

## *The Truth about Tracing*

JOHN MARTIN FISCHER AND NEAL A. TOGNAZZINI  
University of California, Riverside

The Past! the dark, unfathom'd retrospect!  
The teeming gulf! the sleepers and the shadows!  
The past! the infinite greatness of the past!  
For what is the present, after all, but a growth out of the past?

– *Walt Whitman, "Passage to India"*

### **1. The Trouble with Tracing**

It is widely agreed among theorists of moral responsibility that the following two statements, if true and uttered sincerely, will excuse someone from apparent responsibility:

- (1) I couldn't help it.
- (2) I didn't, and couldn't have been expected to, know that it would happen.

If Kevin's kleptomania is so severe that it produces in him a literally irresistible desire to steal, then there is a straightforward sense in which he couldn't help himself (at the time he steals). This sort of severe kleptomania is plausibly thought to excuse Kevin's behavior on the grounds that his behavior was not sufficiently under his control. So, it looks like some sort of "control condition" will be a necessary component of any plausible theory of moral responsibility.

Similarly, if Kevin's friends are planning a surprise party for him but they neglect to tell Dan that it's a surprise and Dan subsequently talks openly with Kevin about the party, Dan's ignorance plausibly excuses his behavior. Since he didn't know (and, we suppose, could not have been expected to

know) that the party was a surprise, he didn't know that talking openly with Kevin about the party would amount to ruining the surprise. His ruining the surprise is excused because of his impoverished epistemic position. So, it looks like some sort of "epistemic condition" will be a necessary component of any plausible theory of moral responsibility, as well.

Arguably, one more component must be added to get a plausible theory of moral responsibility, and it may cause trouble when combined with the epistemic and control conditions. Suppose that Dan gets drunk at Kevin's surprise party and drives home, but his reflexes are so impaired that he runs over a pedestrian on the way. Whatever sort of control is required for moral responsibility, suppose that he didn't have that control at the time he ran over the pedestrian. In a straightforward sense, then, he couldn't help it. But is he therefore not responsible? Of course not, because in another straightforward sense, he *could* help it. After all, presumably he had control of whether to get drunk at the party in the first place. In this case, Dan is morally responsible for running over the pedestrian even though he didn't satisfy the control condition at the time of the accident. To account for this, our theory of moral responsibility should include some notion of *tracing*. In this sort of situation, Dan's moral responsibility can be *traced back* to an action of which he had control—namely, his getting drunk.

Of course, as we've seen, it wouldn't be enough merely to trace Dan's responsibility back to an action of which he had control. He would also need to be in an appropriate epistemic position with respect to the consequences of that action in order for tracing to be appropriate. That is, not only must Dan have had control of his getting drunk, he must have also known that getting drunk might well lead to his driving home drunk and to his running over a pedestrian. It seems that tracing only works if the situation to which we trace is one in which the agent both has sufficient control of some action *and* can be reasonably expected to know the likely results of that action. Drunk-driving cases are unproblematic precisely because everyone knows (or at least *should* know) that too much alcohol will impair the ability to drive a car.

In an influential and highly provocative recent article, Manuel Vargas has presented a number of cases that bring out an apparent tension between the idea of tracing and the epistemic condition on moral responsibility.<sup>1</sup> We will examine these cases in detail below, but the tension raised by each is this: in some situations in which it seems that we need to invoke tracing in order to explain our intuitions of moral responsibility, there is no prior time available at which the agent in question could have reasonably foreseen the consequences of her actions to serve as the terminus of the trace. Indeed, these cases appear to show that we are faced with three, equally troublesome options: (i) the idea of tracing needs to be either abandoned or refined, (ii) the epistemic condition on moral responsibility needs to be either abandoned or significantly refined, or (iii) we should admit that we are morally responsible much less often than we are ordinarily inclined to think.

Now, it is certainly true that a fully adequate theory of moral responsibility will examine the nature of both tracing and the epistemic condition in detail. So we don't want to discourage anyone from conducting these important research projects. However, we are not convinced that the cases Vargas concocts are indeed as troublesome for extant theories as he alleges. In what follows, we will examine Vargas's cases in detail and we will argue that theorists who are attracted to tracing have nothing to fear from them.

The four cases that Vargas presents are quite different and, as we will see, no one response will take care of them all. But there is an abstract structure that they all putatively share. Each is intended to present a scenario according to which the following conditions hold:

- (i) The agent is morally responsible for the relevant outcome
- (ii) The agent does not satisfy the control condition on moral responsibility (at the time of his behavior) with respect to the relevant outcome, and
- (iii) There is no suitable prior time at which the relevant outcome was reasonably foreseeable for the agent

To account for the apparent inconsistency in (i) and (ii), it looks like we need to invoke some sort of tracing. However, when we consider (iii) together with some plausible epistemic condition on moral responsibility, it looks like no attempt at tracing can be successfully completed. In order for Vargas's cases to make trouble for tracing, he needs them to satisfy all of the above conditions. We will argue, however, that each case fails to satisfy at least one of them. In none of Vargas's cases is it plausible to think that all of the conditions are satisfied. We shall then go on to apply a similar analysis to some intriguing examples presented by Angela Smith;<sup>2</sup> in the end, our contention will be that tracing emerges unscathed.

## **2. Jeff the Jerk**

Consider first the case of 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?<sup>3</sup>

This is the initial story provided by Vargas, but he goes on to fill in some pertinent details about how Jeff came to be the jerk that he is. These details are important, as they supposedly provide us with reason to think that we won't be able to account for Jeff's moral responsibility in this case by invoking tracing.

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, ultimately 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.<sup>4</sup>

So goes the case of Jeff the Jerk. Vargas thinks that upon reading this case, we will have the following three reactions. First, we will judge that Jeff is morally responsible for the jerky way in which he laid off his employees. Second, we will judge that due to the specific nature of his jerky character, he doesn't

satisfy the control condition on moral responsibility with respect to laying off his employees in the way in which he did. These two insights together should lead us to invoke tracing in order to account for Jeff's moral responsibility. However, our third reaction should be that when Jeff voluntarily undertook to acquire a jerky character, he could not have reasonably been expected to have foreseen that his acquiring a jerky character would lead to his firing the employees in the precise manner in which he did. Thus, the epistemic condition on moral responsibility will prevent us from tracing Jeff's moral responsibility back to his adolescent decision to acquire a jerky character.

Let's look closely at each of these expected reactions. We think it is quite plausible to suppose that Jeff is indeed morally responsible for the way in which he fired his employees. So we agree that this case satisfies (i) above. What about (ii)? Does Jeff satisfy some plausible control condition on moral responsibility when he fires his employees? To answer this question, we'll need to elaborate a little on what exactly is involved in the control condition.

