a boy who could really dance. Although the relationship eventually ended, it left her with a special place in her heart for men who could dance. Sadly, Britney was not especially fortunate in love. A particularly low moment came during a period of deep loneliness and difficulty

with her job. She invited a childhood friend to Las Vegas for the weekend, and after lots of drinking and flirting they ended up getting a quickie marriage, followed by an annulment within a few days. Though she wasn't aware of it, the bad wedding experience (with a guy who couldn't dance, no less!) made her especially want to settle down and "do it right" as soon as possible. A few months later, she met a charming professional dancer named Kevin. When the relationship turned serious, and Kevin proposed marriage, she instantly said "yes!" Her love of him was so deep that she could not even conceive of having told him no. Is Britney responsible for agreeing to marry Kevin?

### **Paulina the Paralyzed**

Paulina is visiting Florida for the first time. She takes her 2-year-old son, Paul, on a walk in a park just outside of Tallahassee. They eventually find a small clearing and decide to sit down and eat lunch. After a while, Paul gets up and starts wandering around with half-eaten food in his hands, on his face, and on his clothes. Paulina, being a responsible mother, keeps an eye on Paul and warns him against straying too far. Near the edge of the clearing, about 20 feet away, Paul bends over to pick up a rock that catches his attention. Paulina watches him do this, but at that moment realizes that there is an alligator staring at Paul from about 30 feet away. Paulina is paralyzed by surprise and some degree of fear. She feels a rising sense of panic, but remains frozen as the alligator start to move in the direction of Paul. Alas, the alligator moves surprisingly quickly, and snaps up poor Paul. Only then does Paulina finally unfreeze, and then she screams. Is Paulina responsible for not doing more to save Paul?

### **Ruben the Unfortunate**

It is the month of December. Ruben, a life-long fan of the Miami Dolphins football team, has just returned from seeing a real-life, in-the-stands Dolphins game for the first time. Even better, the Dolphins won the game, which in recent years is somewhat unusual, especially in December. As Ruben is wont to do, he enjoyed a few beers during the game. When he gets home from the game, his four-year-old niece is visiting. As is their customary ritual, he picks her up and throws her lightly in the air. Alas, because of his excitement about the football game and the lingering effects of an elevated blood alcohol level, he throws his niece up with more enthusiasm than usual and he fails to catch her on the way down—although she wiggled more than usual because of excitement about the throw. At any rate, sadly, she breaks her leg as a consequence of the fall. He views himself as responsible. But never mind that—do you think Ruben responsible for dropping his niece?

I suspect most of us would be inclined to think that, at least pre-philosophically, most or maybe all of the individuals in these examples are responsible for their actions (or inaction) as well as responsible for any obvious or predictable outcomes of those actions. That is, Jeff's being a jerk to the people

he fires is something for which is morally responsible, Paulina is morally responsible for not even moving to prevent the alligator attack on Paul, Britney is fully responsible for her decision to marry Kevin, and Ruben is genuinely responsible for accidentally dropping his niece.

To the extent that the cases are structurally similar, the similarity is this: they are all cases in which an agent has somewhat unexpectedly found him- or herself in a situation in which he or she may not count as a theory's paradigmatic example of moral responsibility, but for which he or she is nonetheless seemingly responsible.<sup>6</sup> Of course, cases like these are precisely the sorts of cases that the incorporation of tracing components to a theory of moral responsibility is designed to handle. And, it seems that a standard approach to tracing should be able to handle these cases. Take whatever your favored theory is of the conditions on moral responsibility, and apply it. If you accept that Jeff, Britney, Paulina, and Ruben are lacking freedom, responsiveness to reasons, or what have you at the time of action, then (on the supposition that there is nothing bizarre or pathological about these cases) we simply suppose that we can trace back their responsibility to some prior point in time when they did satisfy those conditions.

Before we start thinking too hard about these cases, though, I want to stop and consider another aspect of moral responsibility. We will return to these cases in a moment.

## 2. THE KNOWLEDGE CONDITION ON MORAL RESPONSIBILITY

Traditionally, most of the philosophical debate concerning free will and moral responsibility has focused on how to understand the sort of freedom required by free will and/or moral responsibility. Though some account of the freedom condition is crucial, a satisfactory account of moral responsibility will also include a knowledge condition on moral responsibility.<sup>7</sup> A (non-skeptical) theory of moral responsibility that is deaf to considerations about what an agent has reason to know will be ill-equipped to handle a large range of perfectly ordinary excusing conditions (from "Aaagh! I wouldn't have walked in the room if I had known what you two were doing in here" to "Sorry, I didn't realize you were waiting for me").<sup>8</sup>

Here is a characterization of the knowledge condition that seems to capture those instances, such as the ones mentioned above, when we think an agent ought to be excused on grounds of insufficient knowledge:

6. The characterization of "paradigmatic freedom" will be given varied construals depending on the details of the theory one favors. And, of course, I am assuming that there are no counterfactual interveners, nefarious neuroscientists, evil demons, and other malign-ilk beings lurking in the un-described parts of the examples.

7. This condition has been variously labeled "the epistemic condition" (in, for example, Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* [New York, Cambridge University Press], 13) and "the knowledge condition" (in, for example, Eddy Nahmias's Ph.D. dissertation *Free Will and the Knowledge Condition* [Duke, 2001]).

8. Though a skeptical theory might not care about epistemic conditions if, for instance, the skepticism is sufficiently motivated by a belief that we never have suitable control over our actions, non-skeptics must be prepared to acknowledge the existence of an epistemic condition on moral responsibility.

(KC) For an agent to be responsible for some outcome (whether an action or consequence) the outcome must be reasonably foreseeable for that agent at some suitable prior time.

KC gives expression to an idea that seems deeply embedded in common sense: holding an agent responsible for an action or consequence of some action is mistaken/inappropriate/unfair, and/or largely pointless, when the agent could not have reasonably foreseen the outcome. It is, however, only a construal of the knowledge condition in very broad outlines. Indeed, I take it that *any* extant account of the knowledge condition is provisional and stands in need of further specification, as I argue at the end of the paper.

The knowledge condition, as well as the KC construal of it, seems compatible with tracing. If I cannot currently foresee the outcome of one of my actions, but the reason why is because I knowingly made sure that I would not have that information, we can trace responsibility back to that prior moment when my later ignorance was reasonably foreseeable. Suppose I cannot now reasonably foresee the consequences of not immunizing my children (for example, that they are now at increased risk of whooping cough). If the reason for my inability to foresee this outcome is that I willfully refused to hear out any explanations about the benefits of immunization, deliberately chose not to read about the health needs of my children, and shut my eyes and ears to any discussions about how diseases and vaccines operate, I will hardly have absolved myself of responsibility for any outcomes that result. Why? Well, my failure to foresee the outcome is not reasonable given that I culpably worked myself into a context in which I could not foresee the outcome. So, on KC, we can suppose that any tracing will operate through the clause concerning reasonable foreseeability at some suitable prior time.<sup>9</sup>

The argument of this paper does not rely on a construal of the knowledge condition that is more fine-grained than KC. However, if you are already committed to a more fine-grained account of the knowledge condition, then feel free to substitute it for KC—it should not affect the outcome of the argument. All I require is a characterization of the knowledge condition that reflects the idea that agents ought not be held blameworthy for things they could not be appropriately expected to know, and that admits of the possibility of epistemic tracing.

