

Erkenn  
DOI 10.1007/s10670-014-9693-y

ORIGINAL ARTICLE

# More Trouble with Tracing

Seth Shabo

Received: 11 February 2014 / Accepted: 25 October 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** Theories of moral responsibility rely on tracing principles to account for derivative moral responsibility. Manuel Vargas has argued that such principles are problematic. To show this, he presents cases where individuals are derivatively blameworthy for their conduct, but where there is no suitable earlier time to which their blameworthiness can be traced back. John Martin Fischer and Neal Tognazzini have sought to resolve this problem by arguing that blameworthiness in these scenarios can be traced back, given the right descriptions of these agents' later conduct. I contend that this strategy may succeed against Vargas's particular examples, but that it fails to resolve the larger problem. After clarifying some key issues about derivative responsibility and tracing principles, I develop a case that isn't amenable to Fischer and Tognazzini's treatment. I then suggest the outlines of a compromise solution to the problem for tracing principles.

## 1 Introduction

We normally think it unfair to blame people for what they do while morally incapacitated. A notable exception is when the incapacity results in a foreseeable way from the agent's own free actions. In the classic example, someone is morally blameworthy for her impaired decision to drive because it was a foreseeable consequence of her unimpaired decision to drink. To accommodate such cases, cases of "indirect" or "derivative" blameworthiness, theorists of moral responsibility standardly employ a tracing principle. Such a principle specifies when something done at an earlier time exposes us to legitimate blame for what we do

---

S. Shabo (✉)

Philosophy Department, University of Delaware, 24 Kent Way, Newark, DE 19716, USA  
e-mail: [sshabo@udel.edu](mailto:sshabo@udel.edu)

later, when we would otherwise be exempt from blame because our judgment is seriously impaired.

In an influential article, Vargas (2005) has identified an apparent problem for tracing principles. For such principles to apply, Vargas observes, there must be a suitable earlier time to which the agent's moral responsibility can be traced back. Such a time will be one at which (a) the agent exercised the control required for moral responsibility, and (b) the relevant outcome was a foreseeable consequence of how that control was exercised. Vargas presents several examples in which agents seem to be derivatively blameworthy for their conduct without satisfying (b). If he is right about these cases, that is bad news for tracing principles, and, by extension, for theories that rely on them. For without invoking derivative responsibility, these theories will have trouble accommodating agents' blameworthiness in cases like that of the intoxicated motorist. And without tracing principles' *restrictions* on derivative responsibility, they will have trouble explaining why another motorist—one whose impairment was an unknown side effect of a new medication, say—isn't blameworthy for his impaired decision to drive.<sup>1</sup>

Here I examine Fischer and Tognazzini's (2011) reply to Vargas's challenge.<sup>2</sup> According to Fischer and Tognazzini, careful scrutiny of Vargas's examples shows that there is no real problem for tracing principles. Even in Vargas's most seemingly cogent case, they argue, there is a suitable earlier time to which the agent's blameworthiness can be traced back; for under the right description, the later action is a foreseeable outcome of the free decision that led to it.<sup>3</sup>

As will be seen, Fischer and Tognazzini's response relies heavily on a particular feature of Vargas's example, namely that the agent decides to become a certain way. While some cases of derivative responsibility may be like this, others appear not to be. Thus, even if Fischer and Tognazzini's response succeeds against Vargas's particular example, it won't resolve the larger issue Vargas has raised.

<sup>1</sup> Vargas is officially agnostic about which conclusion we should draw: that tracing principles should be rejected or that some of our considered judgments of blameworthiness are false.

<sup>2</sup> This essay is an updated and expanded version of Fischer and Tognazzini (2009).

<sup>3</sup> For recent discussions of these issues, see Timpe (2011) and Khoury (2012). Khoury (p. 193) proposes that we resolve the puzzles associated with derivative responsibility by denying that we are indeed blameworthy for the consequences of our actions. In Khoury's view (pp. 200–01), consequences are relevant to blameworthiness only insofar as they provide evidence of the agent's quality of will in doing something that had (or might have been expected to have) those consequences. This allows him to hold that the agent is equally blameworthy for her actions whether or not the untoward consequence actually comes about, and thus to avoid the problem of "resultant" moral luck (pp. 195–97). While I find this proposal thoughtful and intriguing, it seems to me that it doesn't cover all the important cases. This is because, in denying that we are blameworthy for consequences, Khoury restricts the denotation of 'consequences': "'Consequence' will be taken to mean an event or state of affairs (causally related in the appropriate way to an action of an agent) under a description that makes no reference to the mental states of the agent in acting' (p. 197). This means that his account won't apply to outcomes such as Brett's *deciding* to drive while intoxicated, and it seems clear that we need to account for derivative responsibility for impaired decisions (or the fact that such a decision is made), as well as for the further consequences of those decisions.

To make the discussion manageable, I focus on a single case.<sup>4</sup> This case differs from Vargas's in two related ways. First, there is no pivotal episode or stage during which the risk of becoming the relevant way—the way that's implicated in the later, bad outcome—should be salient to the agent. Second, the relevant way the agent becomes isn't an obvious continuation of his formative history. Given these differences, it will be harder to identify an earlier time to which the later outcome is related in the way that Fischer and Tognazzini require. Pending a response to this case and others like it, the trouble with tracing remains, and claims to have resolved it are premature.

As I go on to suggest, however, there may be room for a compromise solution. If the scope of tracing principles can plausibly be restricted to a subset of derivative responsibility cases (of which drunk-driving scenarios are typical), we can allow that other cases of derivative responsibility (of which character-forming scenarios are typical) are exempt from the foreseeability requirement. If such a compromise is tenable, we need not choose between tracing principles *tout court* and our firm judgments about who is blameworthy; and so we can resolve Vargas's original worry without denying the force of his challenge. I don't know whether this compromise will ultimately succeed. My aim here is to chart the space it would have to fill. To this end, I suggest that there could be such things as *untraceable responsibility transmissions*. These would occur when someone is derivatively blameworthy for the cumulative effects of her free activities over some considerable span, without there being any particular time at which these cumulative effects were reasonably foreseeable.

After setting out Vargas's challenge in Sect. 2 below, I examine Fischer and Tognazzini's response in Sect. 3. I then renew the challenge in Sects. 4 and 5. In Sect. 6, I anticipate a reply to the renewed challenge. I suggest the outlines of the compromise solution in Sect. 7.

<sup>4</sup> It should be noted that other authors have suggested promising sketches for scenarios involving derivative moral responsibility without foreseeability. Thus Adams (1985, p. 14) writes: "The morally imprudent voluntary omissions, for example, by which a person has failed to pay the price to extricate himself in time from a situation that has left him embittered, cynical about morality, and full of racist sentiments, may be less gravely blameworthy than the attitudes to which they have led. Indeed, we might think them blameless, a successful gamble, if the sequel had not left the person so corrupted." As this passage makes clear, Adams's concern is to show that someone's blameworthiness for his attitudes is not a function of his blameworthiness for the voluntary actions or omissions that led to those attitudes [see also Smith (2005 and 2008)]. If, however, we can plausibly suppose that the "gamble" was one whose true moral risks the agent could not have been expected to foresee, such a case, suitably developed, could support a challenge KC. More recently, McKenna (2008) has provided a template for another such challenge (see note 13 below). It should be noted that KC presupposes a broadly "volitionalist" view of moral responsibility, on which (roughly) the actions, traits and attitudes, and outcomes for which we are fundamentally morally responsible are ones that are subject to our voluntary control. The "attributionist" alternative, endorsed by Smith, Adams, and others is to say (roughly) that what is fundamental for assessments of moral responsibility is whether an attitude or trait can be attributed to us in such a way that we can be morally appraised for it, where the answer depends on its being suitably related to our other attitudes, and not on its being voluntarily formed or maintained. For the purposes of discussing KC, I shall adopt such a broadly volitionalist view.

## 2 Vargas's Challenge

Tracing moral responsibility back to an earlier time requires that the agent meet two conditions at that time. First, she must exhibit the right kind of control over her actions then, the kind required for moral responsibility. While there is much controversy about what such control consists in, and whether its existence is compatible with the truth of causal determinism, there are many ordinary circumstances that uncontroversially impair it, including extreme intoxication and various psychological disabilities and disorders. For present purposes, we may suppose that the control condition is met when and only when none of these circumstances (and no other serious impairment) is present. Second, the relevant outcome must be reasonably foreseeable as a consequence of this exercise of control. Vargas's challenge pertains to this second condition.