There is of course much disagreement about just what capacities an agent needs to have in order to have the control necessary for moral responsibility. Fortunately, there is no need to take a stand on this issue in order to conduct a fruitful discussion of Vargas's cases. There is no need to take a stand here on whether the control condition should be interpreted in a compatibilist or an incompatibilist way.<sup>5</sup> Further, there is no need to decide between an "alternative-possibilities" interpretation and an "actual-sequence" interpretation. So, for example, one could interpret the control condition as requiring freedom to do otherwise (or freedom to bring about a different consequence, and so forth); alternatively, one could interpret it as requiring "acting freely". In what follows, absolutely nothing will hinge on adopting one interpretation over another.

Just for simplicity, we will adopt the "acting freely" interpretation; we emphasize however that, if you prefer the "could have done otherwise" interpretation, you can make the substitution with full preservation of content. Nothing in our critique of Vargas's cases will depend on adopting a particular (plausible) interpretation, rather than another. According to one account of "acting freely", an agent acts freely and thus satisfies the control condition on moral responsibility with respect to some action insofar as the action in question issues from the agent's own, moderately reasons-responsive mechanism.<sup>6</sup> Roughly, the mechanism in question must be one for which the agent has taken responsibility and also one that displays a specific combination of receptivity and reactivity to reasons.<sup>7</sup> Details aside, is it plausible to suppose that Jeff does not satisfy this control condition when he fires his employees?

Vargas indicates that Jeff's action is done "unreflectively", and perhaps this is meant to give us reason to suppose that Jeff fails to satisfy the control condition. But it is not at all clear that unreflective actions are *ipso facto* not under our control. Indeed, it seems perfectly compatible with the situation as Vargas has described it that Jeff's firing the employees in fact issued from

his own, moderately reasons-responsive mechanism. Insofar as it is plausible to suppose that there is a mechanism of *nondeliberative habit*, there seems to be no problem in supposing that this mechanism might be both moderately reasons-responsive and also one for which the agent has taken responsibility.<sup>8</sup> Thus, on a plausible interpretation of the control condition, the mere fact that Jeff's action is unreflective does not entail that the control condition is not satisfied.<sup>9</sup>

Further, we would typically assume that even a very jerky jerk satisfies the control condition. That is, even someone whose jerkiness is ingrained and significant—issuing in strong urges to jerky behavior or a fixed disposition to ignore non-jerky options—is assumed to have the capacity to rise above his jerkiness. We typically hold jerks responsible for their jerky behavior precisely because we assume that their behavior does not result from literally irresistible urges—or anything close. So even as Vargas develops the story, one might be skeptical about whether Jeff really fails to satisfy the control condition, quite apart from issues about the relationship between unreflective behavior and the control condition.

So it is quite unclear that we have here a case in which the control condition is not satisfied. But we think there are deeper problems yet to be exposed, so we will at this point simply grant, for the sake of argument, that Jeff in fact does *not* satisfy the control condition when he fires his employees. Given that Jeff's case satisfies both (i) and (ii), it looks like we will need to invoke tracing for the sake of consistency. So the next question is whether there is any prior time to which we can successfully trace. Vargas maintains that there is no such time, since the most natural candidate—the time when Jeff decides to acquire a jerky character—is a time at which Jeff is in an inadequate epistemic position with respect to his much later action of firing his employees. To assess this claim, we should consider the formulation of the epistemic condition on moral responsibility that Vargas takes as a working hypothesis throughout his paper:

- (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.<sup>10</sup>

What makes a prior time “suitable”? Well, we take it that it is a prior time at which the agent must have been able to foresee the outcome in question and also met the control condition with respect to the outcome.<sup>11</sup> So, more specifically, there must be a prior time at which Jeff's firing his employees is reasonably foreseeable and also at which he freely brings it about that this consequence will occur (later). Now, in Jeff's case, there is surely *some* prior time at which his firing his employees in a jerky way is reasonably foreseeable, if only just a few minutes before it happened. But we take it that this is not a *suitable* prior time, since at that point Jeff's jerky character has already

been formed (and so he doesn't *at that point* freely act in such a way that he subsequently fires the employees).<sup>12</sup>

As Vargas points out, the most natural time to which to trace back is during Jeff's adolescence as he deliberates about whether to become a jerk in order to raise his chances of success with the opposite sex. Surely at that point, Jeff was in control of whether he would later fire those employees in a jerky fashion.<sup>13</sup> But of course it is unreasonable for us to think that 15-year-old Jeff should have taken into his deliberations the fact that if he were to become a jerk, he may well lay off future employees in a despicable manner. Since such an outcome is not reasonably foreseeable for Jeff at that point in his life, the most natural place to trace back to won't serve to ground his later moral responsibility, since (KC) would not be satisfied.<sup>14</sup>

We can now see why Vargas thinks the case of Jeff the Jerk satisfies (iii) as well as (i) and (ii), and hence should lead us to worry about tracing. But we're not convinced that there really is no suitable prior time to ground the tracing in question. We agree, of course, that it's unreasonable to expect a 15-year-old to be able to foresee what sorts of specific actions he will perform some ten years or more down the road while engaged in the daily routine of a career that he has no idea at this point he is even interested in pursuing. This would certainly set the epistemic bar too high. But we also see no reason to suppose that this is the only way to ground the tracing in question.

Now consider what "outcome" must be reasonably foreseeable at the prior time in question in order for tracing to succeed. (KC) itself does not tell us just how finely the outcome in question must be specified. If we specify the outcome narrowly (as in our discussion thus far), then it certainly won't have been reasonably foreseeable for Jeff's 15-year-old self.<sup>15</sup> However, if we specify the outcome more broadly, then it's not clear that the outcome won't have been reasonably foreseeable. Indeed, there will be a range of specifications, each more coarse-grained than the previous, and while some will not have been reasonably foreseeable, others will. For a few examples, consider the following ways of specifying the outcome in question:

- (Outcome1) Jeff fires *those* employees who work for *that* company on *that* precise day in *that* precise manner.
- (Outcome2) Jeff fires some of his employees at some company or other at some point in the future in a despicable manner as a result of his jerky character.
- (Outcome3) Jeff treats some people poorly at some point in the future as a result of his jerky character.

As the outcomes become more general, it becomes much more plausible to suppose that they are reasonably foreseeable to Jeff's younger self. While it's extremely implausible to expect Jeff's 15-year-old self to have been able to foresee Outcome1, it doesn't seem at all implausible to suppose that Jeff's

15-year-old self could have foreseen Outcome<sub>3</sub>. So our question now becomes: which outcome does Jeff's 15-year-old self need to have reasonably foreseen in order for us to successfully trace back to that time without violating the epistemic condition on moral responsibility?