### 3. THE TROUBLE WITH TRACING

Let us now return to those cases I presented in section 1: Jeff the Jerk, Britney the Bride, Paulina the Paralyzed, and Ruben the Unfortunate. Earlier, I claimed that what unified these cases was that they are all cases in which an agent has somewhat unexpectedly found him- or herself in a situation in which he or she seems to lack the sort of freedom paradigmatically required for moral responsibility, but

9. Whether the epistemic condition is treated as a separate condition, or whether it is located in the control condition, is immaterial for my purposes. For an account that puts it in the control condition, see Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Boulder, CO: Westview Press, 2000), 211.

not in a way that seems to undermine his or her responsibility. There is another thing that unifies these cases. In each of these cases, we have an action (or omission) that is brought about in large part by the presence of a disposition, character trait, habit, or other non-deliberative aspect of the agents.<sup>10</sup> This is different than cases, such as the one of Luis the drunk driver, where the outcome is brought about by the agent's specific deliberation about which outcome he or she aims to bring about. Let us call this a difference between deliberative and non-deliberative tracing cases.<sup>11</sup>

My interest is in the non-deliberative cases. These cases can be extremely puzzling, for it is difficult to see how a good many of them satisfy the knowledge condition. In particular, what we find is that there are often going to be non-deliberative cases where the non-deliberative source of behavior was acquired or retained under conditions where the agent could not have reasonably foreseen the later consequences of having that disposition, habit, or character trait. If so, then on standard accounts of responsibility, these agents are not responsible for the outcomes. But I suspect that for many the intuition that the considered agents *are* responsible will persist. So, either our theories or our intuitions must give way.

But first, why think non-deliberative cases are especially troubling? Let's take a closer look at Jeff's history:

As it is for many of us, high school was a crucial formative period in Jeff's life. He does not look back at it with any special fondness. In fact, he has largely forgotten the details. Nonetheless, it played an important role in creating the person who would become Jeff the Middle Management Jerk.

When Jeff was 15, he realized that he was having much less success with members of the opposite sex than he wanted to have. Over time, and through the usual fallible mechanisms of belief acquisition, he came to believe that the only males who consistently had success at gaining the attention of female classmates were those at least *we* might describe as jerks. In Jeff's hormone-ridden 15-year-old mind, this putative insight, coupled with a somewhat enterprising disposition that later served him modestly well in the business world, led him to adopt a plan for self-improvement. To the extent to which he was able, he inculcated in himself all the behaviors and attitudes that we would perceive to be jerk-like, and therefore, ulti-

10. There are some similarities shared by some but not all of the cases. For example, alcohol plays a role in the cases of Britney and Ruben, but it is merely an indirectly contributing or exacerbating factor in both cases (leading to her first marriage, and leading to his increased excitement). It was not a major component of the story in the same way in which alcohol was relevant to the story of Luis at the beginning of this paper. That is, it was not the alcohol that rendered them insensitive to the relevant downstream moral concerns or destroyed their access to alternative possibilities.

11. Fischer and Ravizza discuss a category of "trait actions" in Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York, Cambridge University Press), 87. I have opted for a different terminology so as to remain neutral on whether habits, dispositions, character traits, and other sources of unreflective (but presumably responsibility-bearing) sources of behavior are all traits in the relevant sense.

mately conducive to success with his female classmates. Now this initial decision to undertake the program of self-improvement was not an obvious choice for Jeff. He was worried about what his friends would think when he started attempting to behave differently. Would they make fun of him, noting that he was pretending to be something he wasn't? What if people found out why he was doing it? Could he overcome the shame and humiliation that would result from his female classmates learning of his subterfuge? These were the sorts of thoughts that Jeff had while he deliberated. After a period of uncertainty, however, Jeff decided to undertake the plan. And so he set about becoming, if not a jerk, at least jerk-like.

With surprisingly little effort, he succeeded. In fact he more than succeeded—it didn't even take the whole academic year for him to go from being jerk-like to being a full-on jerk. Part of the reason was undoubtedly rooted in his social context. The context was such that there was little cost to behaving in jerk-like ways when he would have been sensitive to those sorts of feedback. (He was a "latch key" kid, largely ignored by his permissive parents who chalked up his increasingly rude behavior to being a typical teenager. So, they tuned it out and treated him in a largely hands-off fashion). Moreover, what feedback he did receive and did care about was overwhelmingly positive. For whatever reason (perhaps it was merely the growth spurt that struck around the same time), he came to have tremendous success at attracting the attention of the opposite sex. However, at no point during the process of becoming a jerk, and certainly at no point before he undertook the program of becoming jerk-like, did Jeff even conceive that his plan for personal improvement would include in its outcomes that he would some day lay off employees in a despicable fashion. But it did, even though this was not at all reasonably foreseeable given his age and context.

Of course, these are all plausibly contingent facts about Jeff's history. Had he been some place else, perhaps as close as the school on the other side of town, things would have worked out differently. The people and social context would have been such that he would have ultimately abandoned the plan to become a jerk. Perhaps he would have been found out. Perhaps someone would have convinced him that this would have other deleterious effects on his life or on others. Perhaps someone might have even convinced him that it would lead to him becoming a quietly disliked middle-aged middle manager at a mid-sized company in the Midwest. But, alas, he was in the wrong place at the right time. Though he did become a jerk, and though he did become a jerk freely, and though he did foresee of his becoming a jerk that it would help him with his aspirations regarding the opposite sex, given his history and context it was not reasonably foreseeable to his hormone-fogged 15-year-old mind that undertaking this course of action would lead to his becoming the quietly disliked sort of person who would terminate the employment of people in unnecessarily rude and degrading ways.

One thing to note is that this example does not turn on Jeff being non-responsible for his decision to become a jerk. If we decide that Jeff was not respon-



sible for being a jerk, it still shows how satisfaction of knowledge conditions may not be anchored as often as we ordinarily suppose. If, however, we suppose that Jeff was responsible for becoming a jerk—he undertook the decision freely, he could reasonably foresee that his resultant behavior might have benefits with his female classmates, and that some of his friends might complain about his new behavior, and so on—the key thing to notice is that nothing in *this* decision satisfies the knowledge condition for his subsequent uncouth approach to employee termination. What a tracing-sensitive construal of the knowledge condition imposes is a requirement of reasonable foreseeability of the outcome at some suitable prior moment. Since Jeff is a jerk, and unreflective about his behavior, we have to find a prior moment when he could both act freely *and* reasonably foresee the outcome (of wrongfully poor treatment of his employees). The problem is that the only prior moment in his history that seems to be a good candidate—a moment where he freely acquired the trait that gave rise to the unreflective behavior—is a moment where the outcome we are concerned with was not even remotely foreseeable. So, even if we can have success tracing Jeff’s freedom or control over the decision to fire people back to his freedom to become a jerk, we certainly cannot trace reasonable foreseeability back in the same way. If so, then the link required for moral responsibility is broken.<sup>12</sup>

A second thing to note is that this example has a special bite for views that grant that much of our character and self-formation happens when we are young and when our ability to reasonably foresee the future is at its worst. If many of the courses of action we take are ones that seem obviously best to us, and if the reasons they seem obviously best to us have to do with freely created character traits, and if those character traits are often formed before we have a rich complement of beliefs about adult life and consequences of being different sorts of people, we might justifiably worry whether there are even enough cases where the knowledge condition is satisfied to give us even “in general” confidence in responsibility ascriptions.

