

## Culpable Ignorance

The Argument from Ignorance that I have presented in previous chapters culminates in the Origination Thesis, which holds that all blameworthiness or culpability is rooted in witting (or quasi-witting) wrongdoing, and thus that wrongdoing committed from ignorance of the fact that it is wrong can never be something for which one is *directly* to blame. This claim is of course compatible with holding that one can nonetheless on occasion be *indirectly* to blame for such wrongdoing. This possibility is in fact explicitly accommodated in the Argument's second premise, which in its most recent formulation (Draft 5) states:

- (2) if one committed some act or omission because one failed to believe that it was overall morally wrong, one is morally to blame for it, and thereby morally to blame for any of its consequences, only if one is morally to blame for one's failure to believe that it was overall morally wrong.

In brief: blameworthiness for *wrongdoing* committed from ignorance is to be traced to blameworthiness for the *ignorance* (the failure to believe) that gave rise to that wrongdoing. This raises the question that I will pursue in this chapter: What does being blameworthy for such ignorance entail?

Philosophers frequently talk in this context of “culpable ignorance.” Indeed, this has become standard, and I have accordingly employed the phrase in the title of this chapter. But in truth it is not well chosen, since it suggests that culpability attaches directly to a person's ignorance rather than to the person him- or herself *for* his or her ignorance. “Inexcusable ignorance” is more felicitous and thus the term that I will employ in this chapter, which proceeds as follows. In §5.1 I conduct an extended inquiry into whether and how one's being indirectly to blame for something is to be traced to one's being directly to blame for something else. According to the Argument from Ignorance, the distinction between direct and indirect blameworthiness tracks the distinction between direct and indirect control, and I inquire into what this claim means and why we should accept it. I then turn to the question of which consequences of behavior for which one is directly to blame should be said to be consequences for which one is indirectly to blame. It is very commonly held that, among the consequences of behavior for which one is directly to blame, only those that were foreseeable at the time at which one engaged in that behavior can be consequences for which one is indirectly to blame. I argue that this view needs qualification if it is to capture

what its proponents intend. I then address a recent argument for the claim that one is never indirectly to blame for anything. I reject this claim but endorse the related claim that being indirectly to blame for some consequence of behavior for which one is directly to blame cannot increase the degree to which one is already to blame in virtue of having engaged in that behavior. Finally, I focus on indirect blameworthiness for ignorant wrongdoing in particular and respond to a number of objections that have been raised against employing the strategy of tracing in this context.

According to the Argument from Ignorance, ignorance of wrongdoing is an excuse for the wrongdoing that results, unless one is to blame for one's ignorance. In §5.2 I raise the question whether ignorance of wrongdoing is an excuse for the wrongdoing that results even when one *is* to blame for one's ignorance. I distinguish between two interpretations of this question and argue that, although the answer is "No" on one of these interpretations, it is "Yes" on the other. I end by addressing the question of whether ignorance can ever be inculpatory rather than exculpatory. I give reasons for denying that it can be but point out that, even if it can be, this fact is perfectly consistent with the position I have taken on the conditions under which ignorance provides an excuse for wrongdoing.

## 5.1 Tracing blameworthiness

### 5.1.1 Tracing and tracking

As I have just noted, Premise 2 of the Argument from Ignorance implies that blameworthiness for ignorant wrongdoing is to be traced to blameworthiness for the ignorance that underlay that wrongdoing. (This way of putting the point is rough, since it ignores the possibility of quasi-witting wrongdoing, regarding which see §3.3. For simplicity of exposition, I will continue to disregard this possibility in what follows.) The Argument goes on to imply that any blameworthiness one may have for one's ignorance must also be merely indirect. The relevant premises are these:

- (6) one is never directly in control of whether one fails to believe something—that is, any control that one has over failing to believe something is always only indirect,
- (7) if one is morally to blame for something over which one had only indirect control, then one's blameworthiness for it is itself only indirect, and
- (8) one is indirectly morally to blame for something only if that thing was a consequence of something else for which one is directly morally to blame.

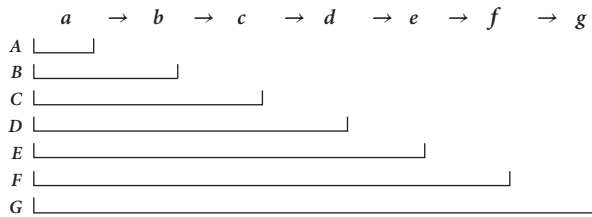


Figure 2.1 BILL AND JILL

I defended Premise 6 in §4.6. In this section I will attend to Premise 7. (Premise 8 is definitionally true.)