We suggest that all tracing requires in this case is that Jeff could have reasonably foreseen Outcome<sub>3</sub> at the time he decides to acquire a jerky character. Why do we hold Jeff responsible for unreflectively firing his employees in such a despicable manner? We hold him responsible partly because he freely decided to become a jerk at some point in the past, and it is reasonable to expect Jeff's younger self to have known that becoming a jerk would in all probability lead him to perform jerky actions. Need Jeff have known that his becoming a jerk would specifically lead to the firing of *those particular employees on that particular day* in the future in order to be morally responsible for firing them in the way he did? Surely not. Need Jeff have known that he would fire *some employees at some point* in the future in order to be morally responsible for firing them when he did? Again, this would likely set the epistemic bar too high. (KC) leaves open just how the outcome in question should be specified, and it seems that there is a perfectly good way of specifying the outcome so that we can successfully trace Jeff's moral responsibility back to the time when he decided to become a jerk. Further, this would seem to be the natural way of understanding our responsibility-practices in general. When you choose to be a jerk, you can be held accountable for your subsequent specific acts of jerkiness—but you might not know in advance what they will be in all their particular glory! Similarly, when you choose to get drunk, you can be held accountable for your subsequent behavior, even though you presumably do not know what it will be in all its florid particularity.<sup>16</sup>

In fairness, Vargas is aware of this sort of response, and more or less concedes in a footnote that the case of Jeff the Jerk needs refinement for precisely this reason. He goes on to suggest a potential refinement.<sup>17</sup> To get around the above problem, Vargas suggests that we amend the story so that Jeff acquires his jerky characteristics “while conceiving of them under a different guise (e.g., being “cool”), while blind to the negative aspects of the acquired trait or characteristic.” So Jeff's 15-year-old self does not think he is becoming a jerk, and hence it is now unreasonable to expect him to have foreseen that he would treat others badly in the future as a result of his jerky character.

But we're not sure how this amendment is supposed to help. First, we simply note that we did not interpret the original case, as presented by Vargas, as requiring that Jeff was aware of the negative aspects of the newly-acquired characteristics; we had already thought of Jeff as simply seeking to be “cool” and being fully oblivious to the dark side (even in the short-term) of what he was doing. Further, whether Jeff conceives of the relevant characteristics as “cool” or “jerky” doesn't seem to change the fact that



acquiring those characteristics will in fact lead to his treating others poorly and that Jeff should have expected that it might do so. When you are a jerk to someone, your behavior is not excused by pointing out that you thought you were just doing the cool thing. It may be cool to be a jerk, but that certainly doesn't excuse jerky behavior. At the very least, we would have to hear more about how this amendment to the case is meant to help.

We think it is fair to take the case of Jeff the Jerk as the most plausible scenario in Vargas's arsenal; Jeff the Jerk is the "poster boy" in the campaign against tracing. Our analysis of the case has issued in the following skeptical conclusions. First, the mere fact that Jeff's relevant behavior is "unreflective" does nothing to show that it does not meet the control condition. Nor would we typically deem agents such as Jeff to be governed by literally irresistible urges to jerkiness (or anything close to irresistible urges). But even assuming that the control condition is really *not* met at the time of the relevant behavior, we have defended the possibility of tracing back to a "suitable" prior time by focusing on whether to define the outcome in question broadly or narrowly. We think that the broader sort of definition is more plausible, and it yields the result that Jeff is indeed morally responsible for his jerky firing of the employees; after all, he could at that "suitable" past time reasonably foresee that his behavior could lead to jerkiness, and he need not be required to envisage the particular details. Thus Jeff can be deemed responsible, and there is no trouble for tracing here.<sup>18</sup>

### 3. Britney the Bride

At this point, let us leave the case of Jeff the Jerk behind and move on to Vargas's other three cases. The next in line is the (strangely familiar) case of 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?<sup>19</sup>

Again, this is the initial description of the case, but Vargas later goes on to fill in some necessary background information:

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.<sup>20</sup>

So goes the case of Britney the Bride. Vargas thinks that upon reading this case, we will have the following three reactions. First, we will judge that Britney is morally responsible for accepting Kevin's marriage proposal. Second, we will judge that due to the dispositions she came to have on the basis of her earlier foray into marriage, she doesn't satisfy the control condition on moral responsibility with respect to accepting Kevin's marriage proposal. These two insights together should lead us to invoke tracing in order to account for Britney's moral responsibility. However, our third reaction should be that when Britney voluntarily agreed to marry her childhood friend, she could not have reasonably been expected to foresee that this action and its consequences would create in her a strong disposition to marry the next attractive professional dancer she came across. Thus, the epistemic condition on moral responsibility will prevent us from tracing Britney's moral responsibility back to her decision to marry her childhood friend.

As we did with Jeff, let's carefully consider each of these reactions. Is Britney morally responsible for accepting Kevin's proposal? Given the way Vargas has described this case, we think it is unclear whether Britney is in fact morally responsible. Of course, we typically think that when someone accepts a marriage proposal, they are morally responsible for having done so. But Britney's case is a bit odd, since it's unclear how large a role the dispositions in question played in influencing Britney's acceptance. Did they produce in her a literally irresistible desire to accept the proposal so that

there was a clear sense in which she couldn't help herself? If so, then it is plausible to suppose that her decision to accept the proposal did not in fact issue from a moderately reasons-responsive mechanism and hence that she is not in fact responsible for it. On the other hand, if the dispositions she acquired from her earlier botched Vegas wedding did *not* produce in her any sort of irresistible desire to accept Kevin's proposal, then why not say that she is morally responsible for accepting, and also that there is no need to invoke tracing?

In order for this case to have any bite, Vargas first needs to secure both the judgment that she is morally responsible for accepting the proposal *and* the judgment that she didn't satisfy any plausible control condition on moral responsibility at the time of acceptance. But we don't see how Britney's case can be described so as to elicit both of these judgments. To the extent that the story of Britney emphasizes the irresistible influence of her dispositions, we will be inclined to excuse Britney from responsibility, and to the extent that the story emphasizes Britney's responsibility, we will start to think that she does satisfy the control condition, after all. Hence it looks like we need not even invoke tracing in this case at all.

It may help in this case to look at what Vargas actually says about Britney's dispositions and their influence on her subsequent action. First, Vargas points out that her accepting the proposal was "spontaneous and unreflective". This alone, however, should not lead us to think that Britney didn't satisfy the control condition when she accepted the proposal. As we pointed out in the case of Jeff the Jerk, there seems to be no reason to suppose that unreflective behavior cannot issue from the agent's own moderately reasons-responsive mechanism, and thus meet the control condition. But Vargas also points out that when she accepts Kevin's proposal, "her love of him was so deep that she could not even conceive of having told him no". Marrying her childhood friend "created in her a powerful disposition" that "left her ready" to accept Kevin's proposal. These descriptions do not help us get a very clear picture of just what is going on when Britney accepts the proposal. She had a powerful disposition, but how powerful? It left her ready, but how ready? She supposedly could not even conceive of declining, but is this meant to be taken literally? Given the unclarity about Britney's precise motivational structure, we suggest that it's equally unclear whether she is morally responsible for accepting Kevin's proposal. Moreover, there does not appear to be an easy way of gaining clarity that will also succeed in sustaining the two intuitions Vargas needs to make his case against tracing.