A similar point can be made without relying on the idea that our earlier selves were simply immature or uninformed about the consequences of life-spanning decisions. The simple fact is that our epistemic powers tend to degrade very rapidly when they have to be projected more than a little into the future. For any theory that relies heavily on self- and character-forming accounts, it seems

12. Since I wrote this paper, Ron Mallon and other audience members at the University of Utah convinced me that the case of Jeff, as described, needs refinement. Briefly, the difficulty with the case stems from the idea that he knowingly acquires jerky characteristics, and that it plausibly follows that he can foresee that he will do jerk-like things as a consequence, even if the particular details (e.g., how he fires someone) cannot be foreseen. I don’t believe this problem plagues the other examples, and I think the main argument goes through in spite of the example in the text. That said, I think the best way to repair the case while preserving the spirit of the example involves the agent acquiring the relevant characteristics (being a jerk) while conceiving of them under a different guise (e.g., being “cool”), while blind to the negative aspects of the acquired trait or characteristic. These issues partly turn on the relationship between responsibility for general events (such as doing jerk-like things) and responsibility for finely grained events (such as firing someone in a particular way at a particular time), but a careful investigation of these issues will have to wait for another time.

extremely unlikely that those prior moments of self-formation were of the sort that included foreseeing the full range of downstream effects that would flow from that character. Even when we are at our epistemic best, our powers of prognostication are extremely limited.

One could object to the way I have described the case in the following fashion: “Surely *somewhere* between high school and that afternoon when he fired those people Jeff had received suitable evidence that his behavior was inappropriate. Indeed, presumably he would have received frequent evidence that he was a jerk, and his failure to respond accordingly is what grounds his responsibility in later life.” In reply: First, even if we suppose that Jeff was such that he eventually came to have sufficient evidence for him to reasonably foresee the consequences of his jerky behavior (by, say, age 27), we will nonetheless have a period of time between, on the one hand, his acquisition of the relevant trait and subsequent mistreatment of others as a consequence of it, and on the other hand, when he came to be in a position to appreciate the consequences of his behavior. During this time, however long it turns out to be, Jeff will fail to be in a position to satisfy the knowledge condition, and thus, on standard theories of responsibility that rely on tracing, if they require something like KC, he will fail to be responsible for the outcomes of behavior that work through his being a jerk. So, perhaps there are comparatively few middle-aged Jeffs who lack sufficient reason to question their jerkiness. Even so, there is some reason to believe that the structures I describe in Jeff could persist in a variety of forms for some time into adulthood, even if it is unlikely that many survive unchallenged into middle age. This ties to a second, and more important, point when thinking about the plausibility of these cases. The nature of being a jerk is often such that one tends to dismiss evidence that one is a jerk and/or that the consequences of being that way should give one a reason to change. This is true of many character deficiencies—they seem to bring with them the resources for dismissing countervailing evidence. Someone prone to anxiety can find her lack of good reasons to be anxious itself a reason to be anxious—perhaps it is a failure to detect things about which one should be anxious! Similarly, a megalomaniac sees criticism of his megalomania as confirmation that his true greatness is not appreciated or understood. (You think: “Vargas should know.” And I laugh that off, chalking up that comment to your shortcomings.) In this vein, we can imagine that Jeff’s jerkiness plays a big role in his not having reason to recognize the effects of his being a jerk, so that he might well have gotten into adulthood without having cause to question his being a jerk. This may become more plausible if we imagine Jeff to have other personality traits (themselves acquired under conditions where he could not foresee their effects), which dampen his susceptibility to these things.<sup>13</sup> Additionally, we might suppose that his imposing demeanor discourages people from complaining. And, we could also suppose that the environment of his job is one in which people do not complain about jerky behavior. So, even if we do not find many cases of people who make it to middle age without having knowingly free choices about being a jerk (anxious, a megalomaniac,

13. Consider this the requisite mention of Aristotle’s discussion of habituation and the acquisition of virtues (and vices) in Aristotle, *Nicomachean Ethics*, trans. Christopher Rowe (New York: Oxford, 2002), 111–13.

maniac, etc.), I suspect that we will be able to find a good number of such cases in younger adults.

Let us now consider the other cases. Here's Britney the Bride's back-story:

Though Britney does not realize it, her spontaneous and unreflective acceptance of Kevin's marriage proposal has roots in her past. In particular it was rooted in that night in Las Vegas when she first agreed to get married to that other man, her childhood friend. Unbeknownst to Britney, her booze-addled decision to walk down the aisle that night carried long-term ramifications for her. In particular, that decision and the unpleasant consequences that followed (together with facts about her personality and personal history) created in her a powerful disposition to accept marriage proposals from men (1) to whom she felt a certain attraction and (2) who were professional dancers. What the idea of a professional dancer did for her was to confirm an earlier, childhood commitment to a certain ideal of marriage and to a certain ideal of what a romantic partner should be like. This was nothing conscious, of course. Britney was unaware that her annulment would trigger a retreat to certain childhood images and ideals. And, it certainly was not foreseeable to her the night she got married in Las Vegas that this would lead her to become the sort of person who instantly accepts a marriage proposal from a professional dancer named Kevin. But it did. When she met Kevin, the dancer, she was already predisposed to find him immensely appealing in light of her earlier disastrous decision to get married. When, after only a small period of dating, Kevin surprised her with a marriage proposal there was, for Britney, nothing left to deliberate. Her hope and fantasies for marriage, operating at subconscious levels, left her ready to accept marriage from anyone who was both a dancer and a source of physical attraction. And so she agreed.

As in the case of Jeff, the case of Britney the Bride shows how there can be instances of non-deliberative actions for which an agent is intuitively responsible, but for which the agent fails to satisfy the knowledge condition. In the case of Britney, her acceptance of the marriage proposal is brought about because of a disposition of which she is unaware, and the acquisition of which she could not have reasonably foreseen. So, again, we have a non-deliberative tracing case in which the knowledge condition is violated.

Though the particulars of Britney's case are idiosyncratic, the general picture should be familiar to all of us. Few of us know why we find the particular things and people we love to be immensely, even irresistibly attractive. And, even when we are fortunate enough to know why we find these things attractive, it often does little to help rid oneself of the attraction. Even if I know what formative event in my life made me find, say, redheaded women to be immensely attractive, almost irrespective of their height, weight, or general physical attractiveness, this does little to give me direct control over my disposition to be attracted to redheaded women. So, whatever the peculiarities of Britney's personal triggers for attraction, the general psychological mechanisms involved in her spontaneous acceptance of marriage should be familiar enough.