Premise 7 gives partial expression to the idea, one that goes back at least as far as Aristotle,<sup>1</sup> that the distinction between direct and indirect blameworthiness *tracks* the distinction between direct and indirect control; that is, if one is to blame for something, then one's blameworthiness for it is direct if and only if, and because, it was in one's direct control, and one's blameworthiness for it is indirect if and only if, and because, it was in one's indirect control. Consider once again the case of BILL AND JILL. In this case, Bill had direct control over *a* but merely indirect control over its consequences, *b–g*. Accordingly, on the present approach, if Bill is to blame for *a*, he is directly to blame for it, whereas, if he is to blame for *b–g*, he is merely indirectly to blame for these events. His control over *b–g* was rooted in his control over *a*, and hence his blameworthiness for the former is correspondingly to be traced to his blameworthiness for the latter

This picture of the relation between blameworthiness and control is certainly attractive, but it faces a number of challenges.

### 5.1.2 Does tracing require tracking?

One question that deserves attention is this. Even if we grant both the distinction between direct and indirect control (as surely we should) and the distinction between direct and indirect blameworthiness (a somewhat more controversial matter, to which I will shortly return), why should we also grant that the latter tracks the former? Why might it not happen on occasion that one is directly to blame for something that is only in one's indirect control, or that one is indirectly to blame for something that is only in one's direct control?<sup>2</sup> This is a good question, since it seems clear that in other moral, or morally relevant, contexts

<sup>1</sup> See Aristotle (1941), 1113b.

<sup>2</sup> Cf. Montminy (2016), pp. 68 f., where it is argued that one can be directly responsible for something that is only in one's indirect control. Montminy's argument founders on the failure to recognize the category of hybrid control, regarding which see §2.2.4. See also Zimmerman (2008), pp. 182 f.

the direct-indirect distinction does not track the distinction between direct and indirect control.

Consider, for example, the distinction between final and non-final value. Suppose, for the sake of illustration, that pain is finally bad, i.e., bad for its own sake.<sup>3</sup> Then, all else being equal, something that causes pain—a blow to the head, say—will be non-finally (more particularly, instrumentally) bad.<sup>4</sup> In such a case, the badness of the blow will be derivative from the badness of the pain and thus indirect, whereas the badness of the pain will be original or non-derivative and thus direct. And notice that control might not enter the picture at all. It could do so (the blow could be one that is delivered during a fight), but it needn't (the blow could be caused by a falling rock).

Or consider the distinction, seldom discussed, between fundamental and incidental moral obligation. Here, it is plausible to think, control must be implicated (as the common claim that “ought” implies “can” suggests). By way of illustration, suppose that Dot is a doctor who is obligated to give her patient, Pat, an injection in order to improve his health. The case may be represented as in Figure 5.1. In this figure, *L* (*l*) is Dot's decision to inject the medication, *M* is her moving her finger (*m* is the movement of her finger), *N* is her plunging the syringe (*n* is the syringe's depression), *O* is her injecting the medication into Pat's body (*o* is the medication's entering his body), and *P* is her restoring Pat's health (*p* is the improvement in his health). As with the case of BILL AND JILL, it is the first of these events (*L* or *l*), and only the first, over which Dot exercises direct control; her control over *m–p* is indirect, while her control over *M–P* is hybrid. Yet, it is plausible to hold, what Dot is fundamentally, originally, and thus directly obligated to do is to perform *P*, that is, to restore Pat's health.<sup>5</sup> If we assume that each

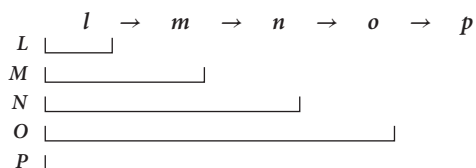


Figure 5.1 DOT AND PAT

<sup>3</sup> A more traditional way of putting this is to say that pain is *intrinsically* bad. But this suggests that the value that something has for its own sake must supervene entirely on its intrinsic properties, a claim that has been disputed. See Rabinowicz and Rønnow-Rasmussen (1999).

<sup>4</sup> Why all else being equal? Because otherwise the blow might have other consequences whose final goodness matches or outweighs the pain's final badness.

<sup>5</sup> This may be a mistake. Perhaps we should say that her fundamental obligation consists in performing some further act, *F*, such as showing respect for Pat or raising his level of welfare. On the distinction between fundamental and incidental obligation, see Ross (2002), p. 46.

of  $L-O$  is a necessary means to her doing  $P$ , then we can presumably say that Dot is obligated to do each of these things, too, but only incidentally, derivatively, indirectly. If this is so, then, even if moral obligation requires control, it will nonetheless be the case that what Dot is directly obligated to do is something that lies only in her indirect control, whereas what is in her direct control is something that she is only indirectly obligated to do.