#### **4. Paulina the Paralyzed**

There are two more cases to consider before our response to Vargas is complete. Next is the case of 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 starts 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?<sup>21</sup>

Vargas fills in the background details as follows:

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.<sup>22</sup>

This is the case of Paulina the Paralyzed. Vargas thinks that upon reading this case, we will have the following three reactions. First, we will judge that Paulina is morally responsible for not doing more to save her son. Second, we will judge that the complex interaction of her surprise with her phobia of large toothy reptiles makes it the case that she doesn't satisfy the control condition on moral responsibility with respect to not doing more to save her son. These two insights together should lead us to invoke tracing in order to account for Paulina's moral responsibility. However, our third reaction should be that there is no plausible prior time at which Paulina could reasonably have been expected to foresee that these facts about her would interact in such a way as to leave her temporarily

paralyzed. Thus, the epistemic condition on moral responsibility will prevent us from tracing Paulina's moral responsibility back to any point in her past.<sup>23</sup>

Once more, it is clear that in order for this case to work the way Vargas intends, *all three* of these reactions must be elicited and sustained. Let's go through them one at a time. First, is Paulina morally responsible for not doing more to save her son? We think the case breaks down even at this first stage. It seems clear to us that given the way Vargas has described the case, Paulina is *not* in fact morally responsible for not doing more to save her son. Would we not excuse her from responsibility upon learning that she was *literally frozen* by her fear and surprise? Indeed, the same problem that haunted the case of Britney the Bride seems to be present here as well. If the case is described with an emphasis on the paralyzing role played by her fear and surprise, we will be inclined to judge her not responsible, and if the case is described with an emphasis on her responsibility, we will be inclined to judge that her fear was not in fact overwhelming and perhaps her omission did issue from a moderately reasons-responsive mechanism after all.<sup>24</sup> So, either we will judge Paulina responsible and in control or else not in control and thus not responsible (on any plausible interpretation of control). Either way, the case does not prove troublesome for tracing, since tracing need not be invoked at all.

Let us suppose for the sake of argument, however, that it's true both that Paulina is morally responsible for the relevant omission and that Paulina fails to satisfy any plausible control condition with respect to that omission. Now we will need to invoke tracing. Is there no plausible prior time to trace back to? Vargas seems to suppose that the most natural way to accomplish tracing in this case is to trace Paulina's responsibility back to a time when she could have undergone therapy to rid herself of her fear of toothy reptiles. He rightly points out, however, that the times when she could have undergone therapy are also times at which she could not have been expected to foresee being in a situation where her fear would play a role in the death of her son. So, this sort of tracing would violate the epistemic condition on moral responsibility.

But we think that, insofar as we are willing to judge Paulina morally responsible at all, there is a much more plausible place to which to trace back. There was no doubt some point prior to her trip to Florida when Paulina could have educated herself about the potential dangers of Florida wildlife. Vargas points out that it was her first trip to Florida and "since she just never bothered to think about it, she did not have any expectations about the native fauna". But surely this was negligent of Paulina. Had Paulina and her son vacationed in Minnesota during the winter, would it not be negligent for Paulina not to have thought about whether she would need to bring a winter coat for her son? Had they gone camping in the California desert and her son had gotten bitten by a rattlesnake because she wasn't expecting to

come across any rattlesnakes, wouldn't this count as some sort of negligence? Insofar as it does seem that Paulina should have known about the dangers of Florida's fauna, we think it is reasonable to suppose that Paulina is to some extent morally responsible for not doing more to save her son. Even if she failed to satisfy the control condition at the time, there was a line of inquiry that she could have (and should have) pursued prior to visiting Florida's parks in the first place.<sup>25</sup>

Note that we think it's plausible to trace Paulina's moral responsibility back to her negligence in not learning more about Florida before visiting *only on the supposition* that she is indeed morally responsible for not doing more to save her son. As we pointed out above, however, we don't think this latter supposition is very plausible in the first place, given the nature of her paralysis.

### 5. Ruben the Unfortunate

Whereas the cases of Jeff and Britney involve responsibility for actions and the case of Paulina involves responsibility for an omission, Vargas's final case appears to involve (most fundamentally) responsibility for a failed attempt. Here is his initial presentation of the case of 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?<sup>26</sup>

Vargas later expands on the story a bit:

[Ruben] 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.<sup>27</sup>

Vargas thinks that upon reading this case of Ruben the Unfortunate, we will have the following three reactions. First, we will judge that Ruben is morally responsible for dropping his niece. Second, we will judge that complex interaction of the booze, the Dolphin win, the presence of his niece, and the habit of throwing her lightly into the air makes it the case that he doesn't satisfy the control condition on moral responsibility with respect to dropping his niece. These two insights together should lead us to invoke tracing in order to account for Ruben's moral responsibility. However, our third reaction should be that there is no plausible prior time at which Ruben could reasonably have been expected to foresee that these facts about him and the circumstance in which he found himself would interact in such a way as to lead to his dropping his niece. Thus, the epistemic condition on moral responsibility will prevent us from tracing Ruben's moral responsibility back to any point in his past.

Unlike the cases of Britney and Paulina, we think it is very plausible to suppose that Ruben is in fact responsible for dropping his niece. It's true that he didn't mean to drop her, but in many cases people are morally responsible for outcomes that are more or less accidental. So we are inclined to grant that this scenario satisfies condition (i). What about (ii)? That is, does Ruben satisfy any plausible control condition on moral responsibility with respect to dropping his niece? To address this question, it may help first to ask exactly what Ruben is responsible for in this case. There are a few different candidates. First, there is his niece's breaking her leg. Second, there is Ruben's failing to catch her on the way down. And third, there is Ruben's throwing her in the air in the first place. The way Vargas describes the case makes it seem as though he is focusing on Ruben's failing to catch her on the way down, so let us consider that possibility.

Now, does Ruben satisfy any plausible control condition with respect to failing to catch his niece on the way down? Given Vargas's set up, it is plausible to suppose that he doesn't. Although he doesn't come right out and say it, a charitable reading of Vargas's case would be to suppose that the mixture of alcohol, the Dolphin win, the presence of his niece, and the standing habit of throwing her in the air, maybe mixed with the extra wiggle she gives in this particular case, all combine to make it the case that Ruben can't help but fail to catch her on the way down. So we'll agree that condition (ii) is also satisfied by this case.

Now the question is whether there is any prior time that can serve as an appropriate terminus for tracing. The most natural place to try to trace back to is to Ruben's throwing her in the air in the first place. Once she was in the air, he couldn't help but drop her, but surely he had control of whether he threw her in the air to begin with. But Vargas argues that this will not do, since at the time he threw her into the air, he could not have reasonably foreseen that he would fail to catch her on the way back down. After all, he threw her up unreflectively and out of habit. One might next try to trace Ruben's responsibility back to the time when he chose to drink the beers at the game, but this move will encounter the same problem. If he couldn't have been expected to foresee dropping his niece at the very time when he threw her into the air, how much more unreasonable would it be to expect him to foresee dropping her when he is in the process of buying beers at the game? After all, he doesn't even think she'll be at home at that point. And the farther away we get from Ruben's dropping his niece, the more difficult it will be to make plausible that the terminus of tracing is an action for which Ruben satisfied the epistemic condition on moral responsibility.