One might try to resist the moral of Britney the Bride in several ways. First, one might object that even if Britney lacked responsibility at the moment of acceptance, surely she would be responsible by the time she took her vows, for by then she would have had sufficient time and opportunity to reflect on and to overturn her spontaneous commitment. And, that further reflection (or what is involved in further reflection) would be what grounds the attribution of responsibility, for at that point she would be well aware of the foreseeable consequences of sticking by her man. In response: What is at stake is not whether Britney could ever come to be responsible for her decision to get married. Rather, what is at stake is whether or not she is responsible for agreeing to marry Kevin *at the moment she accepts his proposal*. Here, it seems to me, that a good many people will think that Britney is responsible at the moment she does accept. Second, we should remind ourselves that it is the nature of some commitments that one closes off further consideration about whether or not to reconsider the commitment.<sup>14</sup> So, even if the particular case of Britney is one where she does reflect further on her acceptance of the engagement, real life is filled with cases where one spontaneously accepts a commitment whose nature or relationship to the agent is such that it closes off further deliberation for a while. (Think about these cases: agreeing to watch someone else's child for a few moments at the park; randomly picking a drink in response to a waiter's query at a bar; and perhaps most interestingly, sudden religious conversion.)

A second way to resist the conclusion that Britney was not responsible when she accepted Kevin's marriage proposal would be to argue that on the first date she had with him, before those non-volitional elements kicked in and rendered her helpless to the clarion call of marriage, a marriage proposal from him at some time in the future was reasonably foreseeable. In reply: I do not know whether such things are reasonably foreseeable on the first date. Certainly they are foreseen by some, but surely not by others. But whether or not this sort of thing is *reasonably* foreseeable is not clear. But even if it were reasonably foreseeable, it is relatively easy to imagine a variant of the scenario I have sketched where such a thing would not have been reasonably foreseeable. Consider Britney\*, who after her first marriage decided she was seriously opposed to marriage, that it was a form of enslavement for women, and that it was part of a long-standing morally objectionable tradition of treating women as less than fully human. She resolved never to marry anyone. In the case of Britney\*, we can suppose that most of the same psychological elements are in play as were in Britney—except that she is consciously against marriage and sub-consciously very much for it. Thus, when Kevin the dancer does propose to her, she astonishes herself and answers yes. It seems to me that Britney\* would be the sort of person of whom, even on the first date, it is not reasonably foreseeable that she would accept a marriage proposal from *anyone*, including the person she is dating. If so, then it seems that we still have a case where someone appears responsible for accepting a marriage proposal, even though in fact she does not satisfy the knowledge condition for moral responsibility.

14. The rationality of disfavoring reconsideration after the formation of plans and intentions has been discussed in Michael E. Bratman, *Intention, Plans, and Practical Reason* (Cambridge: Harvard University Press, 1987), 67–70, 88–91.

Let us now turn to the case of Paulina the Paralyzed, whose imperiling of Paul promoted his perishing. Here's how her dispositions, habits, and character traits converged in her unfortunate inability to act:

Paulina has a standing habit of taking Paul for walks in a park in the afternoon. It was the first time she had been to Tallahassee, and because she just never bothered to think about it, she did not have any expectations about the native fauna. When she did see the alligator, she was paralyzed because of several interacting factors. First, she did not expect to see an alligator. She was simply surprised. Second, it was big—a lot bigger than she thought alligators were. Third, it brought forth a visceral fear she hadn't felt in years—a fear of live reptiles with lots of big teeth. The absence of any one of these conditions would have meant that she moved to attempt to save Paul, but the confluence of these conditions is what left her paralyzed, swamping her desire to try and save her son for just long enough. Here's the thing: she could have undergone therapy that would have reduced or mitigated her fear of reptiles with lots of teeth. After all, her father was a therapist and her mother a herpetologist—so, she grew up knowing that she was afraid of these creatures, and she knew that if she really wanted to she had available to her the means to overcome her fear. But, the last time she thought about this fear she had no reason to believe that it would affect her ability to save her son (indeed, at the time, she didn't even have plans to have children). So, she decided to let sleeping fears lie.

Paulina did have it in her power to arrange things such that she could have avoided the outcome. She just could not have reasonably foreseen that things would work out that way. Similarly, she had it in her power to look up information about the fauna in Florida, but she simply did not foresee that she would have any need to know about it. Moreover, though she at some prior point had it in her power to insure that her child was not an especially tasty-smelling morsel for a hungry gator, again, she could not have reasonably foreseen the situation she found herself in. In short, though Paulina could have done otherwise at some prior point in time, and though at prior points in time she was responsive to considerations that would have avoided the outcome, she could not have reasonably foreseen the interaction of various habits, phobias, and human dispositions that brought about the sad outcome.

The case of Ruben the Unfortunate has many of the same elements, though in this case many of the habits and traits are freely acquired and actively sustained. He freely acquired his fandom of the Dolphins, and we can suppose that he actively took steps to maintain his enthusiasm as a fan. Similarly, he freely started the habit of lightly throwing his niece into the air, and when he has forgotten to do so on the rare occasion, he freely agrees to resume the practice when his niece begs her Uncle Ruben to greet her in the customary way. And, let us suppose that had any of the described elements been missing (the booze, the Dolphins win, the presence of his niece, the habit of greeting her in a particular way), no injury would have been done. What Ruben would have needed to know was that all of these things would interact precisely as they did, when they did. But that was not reasonably

knowable. He had no reason to believe his niece would be home, and that he would unthinkingly lift her up as he habitually did. It simply is not plausible to insist that he should have reasonably known this when he was ordering a beer at the football game, or when he walked in the door, and so on. Even at the moment he is lifting her up, he cannot reasonably foresee what will happen, precisely because what is happening is a habitual, undeliberative action. So again, we have a case where someone seems intuitively responsible but who does not seem to satisfy the knowledge condition at the time of action, and for whom no amount of tracing leads back to a point where the knowledge condition would be satisfied.

Together, the four cases I have been discussing represent a range of cases where a tracing approach seems to break down in light of difficulties satisfying the knowledge condition. In the case of Jeff the Jerk, we have a case where he is aware of the acquisition of the character trait (jerkiness), but unaware of its downstream effects. In the case of Britney, we have someone who freely undertakes a decision that has (unbeknownst to her) the effect of creating a powerful disposition whose operations she does not have knowledge of (the trait of readily accepting marriage proposals by dancers). In the case of Paulina, we have a case where someone is perhaps aware of her traits, but unaware of the effects of the *interaction* of these traits. In the case of Ruben, we have someone who freely acquires a variety of traits, some of which he foresees the consequences of, some of which he does not, and he is certainly unaware of the consequences of the interaction of these traits.<sup>15</sup>

There is, of course, more that can be contested about each of these cases.<sup>16</sup> However, the success of my argument does not depend on your agreement with my treatment of each and every one of these cases. Depending on your account of moral responsibility (on the assumption that it relies on tracing), perhaps one or two these cases can be rescued. All my argument requires is that at least one case (though I suspect that for most theories, it will be most of the cases) raises difficulties for your account. If so, then there is a class of cases where your account of tracing is ill suited to give a verdict that matches what is ostensibly common sense.