As Vargas notes, there is room for disagreement about how strong to make this condition.<sup>5</sup> How likely must a particular outcome be to qualify as reasonably foreseeable? To show that the problem doesn't hinge on how we answer this question, Vargas employs a fully generic statement of the knowledge (or foreseeability) condition,<sup>6</sup> one to which all tracing theorists are presumably committed:

(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 (Vargas 2005, p. 274).

I shall focus on the first of the four cases that Vargas brings against KC, the case of Jeff the Jerk. Here is Vargas's description of the later action for which Jeff is alleged to be derivatively blameworthy:

Jeff is a middle-aged middle manager in a midsize 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 a jerk. He is rude and inconsiderate about the feelings of others. And he is unreflective about it. When people react poorly to his

<sup>5</sup> Ginet (2000) provides an influential discussion of the foreseeability requirement. For some other influential discussions of tracing principles in the free-will debate, see Fischer and Ravizza (1998), ch. 7, Dennett (1984), pp. 131–33 and Kane (1996), pp. 38–40, 77–78. For a recent, more extensive treatment of the knowledge condition for moral responsibility, see Sher (2009).

<sup>6</sup> While Vargas seems to offer KC as a statement of the "knowledge" or "epistemic" condition for moral responsibility in general (Vargas, op. cit., pp. 273–74), this principle is more plausibly understood as a statement of the knowledge condition for *derivative* moral responsibility. If, for example, you take my book, which you (non-culpably) mistake for your copy, you're not blameworthy for this act because you don't meet the knowledge condition in performing it. But your failure to meet this condition isn't a matter of whether this occurrence was foreseeable at some *suitable prior time*. Thus, I submit, KC should be understood as a statement of the knowledge condition specifically for derivative responsibility (or as a *foreseeability* condition). However, nothing in what follows hinges on this.

behavior...he always writes it off as a shortcoming on the part of the others. One afternoon, his superiors tell him that he needs to give notice to a group of longtime employees that they will be laid off. He does tell them, but in an altogether rude and insensitive fashion. (p. 271)

As to why Jeff isn't directly blameworthy for firing the workers in this manner, Vargas continues, this is because Jeff has an impaired sensitivity to the relevant moral considerations when he acts, so that he doesn't meet the control condition for moral responsibility. As to why he is nonetheless derivatively blameworthy, this is because his impairment is a consequence of his earlier, free actions.

To explain the second claim, Vargas describes a pivotal chapter in Jeff's biography (pp. 275–76). In Vargas's story, it is for misguided, adolescent reasons that the 15-year-old Jeff embarks on a plan for "self-improvement," which is aimed at cultivating "behaviors and attitudes that we would perceive to be jerklike" (p. 275). Vargas adds that the "initial decision to undertake the program of self-improvement was not an obvious choice for Jeff," who was worried about what others would think and whether they would ridicule him for his pretense. Vargas continues:

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 jerklike. 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 jerklike to being a full-on jerk. (p. 276)

To repeat, Jeff isn't directly blameworthy for how he fires the workers because he is morally impaired by the time this episode occurs. Fischer and Tognazzini maintain—reasonably, in my view—that Vargas doesn't say enough about how to understand Jeff's supposed impairment (Fischer and Tognazzini 2011, pp. 211–12). However, I shall follow them in granting for discussion's sake that Jeff doesn't meet the control condition when he fires the workers because his doing so exhibits an impairment-level lack of moral sensitivity. As I see it, the crucial claim for the purposes of the example is that people can become *ethically disabled* in a way that deprives them of the competencies required to appreciate the force of a broad range of moral reasons, where their own free, formative actions play a substantial role in producing the disability.<sup>7</sup>

However, once we emphasize that Vargas's challenge depends on Jeff's having an *impairment-level* moral deficiency, a further concern arises. This is because it's natural to understand Vargas's challenge in the following way. In preserving KC by denying Jeff's blameworthiness, we would be "writing off" the moral responsibility of *everyone* who, like Jeff, acts badly due to a deep-seated lack of moral sensitivity

<sup>7</sup> See Jacobs (2001) for a suggestive defense of this possibility. Jacobs contends (p. 3) that someone can be morally responsible for such ethical disabilities and their untoward consequences even if he or she did not aim to become that way. Although Jacobs does not explicitly argue that such agents aren't *directly* morally responsible for such consequences, his view that someone could become ethically disabled with regard to certain kinds of moral reasons, if well founded, could be used to substantiate Vargas's claim that Jeff doesn't meet the control conditions for moral responsibility when he fires the workers.

that stems from his or her free, formative actions. And, the thought continues, such a massive “write-off” would leave a troubling hole in our ordinary responsibility attributions (see Vargas p. 284). But, the further concern is, things look different when we stress that moral responsibility for only a subset of these actions—ones that exhibit impairment-level deficiencies—is at issue. If such cases turn out to be sufficiently rare, denying moral responsibility in them might seem a small price to pay to preserve KC.<sup>8</sup>

Here is how I think Vargas should respond to this further concern. Assuming that such cases exist, the cost of denying moral responsibility in them should be measured not only by their frequency, but also—and more importantly—by the strength of our reasons for judging these agents to be morally responsible for their impaired actions. In particular, if the morally impaired conduct strikes us as a natural outgrowth of the agents’ free, formative behavior—so that these agents could have been expected to sense that they were neglecting some aspect of their personal and moral development, even if they couldn’t have anticipated the specific shortcomings that would or might result—we may see strong, principled reasons not to let them off the hook. As I hope to show, these reasons can be further strengthened by establishing relevant similarities between these agents’ formative histories and those of otherwise similar agents whose shortcomings don’t amount to impairments.<sup>9</sup> If tracing theorists lack satisfactory grounds for denying the blameworthiness of the impaired agents, a lack that is revealed by analogy with other agents, cases like Jeff the Jerk may present a real problem for them, even if they are relatively rare.<sup>10</sup>

Summing up, I have suggested that Vargas should appeal to the notion of an ethical disability in developing Jeff the Jerk, and that he should not tie the significance of such cases too closely to their frequency. Another feature of Vargas’s challenge bears mention as well: we must suppose that the fifteen-year-old Jeff is morally developed *enough* to bear at least some significant degree of moral responsibility for his conduct, so that there is no problem on *this* score about tracing his blameworthiness back to this conduct. While this too might be challenged, I shall grant it in what follows. Let us now return to Vargas’s challenge.

<sup>8</sup> Thanks are owed to two anonymous referees for prompting me to address this concern, and for helping me to see it more clearly.

<sup>9</sup> I present my variation on Vargas’s case in Sect. 4, arguing in Sect. 6 that the judgment of derivative blameworthiness can be bolstered by a comparison with the formative histories of some unimpaired agents.

<sup>10</sup> An ancillary point is that while cases of ethical disability present an interesting and auspicious starting point for a challenge to KC, tracing theorists should not suppose that they will be the only problem cases. If such cases succeed, they may provide a template for other cases that challenge KC. In the interests of space, I won’t try to develop a second example here. However, it seems to me that many of the same points that support the judgment of derivative responsibility in ethical disability cases could also support such a judgment of agents who, though not ethically *disabled*, suffer from intermittent bouts of genuinely compulsive behavior, where the kind of behavior at issue didn’t “come from nowhere,” yet wasn’t specifically foreseeable as a risk of the habits that led to it. If our judgment that such agents are derivatively blameworthy for their compulsive behavior can be bolstered by analogy with agents whose similar habits have led solely to non-compulsive manifestations of such behavior, it seems to me, this should give Fischer and Tognazzini further pause about summarily dismissing derivative blameworthiness in the relevant cases by appeal to KC.

According to Vargas, the younger Jeff doesn't satisfy KC with respect to the later outcome, for

...at no point during the process of becoming a jerk, and certainly at no point before he undertook the process of becoming jerklike, 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. (p. 276)

Assuming that Jeff is blameworthy for how he fires the employees, and that he doesn't meet the control condition when he does so, his blameworthiness must derive from his actions at an earlier time. Yet Vargas rightly denies that the younger Jeff should see that his plan carries the risk of firing people in a despicable fashion in middle age. Thus, if we accept that Jeff is derivatively blameworthy for how he fires the employees, it seems that we must reject KC, and, with it, the constraints on derivative responsibility it imposes. Alternatively, if we insist on KC, Jeff and agents with similar stories won't qualify as morally blameworthy—not even derivatively—for their poor treatment of others, which hardly seems right. If Vargas's argument succeeds, then, we must choose between our commitment to tracing principles and some of our well-founded beliefs about who is blameworthy. I turn now to Fischer and Tognazzini's response to this challenge.