But we think that there is no need to go back very far in order to find a suitable terminus for the tracing. We think it is plausible to trace Ruben's responsibility back to his throwing his niece into the air in the first place. Such behavior, habitual or not, is risky. Surely Ruben knew when he started to acquire this habit that it brought with it certain risks. And while it's not plausible to suppose that he could have foreseen that he *would* drop her in this case, it certainly is plausible to suppose that he could have foreseen that he *might* drop her, and this seems enough to satisfy the epistemic condition.<sup>28</sup> According to (KC), one need not know with certainty that some outcome will occur. The outcome in question must only be *reasonably foreseeable*. And it seems clear that when Ruben throws his niece into the air, his failing to catch her on the way back down is reasonably foreseeable. So we are inclined to trace Ruben's responsibility back to this prior action.

At this point Vargas might respond in two ways. First, he may insist that the outcome in question isn't even reasonably foreseeable at the time when Ruben throws his niece into the air. Why not? Precisely because, as Vargas says in his description of the case, "what is happening is a habitual, undeliberative action". But it's unclear why this should rule out reasonable foreseeability. Although Ruben may not consciously consider it at the time, it's reasonable to suppose that he knows (or should know) that such behavior involves a certain degree of risk, and that there is no way he can *ensure* that his niece will never fall as a result of his throwing her into the air. It certainly doesn't seem a good excuse for Ruben to point out that he wasn't even thinking about the possibility that he might not catch her on the way down. He quite straightforwardly *should* have been thinking about that possibility. Or, at least, that possibility should always have been "in the back of his mind" ever since he started engaging in the risky behavior in the first place.



A second way Vargas might resist is by arguing that Ruben doesn't even satisfy the control condition with respect to throwing his niece into the air, and thus such an action can't serve as an adequate terminus of tracing. But why think this? The only reason it seems Vargas could appeal to is, again, the fact that his throwing her into the air is "a habitual, undeliberative action". But as we have already pointed out in our treatment of previous cases, we don't see any reason to think that habitual actions cannot meet the control condition.<sup>29</sup>

## **6. Vargas's Cases: Summary of Our Critique**

This completes our examination of Vargas's four cases that are supposed to spell trouble for tracing. To briefly rehearse our conclusions, we have agreed that Jeff the Jerk is indeed responsible. The first point to note about Jeff is that mere unreflectivity does not entail that Jeff fails to meet the control condition. And we are skeptical whether the typical jerk—or even most atypical jerks—really fails to meet the control condition. But assuming that Jeff really is not in control of his jerky behavior (at the time of the behavior), there is still a way of tracing his responsibility back to his decision to become a jerk. We need only to specify the relevant outcome broadly. A proper analysis of Jeff the Jerk—and related cases, such as McKenna's young football player—shows that they do not pose intractable troubles for an approach to moral responsibility that has a tracing component.

We argued that Britney the Bride either satisfies the control condition with respect to accepting Kevin's proposal and is therefore responsible, or else she does not satisfy the control condition and is therefore not responsible. We can't see any way of mending the case so that it will elicit both the intuition that she is responsible and that she fails to satisfy the control condition. Paulina the Paralyzed seems straightforwardly excused from responsibility, given the way Vargas described her case, but we argued that insofar as we are willing to hold Paulina responsible, this is because of her negligence in not familiarizing herself with the Florida wildlife. Finally, we argued that although Ruben is indeed responsible for dropping his niece despite his not satisfying the control condition, his responsibility can be naturally traced back to his throwing his niece into the air in the first place. We conclude that none of Vargas's cases requires us to use tracing in a way that runs afoul of the epistemic condition on moral responsibility. The truth about tracing is that its troubles are not as severe as Vargas has suggested, if they are even troubles at all.

## **7. Smith's Cases**

Before closing, we'd like to consider some relevant and very intriguing recent work by Angela Smith.<sup>30</sup> Although Smith does not present a critique of

tracing *per se*, some of the examples she adduces in favor of her position seem to presuppose such a critique. It will be illuminating to see how a proponent of tracing might reply to Smith's cases.

Smith argues against accounts of moral responsibility that she dubs 'volitional'. According to volitional accounts,

what ultimately makes an attitude attributable to a person for purposes of moral assessment is that it is connected in some way to her choices: she made choices in the past which led to the development of the attitude in question, she made a choice in the present to endorse or "identify with" it, or she has the ability to modify her attitudes through the choices she makes in the future.<sup>31</sup>

Smith suggests that most accounts of moral responsibility on the market are volitional, since they reserve a special role for the notion of *choice*. Smith maintains, however, that such views cannot do justice to our actual practices of holding each other morally responsible. In particular, Smith puts forward a number of ordinary examples in which we apparently are inclined to hold someone morally responsible for something that was neither directly the result of any choice nor traceable to any choice in the past or the future.

Rather than making moral responsibility primarily a matter of *choice*, as volitional accounts do, Smith instead argues in favor of a view that makes moral responsibility primarily a matter of *evaluative judgment*. She calls her view 'the rational relations view' and according to this view, "a mental state is attributable to a person in the way that is required in order for it to be a basis for moral appraisal if that state is rationally connected in one of the relevant ways to her underlying evaluative judgment".<sup>32</sup> Smith maintains that her view can account for all the cases the volitional view can account for, *as well as* the cases that apparently cause trouble for a volitional view.

Part of the reason that Smith rejects a volitional view of moral responsibility, it seems, is precisely that the idea of tracing is seemingly impotent when it comes to explaining certain important cases. Since we are attracted to a view that invokes tracing, her cases provide a powerful challenge that is worth addressing. So, let's begin by considering one of her cases:

I forgot a close friend's birthday last year. A few days after the fact, I realized that this important date had come and gone without my so much as sending a card or giving her a call. I was mortified. What kind of a friend could forget such a thing? Within minutes I was on the phone to her, acknowledging my fault and offering my apologies. But what, exactly, was the nature of my fault in this case? After all, I did not consciously *choose* to forget this special day or deliberately *decide* to ignore it. I did not *intend* to hurt my friend's feelings or even *foresee* that my conduct would have this effect. I just forgot. It didn't occur to me. I failed to notice. And yet, despite the apparent involuntariness of this failure, there was no doubt in either of our minds that I was, indeed, responsible for it. Although my friend was quick to pardon my thoughtlessness

and to dismiss it as trivial and unimportant, the act of pardoning itself is simply a way of renouncing certain critical responses which it is acknowledged would, in principle, be justified.<sup>33</sup>

Smith is surely right to point out that we encounter cases like this all the time. To constitute an objection to volitional accounts of moral responsibility, though, the case needs to exhibit the following features:

- (i) Smith is morally responsible for forgetting her friend's birthday.
- (ii) Smith's forgetting was neither the direct result of a choice nor traceable to any past choice.