Some philosophers would reject the view that I have just sketched about the nature of direct and indirect obligation, on the grounds that one cannot be indirectly obligated to do anything, and that what one is (directly) obligated to do is, and can only ever be, to make an executive decision to do something.<sup>6</sup> If this were so, then there would be no discrepancy after all between obligation and blameworthiness and their respective relations to direct control. But their argument itself rests on a mistake, I think, the mistake of confusing the distinction between direct and indirect control with the distinction between complete and partial control. They appear to hold that only executive decisions can be obligatory because only such decisions can, strictly speaking, be (fully) in one's control. But that is not so. As I noted in §4.1.4, complete control is unattainable by anyone at any time; decisions are just as vulnerable to the vicissitudes of nature (e.g., a sneeze that prevents one from making a decision that one was just about to make) as any other event that might lie within one's at-best-partial control. There is therefore no reason, on this score at least, to abandon our normal conception of moral obligation and say that it is not actions (or omissions), but only executive decisions, that we can ever be obligated to carry out. And if we stick with this conception, then we must acknowledge the distinction between direct (fundamental) and indirect (incidental) obligation. But then the question arises once again: Why should it be that the distinction between direct and indirect blameworthiness tracks the distinction between direct and indirect control, when the distinction between direct and indirect obligation does not?

The answer, I suggest, is this. As I noted in §2.1.3, judgments about blameworthiness, and about moral responsibility in general, are agent-focused, whereas judgments about moral obligation are act-focused. One's blameworthiness depends crucially on the quality of will that one displays at the time at which one exercises control and makes the decision to act, and it is therefore at this stage in the history of one's behavior that one's blameworthiness is rooted. One's moral obligations do not depend in this way on one's quality of will. On the contrary, they depend on the actual or probable outcomes of the various options one faces. Some of these outcomes are of fundamental significance, whereas others have only

<sup>6</sup> Cf. Prichard (1949), pp. 31 ff.; Ross (1939), pp. 153 f.; Khoury (2018), p. 1376. These authors talk in terms of "volitions" or "self-exertions" or "willings" rather than "executive decisions." I take these terms to be synonymous.

incidental significance, and it is this difference between them that determines whether an obligation is direct or indirect.

### 5.1.3 Foreseeability

The claim that one is indirectly blameworthy only for what was in one's indirect control of course does not imply that one is indirectly blameworthy for everything that was in one's indirect control. The latter claim is patently absurd, since it implies that one is to blame for all the good things that have happened as a result of one's exercise of control. Also unacceptable is the more restrictive claim that one is to blame for all the bad things that have happened as a result of one's exercise of control. Think of Sally from the first chapter. Her decision to take thalidomide resulted in her child's phocomelia—a tragic consequence. Yet surely Sally is not to blame for her decision and hence not to blame for this consequence of it. Unacceptable, too, is the still more restrictive claim that one is to blame for all the bad things that have happened as a result of an exercise of control for which one is directly to blame. The reason is that, as noted in Chapter 1, some of these things, despite being bad and despite having occurred as a consequence of an executive decision for which one is directly to blame, may not have “fallen within the risk” created by that decision.<sup>7</sup> Indirect blameworthiness, it seems, is restricted to consequences that were *foreseeable* by the agent at the time of the pertinent decision. This is a very common, and seemingly very plausible, claim. But exactly what is it for a consequence to be foreseeable?

One answer to this question, often given in legal contexts, is that a consequence is foreseeable just in case a (or, sometimes, the) reasonable person would have foreseen it. Whatever the merits of this answer for legal purposes (adopting it might facilitate administration of the law, for example<sup>8</sup>), it faces several problems when applied to the present context. One problem is simply that of determining what constitutes a reasonable person. Another problem is that of determining what such a person would foresee under whatever the relevant circumstances may be. But even if these problems can be resolved, the gravest problem remains, and that is that the reasonable-person standard is simply irrelevant to the question of whether and, if so, to what extent one is blameworthy for some decision or the consequences thereof, since one can blamelessly fall short of this standard. Simple Simon, for example, is not to blame, morally or otherwise, for trying to catch water in a sieve, even if a (or the) reasonable person would have foreseen the futility of doing so.<sup>9</sup>

The problem with the reasonable-person approach to foreseeability is, of course, that it fails to accommodate the basic point that, for a consequence to be

<sup>7</sup> Cf. American Law Institute (1985), §2.03(3).