### 3 Fischer and Tognazzini's Response

As noted, Fischer and Tognazzini grant for argument's sake that Jeff is derivatively morally responsible for his later conduct. Even so, they observe, Vargas's cases pose a problem for KC only if all of these conditions are met:

- (i) The agent is morally responsible for the 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 was no suitable prior time at which the relevant outcome was reasonably foreseeable to the agent (Fischer and Tognazzini 2011, p. 208).

And they believe that there is room to resist (iii). The important question, in their view, is how we should describe the relevant outcome of Jeff's earlier actions. They canvass three possibilities:

- (Outcome 1) Jeff fires *those* employees who work for *that* company on *that* precise day in *that* precise manner
- (Outcome 2) 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
- (Outcome 3) Jeff treats some people poorly at some point in the future as a result of his jerky character. (p. 214)

Clearly, Outcomes 1 and 2 are not reasonably foreseeable. To defuse the challenge to KC, then, Fischer and Tognazzini must maintain both that Outcome 3 is reasonably foreseeable, and that nothing more in the way of foreseeability is needed. I shall leave the second claim to one side. As for the claim that Outcome 3 is reasonably foreseeable, this seems plausible enough: Jeff can presumably be expected to understand, if only in a general way, how his “self-improvement” regimen will likely affect his treatment of others. If he ends up acting in ways that typify the personal qualities he seeks to cultivate, this should come as no surprise.<sup>11</sup>

Now Fischer and Tognazzini are not fully explicit about what they take Jeff to be derivatively responsible for. To engage Vargas’s challenge, it seems to me, they must be prepared to hold that Jeff is blameworthy, albeit derivatively, for the fact that he treats *these particular individuals* as he does, and not merely for the fact that he treats someone or other poorly as a result of his jerky character. (Analogously, we would want to say that Brett, an intoxicated motorist, is derivatively blameworthy for his having *given Carmen whiplash*, and not just for the fact that some alcohol-related automotive mishap or other occurred, even though the consequence of his unwise decision to drink wasn’t foreseeable under the description “giving Carmen whiplash by rear-ending her vehicle due to driving while intoxicated.”<sup>12</sup>) Now since the outcome for which Fischer and Tognazzini take Jeff to be derivatively blameworthy is presumably the *same* outcome that was allegedly reasonably foreseeable, how are we to understand their contention that it’s the foreseeability of the more general Outcome 3 that allows Jeff to be derivatively blameworthy? Again, if what Jeff is blameworthy for is that he fires these particular workers in the jerky way he does, how does the foreseeability of the more general state of affairs allow him to be derivatively blameworthy for this specific outcome?

Perhaps the most natural thing for Fischer and Tognazzini to say is that ‘Outcome 3 was reasonably foreseeable’ is short for ‘It was foreseeable that Jeff does something or other with the relevant characteristics’ (i.e., the characteristics expressed in the Outcome 3-description). Insofar as this state of affairs was foreseeable, Fischer and Tognazzini can say, Jeff is derivatively blameworthy for the more specific state of affairs—that he fires these individuals in this way as a result of his jerky character—in virtue of which the more general state of affairs obtains. In short, then, it’s foreseeable *that* Jeff does *something* that satisfies this description, and what he is derivatively blameworthy for is *that* he does the particular thing that satisfies it. (Analogously, given Brett’s history of impaired decisions to drive, it’s foreseeable, given his failure to make suitable arrangements for getting home, that he will have an alcohol-related automotive mishap as a result of starting to drink now. And what Brett is derivatively blameworthy for is that the particular, whiplash-inducing mishap results from this decision.) In any case, this is

<sup>11</sup> It’s worth noting that Fischer and Tognazzini don’t claim that Jeff’s coming to possess *impairment-level* jerkiness (or insensitivity) is foreseeable, or that his blameworthiness depends on the foreseeability of *this* state of affairs.

<sup>12</sup> For a contrasting view, see Zimmerman (1986), p. 211. Zimmerman takes a fine-grained approach to the individuation of outcomes, on which causing Brittany to suffer whiplash and causing someone to suffer whiplash would count as distinct occurrences. By contrast, I see these as two descriptions of the same particular occurrence.



how I shall understand the statements ‘Jeff was derivatively blameworthy for Outcome 3’ and ‘Outcome 3 was reasonably foreseeable’, respectively.

Assuming that Outcome 3 is both reasonably foreseeable and specific enough to satisfy KC, Fischer and Tognazzini need only identify a particular time to which Jeff’s blameworthiness for that outcome can be traced back. And, they believe, such a time is readily discernable in Vargas’s description. They write:

We suggest that all tracing requires in this case is that Jeff could have reasonably foreseen Outcome 3 *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. (Ibid., emphasis added)

It is indeed a stated feature of the case that Jeff’s becoming jerk-like proceeds from a *decision* to become this way, a decision that issues from Jeff’s deliberations. This decision marks a clear turning point in Jeff’s biography; it’s a single, dateable event that embodies his intention to become a certain way, an intention that proves to be effective due to his ongoing efforts to fulfill it. Because Jeff acts with and ultimately achieves his aim, his blameworthiness for Outcome 3 is plausibly traced back to this decision.

As Fischer and Tognazzini go on to argue, it won’t help to suppose that Jeff doesn’t intend to become jerk-like *under that description*. Their reason comes to this. Even if Jeff thinks of the intended traits as “cool” instead of “jerky,” the characteristic manifestations of these traits include treating others poorly, something he can be expected to recognize (p. 215; cf. Vargas 2005, p. 277 n. 12).

I agree that it doesn’t matter whether Jeff intends to become jerk-like under that description, and that Jeff’s blameworthiness for Outcome 3 at least plausibly traces back to his earlier decision. Yet Fischer and Tognazzini’s response to Jeff the Jerk takes us only so far in addressing the larger issue about tracing principles. This is because we can imagine similar cases that exert the same pressure to trace the agent’s moral responsibility back to an earlier time, but where there is no decision or other action(s) with such clear implications for the agent’s moral development. In some of these cases, it will be hard to maintain that there is a suitable prior time, whether an isolated instant or a clearly demarcated span, when the agent can be expected to understand how his action(s) and/or omission(s) then will shape his future trajectory, even in the very general way that Jeff can be said to understand this. If this is right, Fischer and Tognazzini have lit upon an incidental feature of Jeff the Jerk.<sup>13</sup>

<sup>13</sup> Fischer and Tognazzini note that McKenna presents a case that is similar to Vargas’s, in that the agent “*consciously chooses* to cultivate his aggression and thick skin in order to survive in the locker room during his junior varsity football days...with little reason to expect that [this] will be the source of his coolness and tragic distance from his own children” (McKenna 2008, p. 33, emphasis added; cited in Fischer and Tognazzini 2011, p. 209 n. 3.) While Fischer and Tognazzini don’t say how they would address this case, it seems likely that they would see the agent’s conscious choice as the source of his later responsibility (as they do with Jeff the Jerk), still assuming that the agent isn’t directly morally responsible for the later outcome(s) in question. However, we could present a similar case without a pivotal choice, where the agent’s later paternal failings are consequences—but arguably not readily foreseeable consequences—of the earlier pattern of aggressive behavior.

## 4 Renewing the Challenge

To renew the challenge to tracing principles, it will be helpful to consider a case in which the future outcome neither hinges on the agent's decision to become a certain way nor is an obvious consequence of his formative behavior. I submit that Greg the Greedy fits the bill:

Greg has just been tasked with compiling and submitting a report to his company's board, a report that slates positions in the company's 1,200-strong workforce for elimination. He has been chosen for this task because his superiors believe that he will recommend deep cuts, which will free up revenue for salary increases higher up. In return for his cooperation, Greg expects a substantial year-end bonus. His alternatives are clear. If he bucks his superiors' expectations and documents how the existing workforce is already stretched thin, few layoffs will ensue (and he can forget about the bonus—not to mention a promotion anytime soon!). Or he can opt for a compromise, recommending just enough layoffs to appease his superiors. Being greedy, however, Greg surpasses his superiors' expectations, assiduously searching out every last cut in the off chance of netting a slightly larger bonus.