If some volitional view of moral responsibility is correct, then (i) is true only if (ii) is false. So if the birthday case exhibits both (i) and (ii), then no volitional view of moral responsibility can be correct.

Smith contends that what explains her moral responsibility in the birthday case has nothing to do with any choice of hers, but rather with how her forgetting reflects her evaluative judgments. As she puts it,

If we value something and judge it to be worth promoting, protecting, or honoring in some way, this should (rationally) have an influence on our unreflective patterns of thought and feeling. We commonly infer from these unreflective patterns, or from their absence, what a person really cares about and judges to be important.<sup>34</sup>

Since she forgot her friend's birthday, it is reasonable to infer that she doesn't care about her friend as much as a friend should, and it is this fact that best explains why we are inclined to hold her morally responsible for forgetting. Were her forgetting the result of hypnosis, on the other hand, we would of course excuse her from moral responsibility, but again Smith maintains this has nothing to do with any lack of choice. Rather, what explains our willingness to excuse her in the hypnosis case is the fact that her forgetting does not plausibly reflect her evaluative judgments.

While we certainly agree with Smith that whether a bit of behavior or an attitude reflects one's evaluative judgments is an important part of explaining one's moral responsibility for the relevant behavior or attitude, we think that this insight can be incorporated into a certain sort of tracing-based account of moral responsibility. It is crucial here to distinguish a "volitional" approach to moral responsibility from a related but different approach: a "control" approach to moral responsibility. Whereas a volitional approach requires that we trace back (directly or indirectly) to a *choice* in order to justify attributions of moral responsibility, a control model requires that we back (directly or indirectly) to an exercise of control, where such an exercise of control might be a choice, and action, or an omission. In the birthday

case, in particular, we think that it is plausible to suppose that Smith's moral responsibility for forgetting does in fact trace back to some previous free action or omission. If we are right about this, then the birthday case will not after all tell against control-based accounts of moral responsibility.

To see why we think some sort of tracing must be going on in the birthday case, consider the fact that Smith failed to choose to do various things which were such that, had she so chosen, she would have had a better chance of remembering her friend's birthday. So, for example, perhaps Smith omitted to choose to put her friend's birthday on her calendar, or perhaps Smith failed to choose to set up her "email client" to alert her to her friend's birthday, and so forth. Part of what it is to be a good friend is to take these steps to minimize the likelihood that you will forget your friends' birthdays. If you don't take these steps, and then forget, you are legitimately morally assessable for your forgetting *precisely because* you failed to do something to make your forgetting much less likely (and you were in control of this failure). This is just to say that your moral responsibility for forgetting *traces back* to these past omissions. Indeed, we would argue that your forgetting reflects a poor evaluative judgment only insofar as you failed to take the necessary steps that any friend would take to remember friends' birthdays.

Now Smith might respond by adding details to the case so that she *did* in fact take all reasonable precautions with regard to remembering her friend's birthday, but then forgot anyway. Perhaps she did put her friend's birthday on her calendar, she did tell her email client to remind her, and she even put a note on her bathroom mirror! We can suppose that, despite all her efforts, she still simply forgot. Surely in this case there is no plausible place to serve as terminus of any sort of tracing. However, in our view, these added details serve to etiolate (and, arguably, to eliminate) the intuition of moral responsibility that made the case so plausible in the first place. Surely if it's true that Smith did everything she could do to make forgetting her friend's birthday less likely, then she isn't responsible if she accidentally forgets anyway. If Smith were to explain to her friend how much care she put into trying to remember the birthday, and that despite her best efforts, she simply forgot, we maintain that her friend would no longer be justified in holding her responsible for forgetting. (Presumably the case is still under-described, and the friend is still likely to wonder why Smith was forgetful, despite all her efforts to make sure she would remember the birthday; here we assume that the example can be filled in so that it is clear that *all* reasonable efforts were made by Smith to ensure that her memory would be triggered, that she wouldn't have a tendency toward forgetfulness—maybe she has attended a special self-improvement seminar for memory—and so forth.) Perhaps it is only if her forgetting reflects her evaluative judgments that her friend can hold her responsible, but the best way to determine whether her forgetting reflects her evaluative judgments is to see whether it can be traced back to some past free action or omission. If we are right about this, then

the birthday case does not tell against tracing-based control models of moral responsibility after all.

To be fair to Smith, we do not disagree with her rejection of volitional accounts of moral responsibility. It might seem that what is problematic about such accounts is their tracing component. But we have separated the volitional idea from the tracing idea; we have argued that if we replace volition with a broader notion of control, where the pertinent sort of control can be exercised via an omission as well as (say) a choice, one can give an adequate and, indeed, a natural account of Smith's case above. Thus it is not the tracing component, but the volitional element, that is problematic. So, whereas we agree with Smith's rejection of (ii), we would urge replacing (ii) with:

- (ii\*) Smith's forgetting was neither the direct result of an exercise of control nor traceable to any past exercise of control.

Note that not only can (ii\*) give an appealing analysis of Smith's example above, but it can help to explain a range of difficult "manipulation cases" – cases in which an agent's evaluative judgment is induced in some manipulative manner (such as via clandestine advertising, hypnosis, or even direct stimulation of the brain). Smith does have something interesting to say about such cases, but it is not clear that her rational relations model can adequately handle all such cases.<sup>35</sup>

Let's consider another example presented by Smith, just to make our strategy of response clearer. This example comes from a passage in George Eliot's *Scenes of a Clerical Life*:

When the sweets were on the table, there was a mould of jelly just opposite Captain Wybrow, and being inclined to take some himself, he first invited Miss Assher, who coloured, and said, in rather a sharper key than usual, "Have you not learned by this time that I never take jelly?" "Don't you?" said Captain Wybrow, whose perceptions were not acute enough for him to notice the difference of a semitone. "I should have thought you were fond of it. There was always some on the table at Farleigh, I think." "You don't seem to take much interest in my likes and dislikes." "I'm too much possessed by the happy thought that you like me," was the *ex officio* reply, in silvery tones.<sup>36</sup>

Here it is plausible to suppose that Captain Wybrow is morally responsible for his failure to notice that Miss Assher dislikes jelly. It is also plausible to suppose that his failure to notice this is not the direct result of any choice on his part—that is, he didn't *choose* to fail to notice. Smith then argues that the best way to account for this sort of an example is again in terms of Wybrow's evaluative judgments. The fact that Wybrow fails to notice Assher's likes and dislikes is a good indication that he does not care about her as much as

he should, and hence Assher holds Wybrow morally responsible. Again, it seems at first as though a volitional account of moral responsibility will not be able to handle a case like this.