<sup>8</sup> Cf. Prosser (1971), pp. 152 f.

<sup>9</sup> Cf. Keeton and O'Connell (1975), p. 85; Alexander et al. (2009), pp. 81 ff.

foreseeable by someone, it must be the case that *that* person *can* foresee it. But this rather obvious observation doesn't suffice to answer our original question of what it is for a consequence to be foreseeable, unless it is specified what sense of "can" is at issue.

It's safe to say, I think, that, if one actually foresaw a consequence, then one could have foreseen it, in any pertinent sense of "can." But, of course, what we're primarily concerned with in the present context is with one's being to blame for a consequence that one *didn't* foresee, even though it was foreseeable. I think that it will be generally agreed that such a situation cannot arise unless one is to blame for one's lack of foresight; that is, one's blameworthiness for the foreseeable-but-unforeseen consequence is to be traced to one's failure to foresee it. Now, there are perhaps many different senses of "can" that might be at issue when it is claimed that someone could have foreseen something, but, if culpably failing to foresee some allegedly foreseeable consequence is to be the occasion of one's being to blame for that consequence, and if (as stated in Premise 4 of the Argument from Ignorance) one is to blame for something only if one was in control of that thing, then it must be that one could have foreseen the consequence in that particular sense of "can" that expresses the specific ability involved in one's having control over something. (See §4.1.2.) Now, foreseeing a consequence consists in believing that it would or might occur as a result of some other event's occurring, and (as stated in Premise 6 of the Argument from Ignorance and as argued in §4.6.5), one is never directly in control of whether one fails to believe something. Hence one's being to blame for one's lack of foresight is to be traced in turn to one's being to blame for some prior piece of behavior, *B*, consisting in either doing something that caused one's lack of foresight or, more likely, failing to do something that would have prevented it (such as failing to keep one's eyes on the road while driving, for example<sup>10</sup>). If one is to blame for *B*, then *B* must have been an instance of wrongdoing; that is, it must have been something that one ought not to have done. (In saying this, I am for the time being putting aside the possibility of one's having an excuse for *B*, regarding which see §3.4.1.) This observation is of course perfectly in keeping with the common idea that one should have or ought to have foreseen certain consequences that one didn't foresee (one ought to have foreseen the accident that resulted from one's not keeping one's eyes on the road, for example). But here I think it's very important to point out something that is unfortunately frequently overlooked,<sup>11</sup> and that is that from the fact that one *ought* to have foreseen some consequence it doesn't follow that one is to *blame* for failing to foresee it. One might have an excuse! Indeed, the Argument from

<sup>10</sup> Cf. Rosen (2004), p. 301, on what he calls procedural epistemic obligations.

<sup>11</sup> See, among many others: M. Smith (2003), p. 31; Guerrero (2007), pp. 64, 66; Sher (2009), p. 87; Murray and Vargas (2020), pp. 831 ff., 840. For a welcome contrast, see Goldberg (2017), pp. 2885 ff.

Ignorance implies that one *will* have an excuse if one was inculpably ignorant of the wrongness of *B*.

Even given this clarification of the pertinent sense of “can,” there remain complications with aligning blameworthiness for the consequences of one’s behavior with those consequences that were foreseeable. I mentioned these complications in Chapter 1, but it will be useful to repeat them here. Consider again the condition, *C*, of Sally’s poor child. As I noted in Chapter 1, it cannot reasonably be held that any of Grünenthal’s employees could have foreseen that *C* itself would occur, since none of them knew or could have known of Sally, let alone that she was pregnant and thinking of taking Distaval. Any pertinent consequence that can plausibly be said to have been foreseeable by any of these employees must therefore be of a more general character, one that encompassed *C* but also other consequences relevantly similar to *C*. (For example, even if Heinrich could not have foreseen that *C* would occur as a result of failing to disclose the results of his research, perhaps he could have foreseen that the incidence of phocomelia would rise.) Moreover, as I also noted in Chapter 1, it would presumably be too restrictive to insist that, in order for any employee to be to blame for what happened, that person must have been able to foresee that this more general consequence *would* occur; that such a consequence *might well* occur would seem to suffice. These points are commonly recognized,<sup>12</sup> but they raise three problems that are less commonly acknowledged.

The first problem has to do with how to understand the “well” in “might well.” I will postpone discussion of this issue until the next chapter.