Were Greg not so greedy, he wouldn't have done so much extra work for such a modest expected gain. Yet his greed is not ethically disabling, and if it were the only noteworthy factor in his behavior, he would be directly morally blameworthy for submitting the report he does. However, it is not the only noteworthy factor: besides being greedy, Greg has a severe, impairment-level inability to identify with the needs and interests of those he regards as his social inferiors. Without this inability, he would have had qualms about the impact of his report on so many people's lives and reined in his greed, settling on the compromise plan. As it is, though he realizes that his decision will affect many people, this realization simply fails to register as a relevant consideration.

The pressure to treat Greg's conduct as a case of derivative blameworthiness stems not from his greed, then, but from his impairment-level insensitivity to the interests of those who will be affected by his conduct. As with Jeff, we need to know how Greg came to be this way. Here is a synopsis.

Placed in a new school in seventh grade, Greg coped with his many adolescent insecurities by instinctively emulating the mannerisms of his socially dominant classmates, who, as it happens, acted like little jerks! Greg did this without ever *deciding* to do so. Instead, receiving positive reinforcement as he progressed, he gradually and unreflectively adjusted his demeanor to theirs. Naturally, he understood that other students didn't appreciate his disdainful glances, his derisive snorts, or his snide remarks at their expense; yet such behavior passed for normal in his junior high school, and no single incident was ever serious enough to prompt an intervention from school officials. Greg's obnoxious personal style saw him safely through the end of high school, at which point, being more self-assured, he came to rely on it less. Upon finishing college and starting his first real job, Greg was immediately

impressed by the polished and professional manner, genial yet aloof, of his superiors; and he set out—consciously this time—to emulate them. To all appearances, Greg was a new man. Yet unknown to him, something of the old Greg remained. With his instinctive division of the world into big shots and nobodies firmly entrenched, Greg had real trouble registering the latter’s interests as reasons for him to do anything. And there was also this: grasping that his superiors kept score in pecuniary terms, and giddy with his newfound earning potential, Greg became preoccupied with personal enrichment, something that was hard to imagine given his lackadaisical outlook on his post-college finances a short while earlier.

By hypothesis, Greg exhibits an impairment-level insensitivity to the interests of some others when he submits the report, and so we have as much reason to deny that he meets the control condition when he submits the report as we do with Jeff when Jeff fires the employees. We also have good reason to view Greg as blameworthy; his conduct, like Jeff’s, is reprehensible, evincing a moral defect that grew out of his free, formative actions and habits. Thus, if Jeff is indirectly blameworthy for how he fires the workers, it seems that Greg is indirectly blameworthy for submitting the report he does. (I defend this claim in Sect. 6.)

Yet there are two important differences between these cases. First, in contrast with Jeff’s earlier stage, there is no decision or series of actions that the young Greg performs with the aim of becoming a certain way. Second, the relevant way that Greg becomes—the way that’s reflected in his later, reprehensible conduct in submitting the report—isn’t such an obvious continuation of his formative history, even in hindsight. I shall develop these points in turn. If I’m right, both will make it harder to say that Greg’s later conduct satisfies KC, even if Jeff’s does.

## 5 Derivative Responsibility Without Foreseeability

To begin, I take the following to be a reasonable analog of Outcome 3 in Jeff the Jerk:

- (Outcome 4) Greg imposes needless hardship on others at some point in the future, in part because of his indifference to the needs and interests of perceived social inferiors.<sup>14</sup>

Call Greg’s tendency to be unmoved by the needs and interests of perceived social inferiors ‘*T*’. To make the discussion more concrete, suppose that of the many job cuts recommended in Greg’s report, there were six in particular at which even his superiors initially balked. In saying, with regard to these six, that Greg is

<sup>14</sup> Here again I find Jacobs’ discussion of ethical disability suggestive (see note 7). Jacobs writes: “Some agents develop characters that disable them for sound ethical comprehension and action. The agent whose vices are especially entrenched, to the extent that she cannot appreciate ethical considerations in a correct manner, is *ethically disabled*” (Jacobs, op. cit., p. 34, emphasis original). Moreover, Jacobs writes, some agents turn out this way “as a result of habituation and their own voluntary conduct” (p. 37). As will become clear, I’m sympathetic to this general idea; however, Jacobs says little about how he understands the ability to appreciate ethical considerations.

derivatively blameworthy for Outcome 4, I mean that he is derivatively blameworthy for having caused this needless hardship to *these individuals* by slating their otherwise secure jobs for elimination, due at least in part to *T*. It's at least partly his fault, morally speaking, that these six lose their jobs, much as it's Jeff's fault, morally speaking, that the workers are fired rudely (or Brett's fault that Carmen has whiplash).

If this is indeed an apt description of what Greg is blameworthy for, a defender of KC must hold that his conduct was foreseeable under this description. As with Jeff the Jerk, what matters is whether it's foreseeable that Greg does something or other that satisfies this general description, not whether he can be expected to foresee the various and sundry details. My contention is that this outcome, for which Greg is derivatively blameworthy, need not have been reasonably foreseeable, even in these fairly general terms.

To support the unforeseeability claim, it's enough to show that there is no particular time at which the risk of this outcome should have been salient to Greg. Unlike Jeff, Greg didn't decide to become (or even to *try* to become) a certain way; he did not set out to become obnoxious or even someone *we* would regard as obnoxious. *A fortiori* he didn't choose to become this way for reasons he had. Rather, his initial change in behavior upon switching schools was gradual, unreflective, and (so to speak) tropistic. It's true that he eventually became exceedingly obnoxious and was untroubled by how this trait affected others. And like his friends, he eventually outgrew this trait. Unlike them, however, he developed along the way a deep-seated, selective disregard for others' interests, a moral defect of which he remained unaware. (Part of what masked this disregard, we may suppose, is that in the warm glow of post-graduation nostalgia, Greg saw *all* his former high-school classmates as peers, each with his or her place in a fundamentally benign pecking order.) In contrast to Jeff the Jerk, then, there is no obvious time in Greg's history to which his blameworthiness for Outcome 4 can be traced back.

The point is not that Outcome 4 wasn't reasonably foreseeable simply because Greg, unlike Jeff, never intended to become, or entertained reasons for becoming, the way that ultimately led to this outcome. The point is, rather, that these features of Jeff the Jerk make it easy to believe that the relevant risk should have been salient to Jeff when he decided to become jerk-like; and these features are absent in Greg the Greedy. We thus have room to ask whether Outcome 4 was reasonably foreseeable. Was there some particular point(s) or stage(s) at which the risk of Outcome 4 should have been salient to him? If so, where on the timeline does this point(s) or stage(s) fall? Natural candidates include (t1) when Greg first began to emulate the expressions and mannerisms of some of his classmates; (t2) when he first behaved obnoxiously toward an "uncool" classmate; (t3) when he first shrugged off the realization that the recipients of such treatment didn't appreciate it; (t4) when a clear pattern of obnoxious behavior had emerged; (t5) when he first began to understand that, in becoming one of the cool kids, he was taking his place in a pecking order in which smug attitudes and snide remarks were allotted to those lower down; (t6) when, early in high school, the relevant aspect of his moral development began (as seen by an omniscient observer) to lag behind that of his

friends; (t7) when, later in high school, the dispositions that would culminate in *T* were first present (though not evident as such); (t8) when, following his first semester of college, he began to see how his smug and obnoxious demeanor could seem immature and to curb its manifestations, even as he continued to ignore the ethical reasons for doing so; (t9) when, later in college, the relevant dispositions had become settled to the point that *T* could be attributed to Greg without qualification<sup>15</sup>; (t10) when, after college, he decided to retire his old, obnoxious ways for good, while still ignoring their ethical significance; (t11) when the lack of moral awareness embodied in *T* had become ethically disabling; and (t12) when he decided to emulate the polished manner of his superiors at work, while also adopting their measures of personal worth and achievement.