But as above, we think that the best analysis of this case will invoke tracing. To be sure, the fact that Wybrow fails to notice Assher's likes and dislikes does indicate that he does not care about her as much as he should. But, we would argue, one would intuitively hold Wybrow responsible here only if there was something in the past that he did freely (or omitted freely) that led him to be the sort of lover that fails to notice things that he should in fact notice. Perhaps he chose to engross himself in his own affairs rather than think about the needs and wants of Miss Asher, perhaps he omitted asking her certain questions about her emotional life when the opportunity arose, and so on. If there is nothing in the past that he should have done that would have made him much more likely to notice Miss Asher's likes and dislikes, then we would argue that he can't properly be held responsible. But what's more likely is that there were times at which his free actions (or omissions) more or less secured his emotional indifference to her. Again, although his failure to notice her likes and dislikes indicates a poor evaluative judgment, this evaluative judgment does not seem to be properly *his own* evaluative judgment unless we can complete a trace back to some free action (or omission).<sup>37</sup>

Smith puts forward a number of nuanced considerations for the conclusion that the rational relations view provides a better understanding of our practices of moral responsibility than does the volitional view. We agree with her critique of the volitional approach, but we do not agree that one should conclude that a control-based approach with a tracing component is to be rejected. Indeed, such a view gives a natural analysis of Smith's examples, and it also has something plausible to say about a range of manipulation cases. Presumably, a more refined and comprehensive analysis of Smith's cases will invoke the considerations we presented in discussing Jeff the Jerk (and related cases) above: the outcome may be defined broadly and tracing can apply simply because it is reasonable to think that the agent can foresee outcomes that are broadly specified. But perhaps it is interesting to note that, whereas it is plausible to trace back to certain *choices* in the case of Jeff the Jerk, it is more plausible to trace back to certain omissions—certain *failures to choose*—in the cases offered by Smith. Since choices (and, in general actions) tend to be more salient than omissions, it is perhaps a bit more difficult to see how tracing can be applied to Smith's examples than to those of Vargas.<sup>38</sup> But the truth is that tracing can indeed be applied.

## 8. Conclusion

Much is at stake in evaluating a critique of tracing. We do not see how a theory of moral responsibility could adequately handle the range of

drunk-driving cases, “Martin Luther cases”, and manipulation cases without some sort of tracing component; tracing just seems both highly plausible and theoretically indispensable. Indeed, theorists as diverse as the libertarian Robert Kane and the “semicompatibilists” John Martin Fischer and Mark Ravizza have embraced tracing as a component of their theories of moral responsibility.

If tracing cannot be vindicated, this would seem to be a devastating challenge to such theories as Kane’s and Fischer’s—and *any* control-based theory of responsibility. Here we have offered a template for a strategy of defense of tracing against the sorts of examples adduced by Vargas, McKenna, and Smith. A defense of the plausibility of tracing is part of a defense of a plausible control-based account of moral responsibility. Such accounts will need to take seriously the spirit (if not the precise content) of Lady Chiltern’s remark in Oscar Wilde’s *An Ideal Husband*: “One’s past is what one is. It is the only way by which people should be judged”.<sup>39</sup>

### Notes

<sup>1</sup> Vargas 2005. In a recent article (McKenna, 2008, p. 34, fn. 6), Michael McKenna says: “The point I will set out [developing some putative troubles with tracing] is fully and impressively developed with striking results in Manuel Vargas’s recent, “The Trouble with Tracing,” (2005)”.

<sup>2</sup> Smith 2005.

<sup>3</sup> Vargas 2005, p. 271. Michael McKenna develops a case similar to that of Jeff the Jerk:

Consider the young boy who consciously chooses to cultivate his aggression and thick skin in order to survive in the locker room during his junior varsity football days. He will have very little reason to expect that it will someday be the source of his coolness and tragic distance from his own children. In general, so far as various aspects of our own character traits are concerned, we are often the hapless victims of our own unwitting earlier free choices. We do not possess the foresight of gods, and so who we will become, from the vantage point of who we are, is sometimes just a crap shoot. (McKenna, 2008, p. 33).

<sup>4</sup> Vargas 2005, pp. 275–6.

<sup>5</sup> Vargas agrees: Vargas 2005, p. 270.

<sup>6</sup> Fischer and Ravizza, 1998; Fischer and Ravizza call this sort of control “guidance control”.

<sup>7</sup> In particular, the mechanism must display regular receptivity and at least weak reactivity with respect to a range of reasons that include moral reasons. For details, see Fischer and Ravizza 1998, chapter 3.

<sup>8</sup> For the details, see Fischer and Ravizza 1998, pp. 85–89.

<sup>9</sup> Note that, as we stated above, nothing here depends on adoption of the “acting freely” interpretation of control, or this specific account of acting freely; we could simply note that there is nothing in the case, as described by Vargas, that should preclude Jeff’s refraining from firing the employees. After all, the mere fact that behavior is unreflective does not entail that the agent could not have done otherwise.

<sup>10</sup> Vargas 2005, p. 274. Vargas labels this condition ‘KC’ for ‘Knowledge Condition’.

<sup>11</sup> Again, we have adopted the “acting freely” interpretation of the control condition; but one could just as well adopt the “could have done otherwise” interpretation.

<sup>12</sup> Perhaps another way of putting the point is that at the prior time in question, Jeff does not freely bring it about that he later fires the employees in a jerky manner; he does not freely bring about this consequence. For an account of the relevant sort of control of consequences (on which Jeff lacks control of bringing it about that he later fires the employees in a jerky manner, see Fischer and Ravizza 1998, chapter 4.

On the “could have done otherwise” interpretation of the control condition, we trace back to a prior time at which it is both true that the agent could reasonably foresee the outcome in question and also has it in his power to do something that would have prevented the outcome from occurring. On this sort of interpretation, we’ll have to look for a time at which Jeff could both foresee the outcome in question and could have still done something to prevent it from occurring.

<sup>13</sup> For example, to put the point in terms of the “could have done otherwise” interpretation of the control condition, he was at that point able to do something—namely, decide against becoming a jerk—that would have prevented his much later action of firing his employees from occurring.

<sup>14</sup> Again, nothing hangs on the adoption of a particular interpretation of the control condition—so long as it is plausible. We could have just as easily pointed out above that at the prior time he freely did something that brought it about that later he fired the employees in a jerky fashion.

<sup>15</sup> For simplicity’s sake, sometimes we will use locutions such as “Jeff’s 15-year-old self”. We are not seeking to make any special point about personal identity; rather, we use this term as shorthand for “Jeff when he is 15 years old”.

<sup>16</sup> One might be worried that if we are allowed to specify outcomes broadly for tracing purposes, then most people will turn out to be morally responsible for too much. After all, everyone ought to be able to foresee that they might inadvertently offend someone at some point in their lives! While we are inclined to think that our approach does allow for moral responsibility in such contexts, we do not think that this is a problem, since it may well be that you are morally responsible for inadvertently offending someone even though you are not *blameworthy* for doing so. What sort of epistemic requirement is a condition for blameworthiness is an important and interesting question that we do not take up here.