Before turning to a final set of issues, I want to respond to a family of general complaints about these cases and my use of them. First, one could object that these cases are cases where the agent is *not*, intuitively, responsible. So, contrary to

15. Some of these cases are, of course, cases of moral luck. See Thomas Nagel, "Moral Luck," *Mortal Questions* (1979). However, where Nagel seems to treat the problem of free will as one particular variety of moral luck (luck in antecedent circumstances), my discussion is focused on a specific problem generated by some particular features of an account of moral responsibility (namely, tracing and the knowledge condition). Moreover, Nagel's account focuses on luck in control or causal history as such. Here, my focus is not on these issues, but on the way in which even freely chosen features of lives and ourselves can, because of our epistemic limitations, yield unanticipated consequences.

16. In discussions with non-philosophers, these cases sometimes seem to elicit intuitions of strict liability. That is, I am sometimes told that in virtue of Paulina's role as a mother, or Ruben's relationship as an adult to a child, they are to be held responsible for whatever harms befall the child, irrespective of whether or not they satisfy the knowledge condition. It would be interesting to test intuitions on a variety of cases involving well-defined social roles that carry particular responsibility towards the injured party. In particular, it would be interesting to check responses across various demographics (gender, social class, ethnicity, age, etc.) so as to see if there is any uniformity about when people are likely to adopt a strict liability approach to these cases.

the claims of this paper, it is no objection to tracing-reliant accounts of moral responsibility that these are cases where the trace is broken. In response: For those of you who think this, my argument is not for you. However, there are two things to note. First, you do have an issue with everyone else who *does* think that at least some of these cases are cases where the agent is, at least intuitively, responsible. So, that my argument does not reach you does not mean that I have not raised issues that highlight further arguments that you have to make. Second, if you do think that *all*<sup>17</sup> of these are cases in which common sense denies that the relevant agents are responsible, the only way you are likely to persuade those of us who disagree is with some compelling empirical data about commonsense reactions to this case. So even if you reject my reading of the cases, you still have work to do.

A second complaint is this: what all of this shows is that there is no knowledge condition on moral responsibility—these cases are something of a *reductio* on the assumption that there is a knowledge condition. A superficial reading of some accounts of moral responsibility might be taken to suggest that a knowledge condition is spurious. For example, one might suppose that a Strawsonian account of responsibility that emphasizes the quality of will of particular agents need not require that an agent be able to reasonably foresee the consequences of his or her action—all we need is that the agent express the right quality of will (or perhaps more accurately, not express the wrong quality of will). Even on a Strawsonian account, however, one needs some account of how negligence is possible. When my toddler is injured playing with my pocketknife, I might be expressing no negative quality of will—I am simply negligent in leaving the pocketknife some place she might easily get to. To account for this sort of case, it seems the Strawsonian will need to invoke something like a prior failure to attend to some foreseeable outcome. If so, then it looks like even the Strawsonian requires a knowledge condition after all. And, Strawson acknowledges as much when he claims that ignorance is an excusing condition.<sup>18</sup>

A third complaint about these cases has to do with the presentation of the cases. One might object that had we been given the complete background about each of these cases when I originally presented them, it would not be so clear that these are cases of moral responsibility. In response: even if we did have that information, it is not clear why this counts against the power of the examples. After all, in real life, we do not ordinarily have full information about why people do what they do. So, if information of the sort I discussed would, in real life, change our willingness to ascribe responsibility in cases of this sort, and I am right that cases of this sort are reasonably common, this would be an interesting discovery. As it stands, I am agnostic about whether we should think the examples under full information are cases of responsibility, and thus I have endeavored to characterize these as cases of *apparent* responsibility. Indeed, rather than being an objection to the paper, the objection amounts to an endorsement of one of the two possible conclusions I discuss in the next section.

17. And you do have to think it applies to all of the cases, or else there is at least one case where it is a problem, and then the argument of the paper gets going.

18. See P.F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* XLVIII (1962), reprinted in Gary Watson, *Free Will*, 2nd ed. (New York: Oxford University Press, 2003), 73.

A fourth complaint is that what these cases show is that KC is simply an inaccurate construal of the knowledge condition. To this complaint, my reply is George W. Bush's (and perhaps as ill-advised): Bring it on. That is, I suspect that similar difficulties can be raised for any plausible account of the knowledge condition. But, in the absence of specific proposals, it is difficult to see what more can be said, other than to point out that the virtue of KC is that it is a pretty thinly specified knowledge condition, and that a more robustly specified condition seems likely to raise further difficulties, not less. What this points to, I think, is the difficulty of balancing, on the one hand, a plausible account of tracing and the knowledge condition with, on the other hand, the full scope of cases we think of as intuitively responsible. Favoring one seems to come at the cost of the other, and it is not obvious how one might have both.

#### 4. RESTRICTIVISM AND SIGNIFICANCE

The significance of the problem I have been considering turns, in part, on the frequency of problematic cases. At first blush, we might think that there are bound to be only a few cases like Jeff the Jerk, more cases like Paulina, and probably lots of cases like Britney and Ruben—i.e., cases where a complex combination of personal factors works us into a situation where some outcome results for which we lack direct freedom but for which we nonetheless seem responsible. But whether any of these cases happens very frequently or not turns, in part, on how frequently we directly exercise free will.

There is a continuum of possible answers to the question of how frequently we directly exercise free will. However, it is useful to think of two broad answers: restrictive and permissive. Restrictive accounts hold that, at least as adults, our having free will is a relatively rare phenomenon. Its proponents typically characterize it as being had only at moments of considerable indecision or incommensurability of values; it is these moments that underpin our responsibility, for they send us down varied life paths that make us the kinds of persons we are.<sup>19</sup> Restrictivism is typically motivated by the idea that “[t]here are . . . few occasions in life on which—at least after a little reflection and perhaps some investigation into the

19. For restrictivist accounts, see C. A. Campbell, “Is ‘Free Will’ a Pseudo-Problem?” *Mind* LX, no. 240 (1951); Peter van Inwagen, “When Is the Will Free?” in *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, 1989, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1989) reprinted in Timothy O’Connor, ed., *Agents, Causes, and Events: Essays on Indeterminism and Free Will* (New York: Oxford University Press, 1995), 219–269; Peter van Inwagen, “When the Will Is Not Free,” *Philosophical Studies* 75 (1994); and David Vander Laan, “A Regress Argument for Restrictive Incompatibilism,” *Philosophical Studies* 103 (2001). Robert Kane may also be committed to restrictivism. On his account, our moments of free will are phenomenologically distinguished by the experience of uncertainty. See, for example, Robert Kane, “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism,” *Journal of Philosophy* XCVI, no. 5 (1999), reprinted in Laura Waddell Ekstrom, ed., *Agency and Responsibility: Essays on the Metaphysics of Freedom* (Boulder, CO: Westview, 2001), especially p. 164. If these occasions turn out to be comparatively infrequent, then Kane’s view turns out to be restrictivist as well.



fact—it isn't absolutely clear what to do."<sup>20</sup> In contrast, permissive accounts hold that we regularly or frequently exercise free will.