Might there be other candidates besides t1–t12? Perhaps. But it isn't clear that there must be; and if there aren't, Fischer and Tognazzini must select from these twelve. While t1–t12 include significant turning points and choices about how to become, there seems to be no single episode such that Greg could have been expected to recognize that he was risking significant hardship or deep-seated indifference to others (let alone the former as a result of the latter) by his behavior then. At t3, for example, Greg understood that the recipients of his snide remarks didn't appreciate them, as they sometimes made clear with the predictable epithets. But in such exchanges, which fit the competitive social dynamic of his school, Greg detected a grudging respect, and this contributed to his sense that he was doing well. At t5, his place in a pecking order came into view, a place that was maintained by his smug and obnoxious (though not deeply malicious) demeanor. Had he possessed greater insight and sophistication, he might have suspected that many kids go through such a phase. Even then, however, he wouldn't have surmised that this phase might seriously impede his moral development. Indeed, we can readily imagine that nothing about Greg's demeanor suggested future indifference to the hardships endured in the murky world of adults, let alone to hardships meted out by him. At t8 and t10, when his smug and obnoxious demeanor came up for review, we can imagine that some of his longtime friends gently prodded him to rethink some of his attitudes in light of the reasons for not treating others unkindly, but that taking up those reasons would have required a more intensive look at his attitudes than he was inclined to take; and so, as usual, he quickly turned his attention to thoughts that bolstered his self-esteem.

When exactly, then, should the risk of growing inured to such hardships have been salient to the young Greg? Granted, at t4, when the pattern of obnoxious behavior had first emerged, Greg could have worked to curb it, instead of continuing to ignore the reasons for being more sensitive to the feelings of others. But his obnoxiousness isn't the real and lasting problem. Nor does it enable him to identify that problem as an associated risk. On the other hand, his persistently ignoring the reasons for considering others' feelings *does* contribute, albeit not so straightforwardly, to the problem. Perhaps Greg could reasonably have been expected to

<sup>15</sup> It should be noted that I'm adopting the familiar conception of personal traits as stable cognitive, affective, and behavioral dispositions that are formed over extended periods in part by agents' voluntary actions and the resulting habits. Like the writers I'm discussing—and against situationists—I'm supposing that there are such personal traits, including traits like *T* and Jeff's jerkiness.

recognize that ignoring these reasons might result in his turning out to be a hardnosed and insensitive individual, and (say) not to be a pleasant person to work for someday. But the moral defect he exhibits in submitting the report, as well as the seriousness of the consequences, go well beyond this, and it seems a stretch to say that he should have seen these risks in ignoring the relevant reasons. We might say that he was aware of the relevant social and mental habits—habits that in fact contributed importantly to O4—and that he *should* have been aware of the general direction in which they might have been (and indeed were) leading him. But as we follow him down the timeline, there seems to be no point at which the more specific risk of O4 should have been salient to him.

So far, I have argued that there need not be any particular time at which Greg should have foreseen Outcome 4. I now wish to consider a second, related contrast between Greg and Jeff: though Greg’s ethical disability results from his own free choices and freely acquired habits, this disability isn’t such an obvious continuation of his formative history. If prompted, for example, Greg’s classmates could have easily pictured him behaving obnoxiously and superciliously later in life. But no one would have pegged him as a risk for indifference to the crucial needs and interests of others. In hindsight, a perceptive observer—one who knew the details of the report Greg submitted, and who had known Greg during his formative years—might be able to connect the dots. Such an observer might recognize that what looked like typical adolescent cliquishness and smugness were in fact the incipient signs of a profound and selective moral insensitivity. And such an observer might suspect (correctly, we may suppose) that a timely and sustained intervention from Greg’s parents or teachers might well have set him on a better course. But we may plausibly suppose as well that some aspects of Greg’s development would remain opaque, even in light of the successfully connected dots, leaving our observer to speculate about which features were pivotal in explaining why Greg’s moral development faltered as it did. By contrast, Jeff’s “arc” is straightforward: his jerkiness toward the fired employees is an obvious continuation of his initial transition toward jerk-like behavior, quite apart from the fact that he consciously chose this path. I submit that Greg’s more oblique path further underscores the implausibility of supposing that there was some time when its possible culmination in *T* should have occurred to him. This is especially so given that Greg’s greed, which is a pivotal factor in his causing the hardship, isn’t foreshadowed by his impairment-producing behavior.

If Greg is morally blameworthy for Outcome 4 despite its unforeseeability, we have a counterexample to KC. To block this counterexample without insisting that Outcome 4 was reasonably foreseeable, Fischer and Tognazzini must deny that Greg is even derivatively blameworthy for this outcome. I shall now argue that this won’t be easy.

## 6 Responsibility Denied?

To show that Greg is derivatively blameworthy, we need good reason to believe that blameworthiness for his earlier, free actions “travels” down the timeline to Outcome 4—or, as we might say, that his history is “transmission-apt”—despite the

fact that Outcome 4 isn't a reasonably foreseeable upshot of that history. We need to be persuaded that the fate of the six workers is his fault, morally speaking, much as the rude treatment of the employees is Jeff's fault, and Carmen's whiplash is Brett's fault. Importantly, this need not involve denying that there is *some* connection between foreseeability and derivative responsibility for one's future traits and their consequences, but only that KC correctly captures this connection. As will be seen, we can still allow that Greg's having a near-term and partial understanding of the risks of his actions and habits as he gradually shapes his character is part of how blameworthiness is transmitted to his later, impaired actions.

In outline, here is my proposal for how blameworthiness can outstrip foreseeability:

- A. Someone can be blameworthy for a moral deficiency by virtue of persistently ignoring reasons to remedy the emerging deficiency, and for thereby desensitizing herself to further indications that a "course correction" is needed, even if she has limited insight into the nature and possible extent of the deficiency when she ignores those reasons.
- B. If continuing to ignore those reasons causes the deficiency to get worse (more entrenched and more serious), the agent can be blameworthy for the "deeper" deficiency by virtue of persistently ignoring the reasons and not taking up the subsequent warning signs, even if there is no point at which the nature and extent of the deeper deficiency are reasonably foreseeable.
- C. If the resulting deficiency ends up being ethically disabling, the agent can be morally blameworthy for its behavioral manifestations by virtue of being morally blameworthy for the deficiency, notwithstanding that there was no point at which she could have reasonably foreseen the kind of situation in which the deficiency would be manifested.

Applying this outline to Greed the Greedy, we can identify an earlier stage at which Greg should have recognized (and, albeit to a limited extent, did recognize) that he had reason to cultivate greater sensitivity to the feelings and interests of his peers in general, and that doing so would require changes in his attitudes. Because he ignored these reasons, the idea is, he is morally blameworthy for becoming, during an intermediate stage, more established in his lack of sensitivity to the feelings and interests of some of his peers. And he is morally blameworthy, as well, for the consequent difficulty he has, during this intermediate stage, in recognizing warning signs (embodied in personal interactions and his responses to them) of his deepening indifference to the impact of his behavior on those who fall outside his ambit of concern. Insofar as he does recognize these signs, he is disinclined to give them much thought. But even had he worked harder to interpret them during this intermediate stage—and, indeed, even if the signs had more salient, owing to his having cultivated greater sensitivity during the earlier stage—he would have understood only in general terms that something was lacking in his regard for others. This understanding would have enabled him to address his shortcomings, thereby checking his slide toward *T*. But in no case would he have recognized that *T* in particular lay just over the horizon, for he wouldn't have grasped the role of

social status in demarcating the ambit of his concern, or how seriously short he might fall in his concern for those outside that ambit.

Just as Greg is morally blameworthy for how he becomes in the intermediate stage by virtue of his not heeding the relevant reasons during the earlier stage, so he is morally blameworthy for how he becomes in the later stage by virtue of not being more receptive to the warning signs during the intermediate stage. And just as he is blameworthy for how he becomes in the later stage by virtue of his not being more receptive to the warning signs during the intermediate stage, so he is blameworthy for costing the six workers their jobs by virtue of not being more receptive to the warning signs in the intermediate stage, and thereby allowing himself to become desensitized to the countervailing reasons. It is Greg's fault, morally speaking, that he imposes this hardship on these individuals, even though he is not directly blameworthy for this outcome. At the same time, while his blameworthiness for this outcome is derivative, it doesn't stem solely from his conduct during the intermediate stage *or* from his conduct during the earlier stage. Instead, he is derivatively blameworthy for it because the disposition it manifests is the *cumulative effect* of his conduct during both of those stages. This is so even though there is no time at which this cumulative effect or Outcome 4 was reasonably foreseeable.