The second problem has to do with matching that which was foreseeable with that for which one is to blame. I said earlier that it seems that indirect blameworthiness is restricted to consequences that were foreseeable by the agent at the time of the pertinent decision, and yet just now I also said that none of Grünenthal’s employees could have foreseen *C* itself; they could at best have foreseen a more general consequence that encompassed *C*. But if we want to keep open the question of whether any of these employees is to blame for *C* itself, then we *cannot* after all say that indirect blameworthiness is restricted to consequences that were foreseeable by the agent. Perhaps this is not too big a problem, though. To fix it, perhaps all we need say is this: blameworthiness is restricted to consequences that were, at the time of the pertinent decision, either foreseeable by the agent or encompassed by consequences that were foreseeable by the agent. This would allow us to say that, if Heinrich was to blame for the increase in the number of cases of phocomelia that resulted from his decision to keep the results of his research to himself, then he was also to blame for *C* in particular.

<sup>12</sup> See, for instance, Fischer and Tognazzini (2009), pp. 537 f. and p. 546.



The third problem is seldom discussed.<sup>13</sup> It has to do with the fact that there would seem to be a limit to the proper employment of the strategy of generalization that was just used to justify holding Heinrich blameworthy for *C* even though he could not have foreseen *C* itself. As I said in Chapter 1, if further research by Grünenthal's employees would have revealed that taking thalidomide might well result in polyneuritis but *not* that it might well result in phocomelia, then, I suspect, most people would *not* be willing to say that *C* "fell within the risk" created by the decision not to pursue further research, even though there is a general consequence of this decision, namely, the occurrence of harm, which presumably the employees could have foreseen and which of course encompasses not only incidents of polyneuritis but also incidents of phocomelia and, thus, *C* itself.<sup>14</sup> That is why I said that the correct employment of the strategy apparently requires that one restrict its application to a certain "vicinity" beyond whose borders the occurrence of any particular consequence would be something for which one is not to blame. Without further development, this idea of course remains very vague, but I confess that I don't know how to sharpen it. That, as I said, is a problem. Just how big a problem it is is something that I will now briefly address.

#### 5.1.4 Moral luck

I have been assuming throughout that there is good reason to draw a distinction between direct and indirect blameworthiness, but this assumption can be, and has been, challenged. Andrew Khoury has recently argued that one cannot be blameworthy for the consequences of one's behavior and hence, in effect, that there cannot be anything for which one is indirectly blameworthy. This is, of course, a very radical claim. It requires saying, for instance, not only that no one is to blame, but also that no one *could* be to blame, for the tragedy of thalidomide; that no one could *ever* be to blame for anyone's suffering; and so on. This view may seem preposterous, and yet Khoury's argument for it is straightforward. It is this:

- (1) If some thing is not that in virtue of which one is blameworthy then one cannot be blameworthy for that thing.

<sup>13</sup> But not never. See M. S. Moore (1997), pp. 389 ff.

<sup>14</sup> Of course, just how to determine whether some consequence "falls within the risk" created by a decision is a difficult question about which even experts can differ. Consider, for example, the celebrated case of *Palsgraf v. Long Island Railroad Co.* (248 N.Y. 339 (1928)), in which the facts were these. While helping a man board a departing train, an employee of the company dislodged a package wrapped in newspaper that the man was carrying. The package, which contained fireworks, fell onto the rails and exploded. The shock rocked some scales standing some distance away, thereby causing the plaintiff injuries for which she sued. The judges on the New York Court of Appeals were divided 4 to 3 regarding whether the plaintiff's complaint should be dismissed.

- (2) The consequences of one's actions are not that in virtue of which one is blameworthy.
- (3) Therefore, one cannot be blameworthy for the consequences of one's actions.<sup>15</sup>

What should we make of this argument?

One concern that might be raised regarding this argument is that it begs the question at issue. As I noted in §2.2.1, some philosophers have held that the basis of blameworthiness (that in virtue of which one is to blame) and the object of blameworthiness (that for which one is to blame) are one and the same. But, as I also noted there, if this claim is to have any plausibility, it must be restricted to direct blameworthiness. It's clear, for example, that the basis of Heinrich's blameworthiness for the condition of Sally's child cannot simply be that condition itself. The same point applies to Khoury's first premise. But then, it may seem, that premise alone suffices for Khoury's conclusion.

Although I think there may be something to the charge that Khoury's argument begs the question, I am hesitant to endorse it. That is because what constitutes begging the question is a complex and delicate matter, in virtue of which the charge is often very difficult to adjudicate. (Even circular arguments can fail to be question-begging.<sup>16</sup>) I do find Khoury's first premise problematic for a different reason, though. In §2.2.4 I argued that the claim that the basis and the object of blameworthiness are one and the same should be rejected, even when that claim is restricted to *direct* blameworthiness. I won't repeat my argument here, but it applies equally to Khoury's first premise.