These are obviously imprecise categories, but they suffice to illustrate two general points. First, tracing typically plays a crucial role in restrictivist accounts of moral responsibility. Without it, it looks highly implausible that restrictivist accounts of free will provide enough opportunities for agents to be responsible with anything like the frequency that we ordinarily suppose them to be. There simply would not be enough free will to go around. So, restrictivists must appeal to a considerable amount of tracing to account for our moral responsibility. Though this might initially seem like a dubious strategy, restrictivists will also typically help themselves to a view where the exercise of our free will plays an important role in our self-creation.<sup>21</sup> So, even if we must trace responsibility for some outcome back to, for example, the character of the agent, as long as that character was arrived at via some route suitably involving the exercise of free will, then we have something that anchors the tracing of responsibility.

The restrictivist strategy's anchoring of responsibility in character traits may raise a worry about the empirical plausibility of this strategy. There is a concern, familiar to those who have followed recent debates about character traits and their significance for ethical explanations, that common sense and/or many philosophical accounts of character are systematically mistaken.<sup>22</sup> If so, one might wonder whether or not the restrictivist's appeal to character traits and/or the agent's role in forming the "anchoring" features of the self are vulnerable to similar charges of being mired in demonstrably false views about the nature of the self. The significance of situationism for accounts of character is contested, and settling its significance for free will would take us far afield.<sup>23</sup> Whatever the outcome of these

20. Van Inwagen, "When Is the Will Free?" in *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, 1989, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1989), 232.

21. For instance, on p. 237 of "When Is the Will Free?" van Inwagen writes, "It is an old, and very plausible, philosophical idea that, by our acts, we make ourselves into the sorts of people we eventually become. Or, at least, it is plausible to suppose that our acts are *among* the many factors that determine what we eventually become." Kane even adopts the terminology of "Self-Forming Actions" and "Self-Forming Willings" to refer to the exercise of free will. See Kane, *The Significance of Free Will* (Oxford: Oxford, 1996), 124–25.

22. See John Doris, *Lack of Character* (New York: Cambridge, 2002) and Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error," *Proceedings of the Aristotelian Society* 99 (1998–99), reprinted in Gilbert Harman, *Explaining Value and Other Essays in Moral Philosophy* (Oxford: Oxford, 2000).

23. See, for example, Gopal Sreenivasan, "Errors about Errors: Virtue Theory and Trait Attribution," *Mind* 111, no. 441 (2002), and Maria Merritt, "Virtue Ethics and Situationist Personality Psychology," *Ethical Theory and Moral Practice* 3 (2000). As far as I am aware, the only published engagement with these issues in the context of free will and moral responsibility include chapter 7 of Doris, *Lack of Character*, John Doris and Stephen Stich, "As a Matter of Fact: Empirical Perspectives on Ethics," in *The Oxford Handbook of Contemporary Analytic Philosophy*, ed. Frank Jackson and Michael Smith (Oxford: Oxford, forthcoming) and Dana Nelkin, "Freedom, Responsibility and the Challenge of Situationism," in *Midwest Studies in Philosophy* (2005). There are also a number of relevant unpublished items, including John Doris and Dominic Murphy, "War Crimes" (UCSC/Caltech: 2004) and Eddy Nahmias, "Free Will, Knowledge, and the Threat of Social Psychology (unpublished manuscript)" (FSU: 2004).

debates, it is worth reminding ourselves that we ought not to naively suppose that our commonsense views about the self, and about free will, will be fully vindicated by a scientifically credible picture of the world.<sup>24</sup>

A second point we can make about this distinction is that it is orthogonal to the compatibilism/incompatibilism debate. The essential structure of restrictivist arguments is (1) given a particular account of free will and (2) given some prima facie plausible assumptions, it is a contingent empirical fact that we exercise free will only rarely. There is no reason why it could not turn out that, assuming the truth of some compatibilist account of free will, the world is structured such that we only rarely have it. And, it could even turn out that as a matter of contingent empirical fact we *never* have free will.<sup>25</sup> Pessimistic compatibilism is not an especially popular view, but there is nothing that excludes its possibility, nor the possibility of its somewhat less radical and more appealing sibling, restrictivist compatibilism.<sup>26</sup>

If restrictive accounts of free will are true, then given the extent to which restrictive accounts must rely on tracing (and in particular, tracing back to the formation of character traits and habits), the difficulties I have raised for satisfying the epistemic condition strike especially deep. Even on permissive views, though, the difficulties I raise do not disappear. Given that even many permissive views adopt tracing components to handle cases like drunk driving, it appears that even the most permissive views that countenance tracing will have to admit that in a fair number of tracing-invoking cases the knowledge condition on moral responsibility will not be satisfied. Though accurate estimates cannot be attained by mere armchair reflection alone, it does seem plausible that a good chunk of our daily behavior, especially our morally important behavior, consists in actions brought about by the interaction of various personal traits and habits that were individually acquired without any sense of how they would later interact with other traits and habits in novel situations. If so, then we have a significant problem that is pervasive enough, even if there is some sense in which the problem could be more (or less) pervasive.

The extent to which it is a problem will partly depend on the things I have emphasized here: the frequency with which an agent has both freedom and knowledge, the extent to which traces are anchored in features acquired by agents while epistemically immature, and the temporal extendedness of the traces. These elements will be given their contents by the particulars of a given theory of free will and the contingent social and psychological facts of human lives. However, to the

24. See, for example, Henrik Walter, "Neurophilosophy of Free Will," in *The Oxford Handbook of Free Will*, ed. Robert Kane (New York: Oxford, 2002), Owen Flanagan, *The Problem of the Soul: Two Visions of the Mind and How to Reconcile Them* (New York: Basic Books, 2002), and Manuel Vargas, "Reductio ad Libertarianism: Compatibilist Ultimacy and the Plausibility of Libertarianism" (unpublished manuscript).

25. Thanks to Alan Hájek for reminding me of this possibility.

26. Some criticisms of restrictivism can be found in John Martin Fischer and Mark Ravizza, "When the Will Is Free," in *Philosophical Perspectives* 6, *Ethics* 1992, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1992), and Gordon Pettit, "Are We Rarely Free? A Response to Restrictivism," *Philosophical Studies* 107 (2002).

extent that we acknowledge that at least some of our morally significant choices depend on long-standing features of ourselves that we acquired in circumstances that are epistemically remote from our current decisions, we should be troubled by reliance on tracing approaches.