I submit, then, that the above provides us with a way to understand how Greg's blameworthiness can "travel" down the timeline so that he is derivatively blameworthy for Outcome 4, even though there is no earlier time to which his blameworthiness can be traced back. If this is right, KC is false.

But is it really Greg's fault, morally speaking, that the six workers suffer the needless hardship they do? Or can Fischer and Tognazzini plausibly reject this claim on the grounds that Outcome 4 isn't foreseeable? I believe that they cannot. To see why, it will help to consider Greg\*, whose history is very similar to Greg's, but who doesn't end up ethically disabled (so that the question of derivative responsibility doesn't arise). If Greg\*'s formative history is "transmission-apt"—if it has the right features to transmit blameworthiness from earlier to later stages, setting aside the fact that he is directly blameworthy—it will be hard to deny that Greg's history is transmission-apt as well. To resist this conclusion, Fischer and Tognazzini will need to identify precisely where they believe the transmission of responsibility breaks down.

To begin, then, suppose that Greg\* acquires *T* via a formative history strikingly similar to Greg's, but that, unlike Greg, he meets several people along the way who *look down* on his smugness (as opposed to merely disliking it), prompting him to review how he comes across in this regard. And suppose that the result of this self-review is that the *T*-contributing factors are somewhat checked, so that *T* never becomes ethically disabling. Like Greg, the younger Greg\* couldn't have foreseen that his behavior would lead him to acquire *T*. Unlike Greg, however, the older Greg\* can appreciate that the interests of his perceived inferiors do give him practical reasons, even if these reasons are less salient than we would hope. Now, because Greg\* isn't completely insensitive to the interests of those affected, he is not entirely without qualms about slating the six extra jobs for elimination. Let us



suppose, then, that unlike Greg, Greg\* is offered inducements to “go the extra mile” in recommending job cuts, which helps him to overcome these qualms.

To repeat, Greg\* is directly blameworthy, yet we can still ask whether his formative history is “transmission-apt.” I submit that what Greg\* does understand during his formative stages supports an affirmative answer. During these stages, he understands that there are reasons not to subject his “uncool” classmates to his jibes, which they find unpleasant. And like Greg, he freely ignores these reasons and acts in ways that can be expected to reinforce his smug and obnoxious demeanor, a demeanor that will likely persist unless he works to change course. At each of these stages, Greg\*’s moral responsibility for what he is like then plausibly derives from his conduct at previous stages, conduct that extends and further establishes his developing traits. He is directly blameworthy for this free conduct, and derivatively blameworthy for the existence of the deficiencies it expresses.

Suppose that Greg\* proceeds during a subsequent stage to compound his deficiencies—say, by adding a “layer” of smugness to his existing insolence and cliquishness upon ascending to the dominant social group—so that it becomes harder still for him to show age-appropriate sensitivity to the feelings of his “uncool” classmates, where this in turn hampers the development of his concern about their needs and interests. We can plausibly see his blameworthiness for this emerging deficiency as deriving from his actions and habits at several previous stages, even though we cannot trace it back to any one in particular.

We might put the point this way. At any particular point on the timeline, provided that Greg\* can then be expected to understand in general terms how he is comporting himself toward others, he can transmit blameworthiness for the shortcomings he is thereby fostering to subsequent stages. Moreover, he can be derivatively blameworthy for these shortcomings even if the precise cumulative or compound effect of his actions and habits across successive stages isn’t foreseeable at any time. Finally, insofar as his eventually acquiring *T* is the cumulative effect of his habits of mind and action across these stages—stages whose near-term significance for his moral development he could have foreseen—there seems to be nothing in his formative history that would “break the transmission” of blameworthiness from these earlier stages to his acquisition of *T*.<sup>16</sup>

In saying that Greg\* is derivatively blameworthy for having acquired *T*, I mean that it’s his fault, morally speaking, that he possesses this deficiency, a deficiency that disposes him to callousness in his deliberations regarding perceived inferiors. Although we wouldn’t normally hold him to account for producing this outcome (except insofar it’s manifested in his behavior), it remains a bad outcome for which he can be blamed, much as we can fault Jeff for having become a jerk. (Analogously, in saying that remiss parents bear significant moral responsibility for a child’s objectionable traits, we mean that the untoward developmental outcome is

<sup>16</sup> In this connection, Bill Pollard’s discussion of moral responsibility for habitual actions is helpful. According to Pollard, we can be morally responsible for habitual actions because we have “intervention control” over whether we perform them—the ability to stop ourselves at will before performing or completing a habitual action, however natural its performance may be (See Pollard 2006, pp. 59–60). I shall suppose that such control allows us to see habitual action as voluntary in a way that supports attributions of derivative moral responsibility.

their fault, morally speaking, at least to some significant extent. This is some indication that we view personal dispositions as the kinds of outcome for which moral blame can be assigned.) Like Jeff's jerkiness, this state results from a pattern of voluntary behavior over an extended period, albeit without a key turning point or pivotal moment, much less one that Greg\* should have recognized as such. If called to account for acquiring *T*, Greg\* could accurately reply that he didn't realize—and could not have been expected to realize—that he was at risk for *T* in particular.

But would this be a valid excuse, a basis for denying that the outcome was his fault? A defender of KC may believe so. In response, however, it can be observed that Greg\* voluntarily continued on the course that led to *T*, despite indications of the general kind of moral deficiency—a lack of concern with the interests of some others, of which *T* is a more specific instance—that was emerging. If KC's defenders claim that Greg\* isn't blameworthy for *T*, but only (perhaps) for this more general deficiency, they should be able to tell us where the transmission of responsibility for the more specific outcome breaks down. For example, they might say that when Greg\* added the aforementioned layer of smugness, he was derivatively blameworthy for the combined effects of this smugness and his existing insolence and cliquishness, since he should have foreseen these combined effects. But, the thought would continue, there must be a "transmission breakdown" before *T* developed, since *ex hypothesi* Greg\* couldn't have foreseen how *T* might proceed from this combination plus his subsequent behavior. It won't be a *complete* breakdown, since he will still be derivatively blameworthy for the more general deficiency he possesses at the later time; but the voluntary behavior that led to *T* must ultimately fail to transmit blameworthiness *for* his becoming *T*.<sup>17</sup>

Where, then, might this breakdown occur? A natural place to look is the last stage at which Greg\*—albeit without recognizing that he was becoming *T*—could have avoided this outcome by heeding his reasons for greater self-reflection. This was the stage when Greg\* was about to make the transition from possessing *T*-type tendencies to genuinely embodying *T*, a transition he couldn't have recognized, except perhaps in superficial terms, such as "becoming more comfortable with who he is." Should we conclude from Greg\*'s inability to foresee this transition that his becoming truly *T* isn't his fault, morally speaking, so that the transmission of responsibility breaks down here? I believe not. This conclusion may seem plausible if we suppose that the "source" of the transmission is just Greg\*'s conduct during this penultimate stage. But this is not the case; on the contrary, what he is blameworthy for at this stage is not becoming *T tout court*, but only for the incremental step that makes the difference between his having *T*-type tendencies and his becoming *T* "full stop." His blameworthiness for becoming *T* derives, then, not solely from his conduct during this stage, but from the patterns of behavior over a much longer span by which Greg\* develops the various elements of *T*.

<sup>17</sup> The claim that we are morally responsible for personal traits, unforeseeable though they are, may be grist for the attributionist's mill (see note 4). The attributionist may say that the correct lesson from such cases is that neither voluntary control nor foreseeability is required for moral responsibility. Be this as it may, I will continue to suppose that moral responsibility for such traits depends on our having some degree of voluntary control over the actions and habits that produce them.

If Greg\* is blameworthy for bringing himself to the verge of *T*, owing to the free, formative actions that led him there, why believe that the transmission of responsibility breaks down there, so that he isn't blameworthy for becoming *T*? If the transmission of responsibility doesn't break down there, despite Greg\*'s inability to see *T* (let alone his becoming *T* "full stop") as the outcome, why think that it will break down somewhere else on the timeline—say, just before Greg clearly comes to possess *T*-tendencies, so that it isn't his fault that he comes to possess them? As long as he can be expected to have a suitable general awareness of the direction in which he is headed, why think it won't be his fault, morally speaking, that he embodies *T*-type tendencies? If these tendencies emerged gradually, we should conclude that he is morally responsible for them by virtue of his self-shaping actions and habits during an extended period, not that he is off the hook for them. Finally, if Greg\* is morally responsible for acquiring *T* by virtue of his being morally responsible for the formative history that shaped this tendency, it's hard to see why he couldn't be morally responsible in this way for expressions of *T*. That is, it's hard to see why he couldn't be derivatively blameworthy for the relevant conduct by virtue of being blameworthy for forming the disposition toward such conduct. Setting aside the fact that Greg\* is directly blameworthy for this conduct, his formative history seems to be "transmission-apt."