I am strongly inclined to think that Khoury's second premise is *true*, however. He bases it on the view that the basis of blameworthiness is the quality of will that informs one's behavior. If that is so, then, since nothing that occurs as a consequence of one's behavior can alter or affect the quality of will that informs that behavior, nothing that occurs as a consequence of one's behavior can alter or affect one's blameworthiness. I find this line of reasoning very appealing. Now it's true, as I pointed out in §2.2.3, that, unless quality of will is understood to comprise not just the mental but also the agential condition of moral responsibility, it would be a mistake to say that the basis of blameworthiness has *exclusively* to do with one's quality of will, and it might seem that this fact undermines Khoury's second premise; but I don't think it does, for a complementary line of reasoning can be offered regarding the agential condition. Let me explain.

In a famous paper about moral luck, Thomas Nagel asks, "How is it possible to be more or less culpable depending on whether... a bird [gets] into the path of one's bullet?" I take the answer to this question to be that this is *not* possible. Consider again, from §4.5.1, the case of BILL AND THE BIRD, represented in

<sup>15</sup> Khoury (2012), p. 195. Cf. also Khoury (2018), p. 1363.

<sup>16</sup> See Sinnott-Armstrong (1999) for helpful and insightful discussion.

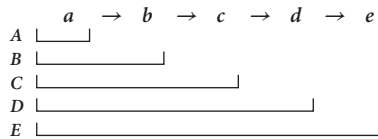


Figure 4.5 BILL AND THE BIRD

Figure 4.5. As I noted before, there is of course more to the case than what is represented in this figure. For present purposes, however, I will assume that what is left out is morally irrelevant, in that nothing further occurred for which Bill may plausibly be said to be culpable. (If the lives of birds have value, then this may be a mistake, but let's just put this concern to one side.) Question: Would there be any reason to declare Bill less culpable for what happened in BILL AND THE BIRD than for what happened in BILL AND JILL, depicted in Figure 2.1, in which he succeeded in killing Jill? Answer: Not that I can see. First, in keeping with Khoury's point, the absence in BILL AND THE BIRD of events *f* and *g* (the events of the bullet's piercing Jill's body and of Jill's death, respectively) does not affect or alter the quality of will that attached to his decision, *a*. Second, since all the events that occurred as a consequence of *a* were in Bill's control only to the extent that *a* itself was, the absence in BILL AND THE BIRD of events *f* and *g* does not affect or alter the extent to which Bill exercised control over what took place. True, in BILL AND JILL these events fell within the *scope* of his control, whereas in BILL AND THE BIRD they did not, but the *degree* to which Bill exercised control was the same in both cases (in *one* important sense of "degree of control"; for discussion, see §4.5). For this reason, and for the reason stated by Khoury, there is strong pressure to say that the *degree* to which Bill is to blame for what happened is the same in both cases. Thus, even if we reject Khoury's first premise (as I do) and allow for the possibility of indirect blameworthiness and thereby concede that the *scope* of Bill's blameworthiness in the original case (encompassing, as it does, all the events depicted in Figure 2.1) is greater than the scope of his blameworthiness in the modified case—so that, as we might put it, he is *to blame for more* in the former than in the latter—it seems nonetheless very plausible to maintain that he is *no more to blame* in the former (or, if that sounds objectionably lenient, that he is no less to blame in the latter), and thus that, more generally, one cannot, in Nagel's terminology, be *morally* lucky (or unlucky) in virtue of what happens to transpire as a consequence of the decisions one makes.

Once again, it may help to appreciate the point that I am trying to make if the case of BILL AND JILL is compared with the case of BILL'S PERSISTENCE, represented in Figure 4.6. I noted in §4.5.1 that, given the repeated opportunities that Bill has in BILL'S PERSISTENCE to administer an antidote to the poison that he has injected into Jill and his repeated decisions to refrain from doing so, his degree of control over

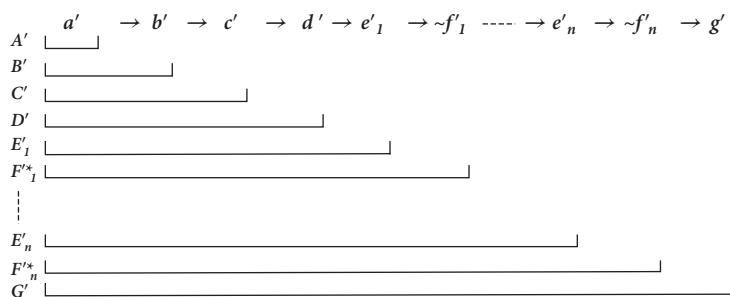


Figure 4.6 BILL'S PERSISTENCE

Jill's death in this case is greater than it is in BILL AND JILL. Given that his quality of will remains constant throughout, he incurs culpability afresh with each decision, and so his culpability for her death will be correspondingly greater, too.