There is at least one class of theories that are immune to these worries about tracing: theories that do not rely on tracing. “Structuralist” theories that take their cue from Harry Frankfurt’s early work hold that assessments of responsibility are settled entirely by the non-historical structures of the agent (such as hierarchies of desires). A purely structuralist account of responsibility is immune to the worries motivated by Jeff the Jerk and related cases because Jeff’s responsibility for the manner in which he dismisses the employees will be accounted for purely in terms of the structure of Jeff’s agency at the time of action. These worries about tracing might therefore seem to constitute something of an argument in favor of structuralist theories of responsibility. However, the absence of any role for tracing in structuralist accounts seems at least as problematic as the troubles caused by tracing in non-structuralist accounts. Pedestrian cases (so to speak) involving drunk drivers, as well as a range of somewhat more esoteric manipulation cases, seem deeply problematic for structural theories.<sup>27</sup>

Since the trouble with tracing is not restricted to restrictivism, and since the only theories that seem entirely immune to it are widely criticized for lacking the element that gives rise to the problem, what are we to make of the dialectical situation? There seem to be two main options: we can abandon our current theories or we can abandon common sense and embrace the idea that we are less frequently responsible than we suppose.<sup>28</sup>

Suppose that we decide the problem is with our theories of moral responsibility. Perhaps tracing is the wrong way to address the cases to which it has been applied. If so, it remains to be seen just what alternatives there are that do not collapse back into purely structural theories. However, rather than abandoning reliance on tracing the most promising route seems to be careful refinement of the knowledge condition. The challenge is to do it without making our accounts of it *ad hoc*. If I am right that there is a tension between that aim and the aim of having the cases I have discussed turn out to be cases of responsibility, then this will not be easy. Indeed, I suspect that what we will find is that a suitably refined account of the knowledge condition can account for many but not all cases. If so, we may simply have to accept some degree of counterintuitiveness about the cases it does

27. Cases of manipulation are often taken to be decisive against structural theories. However, the significance of manipulation cases for these issues is complicated, as I argue in Manuel Vargas, “On the Importance of History for Responsible Agency,” *Philosophical Studies* (forthcoming).

28. Another option would be to say that we are less (as in “not fully”) responsible. As I currently think about these things, I am inclined to think of responsible agency as a binary concept—you either are or are not a responsible agent—although issues of blameworthiness are a matter of degree. If so, perhaps we should think of the freedom condition as a condition of responsible agency and knowledge as a condition of blameworthiness. This would be consistent with thinking of blame as a matter of degree, and that failure to fulfill the knowledge condition might also come in degrees, in addition to being a matter of frequency. So perhaps there are really three options.

not accommodate. But these are issues that await a more careful development of the knowledge condition.

There is, however, another possibility. We might instead suppose that there is nothing substantially wrong with our accounts of tracing or the knowledge condition. Instead, what these thoughts could be taken to show is that we really are less frequently responsible than we might have thought. Claims of this sort are always regarded with skepticism in some quarters. Suppose one believed that it is a shortcoming of a philosophical theory if it does not explain how what we already believe could be true. For any such view about the aim of philosophizing, an account that calls for a revision in our beliefs or practices will therefore be viewed as an unsatisfactory or implausible account because it fails in the aim of an adequate philosophical explanation. And, since these results seem to show that we are less responsible than we think, the considered account must be rejected. Though perhaps not indefensible, this conception of the aim of philosophical theorizing seldom receives much defense, and I have argued against in several places.<sup>29</sup>

However, even if we do accept that we are less frequently responsible than we suppose, it is not obvious what follows. Given the immensely complex task it would be to analyze every case of non-deliberative action for its status with respect to the knowledge condition, and given the role our responsibility ascriptions play in regulating social practices, there might be good pragmatic reasons to—at least in ordinary cases—continue as we were. We should be careful about being too conservative, however. Unjustified blame seems to carry a moral cost, subjecting its target to harm. In contrast, unjustified praise is usually comparatively harmless. If this is correct, we would have some reason to favor an asymmetry in praising and blaming—permissiveness in praise, and caution in blame.<sup>30</sup> This might allow us to conserve some of our practices while acknowledging that we are less frequently responsible than we currently tend to believe. However, the centrality of blame and punishment in our everyday practices make such revisions difficult to readily accept. So here, too, a revisionist proposal seems subject to doubt. However, it seems to me that until we have some reasonably good answer to the challenges raised by tracing and the knowledge condition, these are possibilities that merit further exploration.

In many domains of philosophy it is unclear at what point theory overrides intuition, and when intuitions are sufficiently strong so as to justify the dismissal of favored theories. It is not different here, but unlike many areas in philosophy, we have spent entirely too little time articulating the compromises and standoffs raised by various conceptions of tracing and the knowledge condition.

29. Vargas, “On the Importance of History for Responsible Agency,” Manuel Vargas, “Responsibility and the Aims of Theory: Strawson and Revisionism,” *Pacific Philosophical Quarterly* 85, no. 2 (2004), Manuel Vargas, “The Revisionist’s Guide to Responsibility,” *Philosophical Studies* (forthcoming).

30. A related but distinct idea of asymmetry in praise and blame is developed in Susan Wolf, “Asymmetrical Freedom,” *Journal of Philosophy* 77 (1980). However, Wolf’s account has to do with an asymmetry in the conditions for freedom, not an asymmetry because of pragmatic considerations of a difficulty raised by the knowledge condition.

## APPENDIX

## Further Thoughts on the Knowledge Condition

Consider Carl Ginet's account of "K," a construal of the knowledge condition that is more finely specified than KC:

[Where "know\*" and cognates refer to something that is actively believed,] either (i) S knew\* at t1 that his moving in way M then would or might bring it about (or, in the case of omission, prevent) a harm of sort H, or (ii) S did not then know\* this but there is a sequence of one or more prior acts (or omissions) that ends with the act (or omission) at t1 and is such that (a) each member before this last member benights the subsequent member, (b) the first (earliest) member of the sequence was *not* a benighted act (or omission) that is the next member of the sequence—and (c) at the time of each benighted act (or omission) in the sequence S should have known\* (was blameworthy for not knowing\*) that it would or might lead to the sort of benighted act (or omission) that it in fact led to in the next member of the sequence.<sup>31</sup>

As with KC, it is compatible with a notion of epistemic tracing. Indeed, it specifically builds it in clause (ii).<sup>32</sup> Though there are many excellent features of Ginet's account, it seems that further specification is possible and even desirable. For instance, it seems that we demand that agents be differentially responsive to the probabilities of morally significant outcomes (Ginet's "harm of sort H") in complex ways. As Fischer and Ravizza have noted,

In some contexts, it seems appropriate to hold an agent responsible for a later action (or omission or consequence) that is extremely unlikely to occur, whereas in other contexts the extreme unlikelihood of (say) the action seems to rule out responsibility. This makes it reasonable to think that a full and explicit tracing approach would not specify a degree of likelihood that is always employed straightforwardly to ascertain responsibility; rather, the degree of likelihood employed by the tracing approach would need to be context-relative.<sup>33</sup>

If Fischer and Ravizza are right, then even an account such as Ginet's will require further specification about why different probabilities matter in the contexts that

31. Carl Ginet, "The Epistemic Requirements for Moral Responsibility," in *Philosophical Perspectives 14: Action and Freedom*, 2000, ed. James Tomberlin (Malden, MA: Blackwell, 2000), 275.