Similar considerations apply to Greg, who (unlike Greg\*) isn't directly blameworthy for having submitting the report. After all, Greg too can be expected to understand the near-term consequences of his conduct at the various stages of his development, he too acts in ways that reinforce his shortcomings and inhibit his moral growth, he too fails to notice warning signs because of his pattern of ignoring moral reasons, and he too acquires *T* as the cumulative effective of his behavior at these earlier stages. Why think, then, that his actions at the earlier stages wouldn't transmit blameworthiness to his later stages? Why should the cumulative effect's being unforeseeable block the stage-wise transmission of derivative responsibility in his case but not in Greg\*'s? If we accept that Greg\* *could* be derivatively blameworthy for Outcome 4 by virtue of his free formative activities (or that his formative history is "transmission-apt"), why should Greg get off the hook merely because his shortcoming ultimately reaches the level of a disability? That he is ethically disabled in this respect prompts the question of whether he is derivatively blameworthy for Outcome 4; it doesn't by itself support a negative answer to this question. Assuming that he, like Greg\*, is morally blameworthy for having acquired *T*, it's hard to see why that blameworthiness won't extend to his exhibiting *T* in submitting the report he does.

If this is right, denying Greg's blameworthiness for Outcome 4 won't be plain sailing. Just as Jeff is derivatively blameworthy for having become a jerk, so Greg\* and Greg are derivatively blameworthy for having acquired *T*. And just as Jeff is derivatively blameworthy for Outcome 3 (and as Greg\*'s history is suitable to transmit blameworthiness to Outcome 4, though he is in fact directly blameworthy for this outcome), so Greg is derivatively blameworthy for Outcome 4. The limits of his understanding do not get him off the hook; this outcome is still his fault, morally speaking, even if, perhaps, he doesn't merit as *much* blame as he might have if the outcome had been foreseeable. It seems, then, that someone who denies that

Outcome 4 is Greg's fault on the basis of KC will have to tell us where the transmission of blameworthiness breaks down. As I have suggested, this won't be easy given the relationship between *T* and what Greg does understand as he moves down the timeline. Indeed, instead of denying that Greg is blameworthy for Outcome 4 because KC isn't met, it seems more plausible to deny that Jeff's blameworthiness for Outcome 3 really *depends* on KC's being met; perhaps it's sufficient that Jeff ignores indications of a moral deficiency he is reinforcing, so that his being able to see the exact nature of the deficiency isn't truly essential.

In any case, if our verdict on Outcome 4 cannot easily be separated from our verdicts on the other outcomes, it won't be easy to deny Greg's blameworthiness by appealing to KC. While cases of ethical disability may be rare, then, they will have implications for other cases, and theorists who aren't firmly wedded to KC may wish to entertain alternatives to it.

## 7 Toward a Compromise Solution

I have argued that there is no suitable earlier time at which Outcome 4 is reasonably foreseeable, and that defenders of KC cannot easily deny that Greg is derivatively blameworthy for this outcome. Does accepting both of these claims force us to reject tracing principles? Not necessarily, for there is room to explore a compromise position. I shall suggest only the bare outlines of such a position here, but I hope this will suffice to show that further exploration is warranted.

In the classic drunk-driving case, it seems clear that the agent's blameworthiness for the impaired decision to drive (and for any untoward consequences thereof) depends on the impairment's being reasonably foreseeable. If, for example, another agent had surreptitiously spiked the driver's iced tea with an intoxicating substance, we wouldn't see the driver as blameworthy for his impaired decision to drive *because* this wasn't a foreseeable outcome of drinking the tea. By rejecting tracing principles altogether, we would be abandoning this very plausible explanation of the difference between these cases.

The first part of the suggested compromise is to say that the foreseeability requirement applies unproblematically in cases like these, where the relevant outcome hinges on the agent's conduct at some particular moment or during some isolated episode. If what the agent does (or doesn't do) at that time is pivotal to the occurrence of the bad outcome, it seems unfair to blame the agent unless the risk involved in so acting should have been salient to her then. The agent who knowingly drinks the alcohol plausibly meets this condition, while her counterpart who drinks the spiked tea does not. Similarly, given that Jeff's decision to become jerk-like is pivotal to Outcome 3, it would be unreasonable to trace his blameworthiness for this outcome back to this decision if the outcome were unforeseeable. According to the compromise position, then, the foreseeability requirement reflects the unfairness of blaming someone for the consequences of her actions at a pivotal time, when she couldn't have been expected to recognize that those actions might prove to be pivotal in this way.

The second part of the compromise is to deny that scenarios in which someone has gradually shaped her character over an extended period feature such pivotal episodes, and thus that they are subject to this requirement. Since Greg's submitting the report he does won't be subject to the restricted requirement, we won't have to make the hard choice between tracing principles *tout court* and our well-founded belief that Greg is blameworthy for Outcome 4.

Of course, this compromise position will be appealing only in conjunction with both a more general account of derivative moral responsibility for the consequences of our character-forming behavior, and a plausible specification of which times count as pivotal in the relevant way. While both tasks are beyond the scope of this article, I wish to make a start on the second. A time will be *pivotal* in the relevant way, I suggest, if there's a type of action such that performing (or failing to perform, as the case may be) an action of that type at that time involves a recognizable risk that so acting will preclude the agent's having a reasonable opportunity to avoid the bad outcome in question. The story of our intoxicated motorist will feature such a pivotal time, since, when this motorist commences (or decides to continue) drinking without a plan to get home, he then performs an action of a kind that involves a recognizable risk that his driving-related judgment will be impaired in a way that precludes his having a reasonable opportunity to avoid an automotive mishap. The story in which the agent's iced tea is spiked without his knowledge will also feature such a pivotal time: he too performs a kind of action—ingesting an intoxicating substance—such that his performing an action of that kind then involves the same recognizable risk. Thus the restricted foreseeability requirement will apply to both motorists. The difference, of course, is that the first motorist will meet the requirement (and so will presumably be derivatively blameworthy for his mishap) while the second won't, since he, being unaware that he was performing such a risky action, couldn't have been expected to foresee the mishap.

Relying on this notion of a pivotal time, then, I tentatively propose the following as an alternative to KC. According to the restricted foreseeability requirement,

(RFR)            If there is a pivotal time in the history of an untoward outcome, the agent won't be morally responsible for that outcome unless it's reasonably foreseeable at the pivotal time

Part of the notion of a pivotal time involves being precluded from having a *reasonable opportunity* to avoid the untoward outcome. A detailed examination of RFR would require a closer look at what such an opportunity consists in. For present purposes, however, a flatfooted construal of a reasonable opportunity will suffice. When someone's judgment is impaired by alcohol, she may lack a reasonable opportunity to avoid behaving badly. When someone has accidentally boarded the wrong bus, he may lack a reasonable opportunity to reach his destination on time. There are clearly times at which performing an action of one of these kinds involves a recognizable risk that doing so will preclude one's having a reasonable opportunity to avoid the bad outcome in question (behaving badly, arriving late). RFR implies, in effect, that one is blameworthy for the outcome of such a pivotal time only if one *should* have recognized that the time *was* pivotal to that outcome. In other words, if the causal history of a bad outcome features such a pivotal time,

one isn't blameworthy for that outcome unless it was foreseeable at the pivotal time. Because there is a pivotal time in the bus example, whether the agent is derivatively blameworthy for arriving late will depend on whether he could have reasonably been expected to realize that he might be boarding the wrong bus. There is also a pivotal time in the spiked tea example, since performing an action of the kind *ingesting an intoxicating substance* involves a clear risk in the agent's circumstances (a risk that would be clear to *him* if he knew more about his circumstances) of depriving him of a reasonable opportunity to avoid an automotive mishap. Hence, this case, too, is subject to the RFR's requirement for blameworthiness (though the requirement isn't met).