<sup>17</sup> Vargas 2005, p. 277, n. 12.

<sup>18</sup> Our analysis should apply also to McKenna’s example of “the young boy who consciously chooses to cultivate his aggression and thick skin in order to survive in the locker room during his junior varsity football days” and to a whole class of similar examples.

<sup>19</sup> Vargas 2005, pp. 271–2.

<sup>20</sup> Vargas 2005, p. 279.

<sup>21</sup> Vargas 2005, p. 272.

<sup>22</sup> Vargas 2005, p. 281.

<sup>23</sup> Note that whereas the cases of Jeff and Britney involved responsibility for a particular *action*, the case of Paulina is supposed to involve responsibility for a particular *omission*. We are invited to have the intuition that Paulina is morally responsible for *not doing more to save her son*. Whereas this is indeed a difference between the cases, it’s unclear whether it makes a difference, so we will just flag it for now and move on.

<sup>24</sup> On how omissions can issue from moderately reasons-responsive mechanisms, see Fischer and Ravizza 1998, chapter 5. Here again, nothing depends on accepting the “acting freely” interpretation of the control condition, or any specific account of acting freely. On the “could have done otherwise” interpretation, one would say that Paulina is morally responsible for not doing more to help her son insofar as she could have done more to help her son.

<sup>25</sup> Suppose the case were amended so that Paulina and her son happen upon an alligator while walking the streets of downtown Manhattan (perhaps it escaped from a zoo nearby). We would argue that it is even clearer in this case that Paulina is not responsible for not doing more to save her son. Insofar as this case is a clearer case of non-responsibility, this seems to show



that what may have been leading us to judge Paulina responsible in the first case was the fact that she and her son were in Florida, and it is reasonable to think that people who visit Florida should know something about the wildlife there.

<sup>26</sup> Vargas 2005, p. 272.

<sup>27</sup> Vargas 2005, pp. 281–2.

<sup>28</sup> Of course, bare possibility cannot be the relevant notion of “might”, if tracing is to be plausible; this is a delicate matter, which we put aside here insofar as it is not the focus of Vargas’s critique. A fully adequate and completely reductive account of tracing would have to address the issue of interpreting the “might” in question.

<sup>29</sup> For example, we do not see why they cannot issue from an agent’s own, moderately reasons-responsive mechanism. (Alternatively, one might point out—as above—that the mere fact that an action is habitual does not rule it out that the agent could have done otherwise.) And if these contentions are right, then Ruben’s throwing his niece into the air may well satisfy the control condition on moral responsibility after all.

<sup>30</sup> We will focus on Smith 2005, in particular.

<sup>31</sup> Smith 2005, pp. 238–239.

<sup>32</sup> Smith 2005, p. 262.

<sup>33</sup> Smith 2005, p. 236.

<sup>34</sup> Smith 2005, p. 247.

<sup>35</sup> Smith says:

One might well ask what it is in these “implantation” cases that makes the judgments in question “not the agent’s own.” And one might also wonder how these cases differ from cases of irrationality, in which the judgments reflected in a person’s attitudes run counter to some of the other judgments she claims sincerely to accept. What differentiates implanted attitudes from these others, in my view, is that such attitudes are not based upon the agent’s own evaluative appraisal of her situation and surroundings but are induced in a way that bypasses her rational capacities altogether. They do not, therefore, reflect the agent’s own evaluative judgments and commitments. Even in cases of irrationality, I would argue, the irrational attitude reflects a genuine tendency on the part of the agent to see certain considerations as counting in favor of certain responses. Implanted attitudes, by contrast, tell us nothing about the agent’s rational or evaluative dispositions. I see no other way of giving content to the expression “the agent’s own” here, however, except in a way which makes reference to the very network of beliefs and attitudes which I am suggesting ground our attributions of responsibility. This may seem to make this account of responsibility objectionably circular and uninformative, but I think that it is what makes it specially plausible and attractive: the seeming circularity is itself a reflection of what many philosophers have referred to as the “holistic” character of the mental. (Smith 2005, p. 262)

This passage raises many difficult issues. We certainly do not have a decisive argument that Smith’s way of handling manipulation cases cannot work for all such cases. We do however wish to raise the concern that there are such cases in which the relevant rational capacities are *not* bypassed in the way suggested by Smith, but in which the agent is nevertheless intuitively considered to be *not* morally responsible. This would involve thoroughgoing manipulation of the mind, including the comprehensive network of evaluative beliefs and attitudes. Of course, such examples are contentious, and it is perhaps difficult for *any* theory of moral responsibility to give a plausible account of them.

<sup>36</sup> As quoted in Smith 2005, p. 242.

<sup>37</sup> An anonymous referee points out the following distinct interpretation of this case. Perhaps Miss Assher is not holding Wybrow responsible for failing to take note of her likes and dislikes and is instead holding him responsible for *pretending* to care about her when he doesn’t really. This interpretation would also support our contention that Wybrow is

responsible only insofar as there is something he did freely onto which his responsibility can latch.

<sup>38</sup> Christopher Franklin similarly argues that the difference between tracing back to actions and tracing back to omissions helps to understand the initial appeal of “Buffer Zone” versions of the Frankfurt-type cases. See his unpublished paper, “Another One Bites the Dust: Why Buffer Cases Fail to Refute the Principle of Alternative Possibilities”. He contends that it is difficult at first to see that one should apply tracing to such examples because the relevant item—the item to which one traces back—is always an omission in such cases, rather than an action (as in the drunk-driving cases). For a buffer zone version of the Frankfurt-type cases, see Pereboom 2001, pp. 20–21.

<sup>39</sup> We have read versions of parts of this paper to the philosophy departments at: the University of Notre Dame; the University of Calgary; the University of California, Irvine; and the University of Washington. We are very grateful to the members of those departments for their helpful comments. Also, thanks very much to the following for their comments on previous drafts of this paper: Manuel Vargas, Michael McKenna, and Angela Smith.

### References

- Fischer, John Martin and Ravizza, Mark. (1998) *Responsibility and Control: A Theory of Moral Responsibility*. (Cambridge: Cambridge University Press).
- Kane, Robert. (1996) *The Significance of Free Will*. (Oxford: Oxford University Press).
- McKenna, Michael. (2008) “Putting the Lie on the Control Condition for Moral Responsibility”, *Philosophical Studies*, 139 no. 1: 29–37.
- Parfit, Derek. (1984) *Reasons and Persons*. (Oxford: Clarendon Press).
- Pereboom, Derk. (2001) *Living Without Free Will*. (Cambridge: Cambridge University Press).
- Smith, Angela. (2005) “Responsibility for Attitudes,” *Ethics*, 115: 236–271.
- Vargas, Manuel. (2005) “The Trouble With Tracing,” *Midwest Studies in Philosophy* 29: 269–291.