In an earlier work, I summarized the view that there is no moral luck as far as responsibility is concerned as follows: "Degree of responsibility counts for everything, scope for nothing, when it comes to [the] moral evaluation of agents."<sup>17</sup> Khoury has himself objected to this statement (which is in tension with the first premise in his argument), on two grounds. He claims, first, that it is implausible to say that degree of responsibility counts for everything, and, second, that it is implausible to say that scope of responsibility counts for nothing.<sup>18</sup>

Regarding the first point: I have claimed that Bill is blameworthy to the same degree in both BILL AND JILL and BILL AND THE BIRD. Khoury agrees with this verdict. But, he says, there being no difference in Bill's blameworthiness in either case isn't merely a matter of Bill's being blameworthy to the same degree; it's also a matter of Bill's exhibiting the same quality of will. Another agent could exhibit a different quality of will, and thus be blameworthy in a different way, and yet be blameworthy to the same degree. Since one's being blameworthy is a function of the way in which one is blameworthy as well as the degree to which one is blameworthy, it's false to claim that degree of responsibility "counts for everything" in the moral evaluation of agents.

I think this objection misfires. There are two interpretations to be given to the phrase "the way in which one is blameworthy." On one reading, this refers to the *basis* of one's blameworthiness—the way in which one incurs blameworthiness; on another reading, it refers to the *mode of blame* of which one is worthy—the way in which one is to be blamed. Khoury intends the phrase to be understood according to the first interpretation (he holds fixed the mode of blame he takes into

<sup>17</sup> Zimmerman (2002), p. 568. The kind of moral evaluation in question is, of course, hypological rather than aretaic. See §2.1.3 for discussion.

<sup>18</sup> Khoury (2018), pp. 1361 ff.

consideration, namely, resentment); he is concerned with the basis of one's blameworthiness, which, as I noted, he takes to be the quality of will that one exhibits in one's behavior. Now, I am quite willing to concede that two agents can exhibit different qualities of will and yet be blameworthy to the same degree, but, if they are also worthy of the same mode of blame, then that, as far as I can see, exhausts the question of how they are to be evaluated morally: they are to be blamed in the *same* way and to the *same* degree. The fact that this evaluation has a different basis in one case than in the other doesn't feature in the evaluation itself.

But what if two agents are to be blamed in different ways? What if, for example, *P* is worthy of resentment whereas *Q* deserves a reprimand? Then, as I noted in §2.5.3, it's not clear how the degree to which *P* is to blame can be compared to the degree to which *Q* is to blame. Even so, it is presumably the case that *P* is to blame to some degree and that *Q* is also to blame to some degree, and, given that the mode of blame of which each is worthy has been established, once again that would appear to exhaust the question of how each of these agents is to be evaluated morally. Still, I concede that, given that there are different kinds or modes of blame and that someone can be worthy of one kind but not of another, I oversimplified matters (in a way that Khoury does not accuse me of) when I said that degree of responsibility counts for everything in the moral evaluation of agents. The kind of response of which they are worthy counts, too.

Oversimplified or not, my main aim in stating that degree of responsibility counts for everything in such evaluation was simply that of driving home the point that scope of responsibility counts for nothing. But this brings us immediately to Khoury's second complaint:

If, in light of the rejection of...moral luck, we can retain the possibility of responsibility for external objects [i.e., objects that are the consequences of "internal" decisions] only at the cost of making such responsibility empty, why bother?<sup>19</sup>

This is a fair question. Here's my answer.

The idea that indirect responsibility is "essentially empty," as I have put it in previous work,<sup>20</sup> may be initially disconcerting, but one's unease with it should diminish when one recognizes that the emptiness of indirect responsibility is just one instance of a phenomenon that is exemplified elsewhere. Consider, for example, the distinction I mentioned earlier between final and non-final value. Suppose, as before, that pain is finally bad, i.e., bad for its own sake. Then, I said, something that causes pain will, all else being equal, be non-finally (more particularly, instrumentally) bad. Now consider these two scenarios. One is

<sup>19</sup> Khoury (2018), p. 1363.