If RFR is right, it's unfair to blame the driver in the spiked tea example not simply because the resulting mishap was unforeseeable, but because the unforeseeable mishap was the result of his conduct *at a pivotal time*. If someone's conduct at a particular time both deprives her of a reasonable opportunity to avoid a bad outcome *and* grounds her blameworthiness for that outcome if it ensues, it must be reasonable to expect her to understand the significance of her conduct at that time. In particular, it must be reasonable to expect her to understand that she is engaging in conduct of a type that, in circumstances like hers, involves a significant risk of depriving her of a reasonable opportunity to avoid such an outcome.

Let us now return to Greg the Greedy and Jeff the Jerk. When Greg slides into ethical disability, there is no time and act-type such that Greg should have realized that performing (or failing to perform) such an action then would preclude his having a reasonable opportunity to avoid Outcome 4. Earlier in his history, he had numerous opportunities to change course and avoid this outcome, even though he had only a hazy sense of the reasons for changing course and the personal qualities that were taking root. At the last point on the timeline when he could have avoided becoming ethically disabled (and thus being precluded from having a reasonable opportunity to avoid Outcome 4), he still doesn't perform an action of the relevant type: nothing he does then is the *type* of thing that, by itself, embodies a recognizable risk in circumstances like his of depriving him of a reasonable opportunity to avoid Outcome 4. Granted, he may perform actions of the type *reinforcing bad habits* and *further neglecting his moral development*, but these aren't types that embody the relevant risks. In this respect, they are unlike *ingesting an intoxicant* or (say) *performing a pyrotechnical stunt*. Since there is no pivotal time in Greg the Greedy, RFR doesn't imply that he isn't derivatively blameworthy for Outcome 4.

Notably, the same turns out to be true of Jeff. Although Outcome 3 was foreseeable when Jeff decided to become jerk-like, this decision did *not* deprive Jeff of a reasonable opportunity to avoid this outcome; on the contrary, he undoubtedly had a great many later opportunities to change course. Moreover, it was only after a long pattern of reinforcing behavior that Jeff acquired an impairment-level lack of sensitivity, and there was presumably no point when he should have recognized the risk of becoming morally impaired (and thus potentially losing a reasonable opportunity to avoid Outcome 3). Since there is no pivotal time (in my sense) in Jeff the Jerk, then, RFR doesn't imply that Jeff's derivative blameworthiness for Outcome 3 hinges on its foreseeability. Consistent with RFR, then, we can say that

Jeff is blameworthy for this outcome not simply in virtue of his blameworthiness for making the decision to become jerk-like, but in virtue of that decision *together with* the many subsequent actions and habits by which he came to be morally impaired. Because he isn't blameworthy for Outcome 3 solely in virtue of that decision, a defender of RFR will say, the foreseeability of this outcome isn't essential to Jeff's blameworthiness for it after all.

More generally, with regard to choices to influence one's personal development, a defender of RFR should allow that such choices can have profound consequences for how one turns out and what one does as a result, while denying that such choices typically occur at pivotal times, since, like Jeff, agents will typically have opportunities to reflect on and modify the paths they have chosen.<sup>18</sup> Assuming that the outcomes for which we are derivatively morally responsible include character traits, most instances presumably won't involve pivotal episodes and/or decisions about how to become. Instead, they will involve what I earlier called *untraceable responsibility transmissions*: moral responsibility for outcomes by virtue of those outcomes' being the cumulative effects of morally significant patterns of voluntary action and the resulting habits, without there being any particular time to which moral responsibility can be traced back. Such a formative history can transmit blameworthiness for objectionable qualities, I have suggested, even if the agent's insight into the nature of these qualities is limited, and even if she doesn't intend to reinforce them. Such a history is "transmission-apt" provided that the agent can be expected to have some awareness of the emerging qualities, and to see that she therefore has reason to change course.

To illustrate, consider a significant three-year span in someone's adult life. If the result of the agent's behaving in an objectionable way during Year One is that her disposition to exhibit such behavior is more fully settled by Year Two, she could be morally responsible for embodying this disposition to the extent she does in Year Two because it's a consequence of her behavior during Year One. And if, during Year Three, this disposition is even more entrenched, but its manifestations have changed in response to new circumstances (e.g., she is now at a point where intimidation is more effective than back-stabbing for neutralizing rivals, though she has previously balked at direct confrontation), we can see her as being morally responsible, both for how settled this disposition has become, and for the behavior that reflects its depth and nature in her new circumstances, by virtue of her conduct during Years One and Two. Setting aside the extent to which she is directly blameworthy for her actions during Year Three, these actions are candidates for derivative blameworthiness insofar as the dispositions they manifest are the cumulative effects of her free conduct during the previous two years. Although some of these dispositions and their manifestations were not foreseeable during Years One and Two, her conduct during these stages nonetheless seems to be "transmission-apt," for she has some awareness of the qualities she has cultivated and can be expected to understand the reasons against cultivating such qualities.

---

<sup>18</sup> Thanks to an anonymous referee for prompting me to consider the implications of the revised requirement for character-shaping decisions, including ones whose effects are gradual.

If such a history is indeed “transmission-apt,” there is reason to believe that there could be untraceable responsibility transmissions. If so, it’s hard to see why supposing that the agent has become ethically disabled by Year Three should serve to block these transmissions. She will have made herself into the ethically disabled person she has become, and her being disabled in that way doesn’t seem to be a valid excuse for what she does as a result. Manifestations of her ethical disability will be her fault, morally speaking. To deny that this is her fault, it seems, we would have to deny that her history is transmission-apt after all, and that would require identifying where the transmission of responsibility breaks down. And I believe that this won’t be easy.

Summing up, the compromise I have proposed has two parts: a restriction of the foreseeability requirement to consequences whose risk should be salient in light of one’s actions at a pivotal time, and a suggestion as to how we might understand derivative moral responsibility for outcomes (such as Outcome 4) to which the restricted foreseeability requirement doesn’t apply. Both parts of the proposal require further elaboration and defense. My aim here has been to show that such a project is worth pursuing. If viable, such a compromise may yield a solution to Vargas’s problem. If, on the other hand, untraceable responsibility transmissions should be rejected, showing this will be a step toward vindicating Fischer and Tognazzini’s response to this problem; for then they can deny that Greg the Greedy is indeed a case of derivative blameworthiness without foreseeability (so that KC is safe). As I hope to have shown, it’s simply too early to rule that tracing principles are off the hook.<sup>19</sup>

## References

- Adams, R. (1985). Involuntary sins. *Philosophical Review*, 94, 3–31.
- Dennett, D. (1984). *Elbow room*. Cambridge, Mass.: MIT Press.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control*. New York: Oxford University Press.
- Fischer, J. M., & Tognazzini, N. (2009). The truth about tracing. *Noûs*, 43, 531–556.
- Fischer, J. M., & Tognazzini, N. (2011). The triumph of tracing. In J. M. Fischer (Ed.), *Deep control: Essays on free will and value*. New York: Oxford University Press.
- Ginet, C. (2000). The epistemic conditions for moral responsibility. *Philosophical Perspectives*, 14, 267–277.
- Jacobs, J. (2001). *Choosing character: Responsibility for virtue and vice*. Ithaca, N.Y.: Cornell University Press.
- Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.
- Khoury, A. (2012). Responsibility, tracing, and consequences. *Canadian Journal of Philosophy*, 42, 187–207.
- McKenna, M. (2008). Putting the lie on the control condition for moral responsibility. *Philosophical Studies*, 139, 29–37.
- Pollard, B. (2006). Explaining actions with habits. *American Philosophical Quarterly*, 43, 57–69.
- Sher, G. (2009). *Who knew? Responsibility without awareness*. New York: Oxford University Press.
- Smith, A. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, 115, 236–271.

<sup>19</sup> I am deeply indebted to two anonymous referees for *Erkenntnis* for exceptionally helpful and generous comments, which resulted in numerous improvements to the manuscript. I would also like to thank John Fischer and Kai Draper for valuable feedback on an earlier draft of the manuscript.



- Smith, A. (2008). Control, responsibility, and moral assessment. *Philosophical Studies*, 138, 372–392.
- Timpe, K. (2011). Tracing and the epistemic condition on responsibility. *Modern Schoolman*, 88, 5–28.
- Vargas, M. (2005). The trouble with tracing. *Midwest Studies in Philosophy*, 29, 269–291.
- Zimmerman, M. (1986). Negligence and moral responsibility. *Noûs*, 20, 199–218.