32. As I understand it, Ginet's account does not specify what degree of epistemic possibility matters, or whether any possibility at all should be significant. Nonetheless, I take it that there is something promising about the general framework of his account of the knowledge condition.

33. Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (New York: Cambridge University Press), 50.

they do. In turn, this seems to require some more systematic account of what makes some morally significant possibilities epistemically relevant and others (of equal likelihood) not. If we were presented with such an account, one where some morally significant possibilities of a particular probability are to be ignored and others of equal probability are not to be ignored, a natural question to ask would be whether or not this is evidence for our being morally unprincipled. And, if we are being morally unprincipled, this might suggest that we need to go back and revise our understanding of the knowledge condition. To block this conclusion we would need (i) some account of what justifies differences in epistemic and deliberative relevance of equal moral possibilities, and (ii) proof that the differences in epistemic relevance that we do acknowledge as intuitively right are indeed covered by (i). Surely (i) is relatively easy to show for a range of cases (e.g., remote possibilities of directly causing harm are justifiably more important to attend to than remote possibilities of indirectly causing harm, at least if we put aside some consequentialist accounts). I am far less convinced that (ii) would be as easy to show.

These issues are complex and difficult to sort out. I suspect that we will find that careful work on the knowledge condition will face difficulties no less challenging than those that face accounts of the freedom condition.<sup>34</sup>

## REFERENCES

- Aristotle. 2002. *Nicomachean Ethics*. Trans. Christopher Rowe. New York, NY: Oxford.
- Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Campbell, Charles Arthur. 1951. "Is 'Free Will' a Pseudo-Problem?" *Mind* LX, no. 240:441–65.
- Dennett, Daniel. 1984. *Elbow Room*. Cambridge: MIT.
- Doris, John. 2002. *Lack of Character*. New York, NY: Cambridge.
- Doris, John, and Dominic Murphy. 2004. "War Crimes." Unpublished manuscript, UCSC/Caltech.
- Doris, John, and Stephen Stich. "As a Matter of Fact: Empirical Perspectives on Ethics." In *The Oxford Handbook of Contemporary Analytic Philosophy*, ed. Frank Jackson and Michael Smith. Oxford: Oxford, forthcoming.
- Ekstrom, Laura Waddell. 2000. *Free Will: A Philosophical Study*. Boulder, CO: Westview Press.
- , ed. 2001. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder, CO: Westview.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York, NY: Cambridge University Press.
- . 1992. "When the Will Is Free." In *Philosophical Perspectives 6, Ethics 1992*, ed. James E. Tomberlin, 423–51. Atascadero, CA: Ridgeview.

34. The ideas in this paper have benefited from conversations with many students and several colleagues over the past few years, including participants in seminars conducted at the California Institute of Technology and the University of San Francisco. Thanks to Eddy Nahmias (twice over!), Dan Speak, for very helpful comments on earlier versions of this paper. I am sure I have not sufficiently accommodated their concerns (to the detriment of this paper). If memory serves, some of the credit for these ideas can be traced back to Derk Pereboom. At the excellent Inland Northwest Philosophy Conference in 2001, we learned that we both had stumbled across this issue and spoke about it briefly on the way to lunch. So, what good there is in the ideas here may be to his credit, though as I write this I do not know if he still holds the view or has abandoned it for other reasons. The errors and confusions, however, are undoubtedly all mine. Finally, my thanks to the California Institute of Technology for financial support while working on parts of this paper.

- Flanagan, Owen. 2002. *The Problem of the Soul: Two Visions of the Mind and How to Reconcile Them*. New York: Basic Books.
- Ginet, Carl. 2000. "The Epistemic Requirements for Moral Responsibility." In *Philosophical Perspectives 14: Action and Freedom, 2000*, ed. James Tomberlin, 267–77. Malden, MA: Blackwell.
- Harman, Gilbert. 2000. *Explaining Value and Other Essays in Moral Philosophy*. Oxford: Oxford.
- . 1998–99. "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error." *Proceedings of the Aristotelian Society* 99:315–31.
- Kane, Robert. 1999. "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism." *Journal of Philosophy* XCVI, no. 5:217–40.
- . 1996. *The Significance of Free Will*. Oxford: Oxford.
- Knobe, Joshua. 2003. "Intentional Action in Folk Psychology: An Experimental Investigation." *Philosophical Psychology* 16:309–24.
- Merritt, Maria. 2000. "Virtue Ethics and Situationist Personality Psychology." *Ethical Theory and Moral Practice* 3:365–83.
- Nagel, Thomas. 1979. "Moral Luck." In *Mortal Questions*, 24–38.
- Nahmias, Eddy. 2001. *Free Will and the Knowledge Condition*. Ph.D. dissertation, Duke University, Durham, NC.
- . 2004. "Free Will, Knowledge, and the Threat of Social Psychology" (unpublished manuscript). FSU.
- Nelkin, Dana. 2005. "Freedom, Responsibility and the Challenge of Situationism." *Midwest Studies in Philosophy* 29:181–206.
- O'Connor, Timothy, ed. 1995. *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York, NY: Oxford.
- Pettit, Gordon. 2002. "Are We Rarely Free? A Response to Restrictivism." *Philosophical Studies* 107: 219–37.
- Sreenivasan, Gopal. 2002. "Errors about Errors: Virtue Theory and Trait Attribution." *Mind* 111, no. 441: 47–68.
- Strawson, P. F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* XLVIII: 1–25.
- van Inwagen, Peter. 1989. "When Is the Will Free?" In *Philosophical Perspectives, 3, Philosophy of Mind and Action Theory, 1989*, edited by James E. Tomberlin. Atascadero, CA: Ridgeview.
- . 1994. "When the Will Is Not Free." *Philosophical Studies* 75: 95–113.
- Vander Laan, David. 2001. Regress Argument for Restrictive Incompatibilism." *Philosophical Studies* 103: 201–15.
- Vargas, Manuel. "On the Importance of History for Responsible Agency." *Philosophical Studies* (forthcoming).
- . "Reductio ad Libertarianism: Compatibilist Ultimacy and the Plausibility of Libertarianism." (unpublished manuscript).
- . 2004. Responsibility and the Aims of Theory: Strawson and Revisionism." *Pacific Philosophical Quarterly* 85, no. 2:218–41.
- . "The Revisionist's Guide to Responsibility." *Philosophical Studies* (forthcoming).
- Walter, Henrik. 2002. "Neurophilosophy of Free Will." In *The Oxford Handbook of Free Will*, ed. Robert Kane, 565–76. New York, NY: Oxford.
- Watson, Gary. 2003. *Free Will*, 2nd ed. New York, NY: Oxford.
- Wolf, Susan. 1980. "Asymmetrical Freedom." *Journal of Philosophy* 77:150–60.