<sup>20</sup> Zimmerman (1988), p. 55.

straightforward: Moe slaps Curly on the back of the head, causing him pain. The other involves an elaborate Rube Goldberg-type of apparatus: Larry lifts his finger, thereby rolling a ball down a chute, thereby setting a wheel in motion, thereby tripping a switch, thereby . . . , thereby releasing a paddle that delivers a slap to the back of Curly's head that causes exactly the same quantity and quality of pain to him as in the first scenario. Larry clearly does *more bad things* than Moe; all the actions mentioned are instrumentally bad. But, equally clearly, Larry does not do *more bad* than Moe;<sup>21</sup> the sum total of value in the world is reduced to the same extent in each scenario. Non-final value, being merely indirect, is thus essentially empty. Note that this point applies to positive value just as it does to negative value. Just as one person can do more bad things than another without doing more bad, so, too, one person can do more good things than another without doing more good. Suppose that pleasure is finally good, and then consider two different ways in which the same quantity and quality of pleasure might be caused, one straightforward, the other elaborate.

There is another way in which doing more good (or bad) things does not entail doing more good (or bad). Suppose that, in addition to pleasure (pain), knowledge (ignorance) is also finally good (bad). Then imagine that, in doing some act *A*, John brings about a trivial instance of each of these goods (bads), while, in doing some other act *B*, Jane brings about a substantial instance of just one of them. Then John does more goods (bads) than Jane,<sup>22</sup> but Jane does more good (bad) than John. This kind of case is quite different from the kind just mentioned. It is only non-final (indirect) goods and bads that are essentially empty. When it comes to final goods (bads), the more, the better (worse),<sup>23</sup> even if one good (bad), such as that brought about by Jane, can outweigh two goods (bads), such as those brought about by John.

We can couch the case of John and Jane in terms of reasons. John has *more reasons* (not) to do *A* than Jane has (not) to do *B*, yet Jane has *more reason* (not) to do *B* than John has (not) to do *A*. This fact does *not* point to any essential emptiness in the nature of reasons. Nonetheless, I believe that there *is* such a phenomenon: we can distinguish between direct and indirect reasons, and the latter are essentially empty. For example, one may have a direct reason (not) to cause pleasure (pain) and only an indirect reason to perform any of the means at one's disposal to accomplish this. As far as I can tell, however, the term "more reasons" as opposed to the term "more reason" cannot be used in any natural way to capture this distinction.

<sup>21</sup> Please pardon my English. "Do more bad" is a somewhat questionable phrase. "Do more evil" is less questionable, but "evil" itself carries connotations that I wish to avoid in this context.

<sup>22</sup> Once again, please pardon my English.

<sup>23</sup> This is an oversimplification. The claim should be restricted to *basic* final goods and bads. For clarification, see Zimmerman (2001), ch. 5.

Now consider the distinction, also mentioned earlier, between fundamental and incidental obligation. Suppose that Dot has two patients to whom she must administer an injection, Pat and Paul. Pat's case, as depicted in Figure 5.1, is relatively straightforward. Paul's case is quite different; Dot must make use of an elaborate method that involves many different steps. Suppose that in each case Dot does *none* of the things needed to ensure that either Pat or Paul receives the required medication. Then clearly she will have done *more wrongs* in the case of Paul than in the case of Pat. But, equally clearly, she will not have done *more wrong* in the former case than in the latter. Incidental obligation, being merely indirect, is essentially empty.

Should the essential emptiness of non-final value, of indirect reasons, and of incidental obligation lead us to reject the practice of making judgments in terms of these concepts? I see no reason to think so; indeed, it's hard to imagine how we could manage our lives without recourse to them. (Try eschewing such judgments as that it is good to donate to charity, wrong to beat a child, and so on.) So too, I think, for the concept of indirect moral responsibility. I see no need to follow Khoury and deny that we can ever be morally responsible for the consequences of our actions, especially when it's clear that we can be, indeed are, causally responsible for these consequences and it is at least arguable that the conjunction of moral with causal responsibility can render certain responses fitting that causal responsibility alone would not.<sup>24</sup> Nonetheless I concede that, in rejecting the possibility of moral luck, my view, though shared with several others,<sup>25</sup> is pretty radical, since it is in conflict with conventional "wisdom," according to which, for example, Bill would be far more culpable if he had succeeded in killing Jill than if he had tried to do so but, due to the passing bird, failed in his attempt.

I have said that in the case of BILL AND JILL, Bill is to blame for Jill's death. Clearly this is not true in the case of BILL AND THE BIRD, since in that case Jill didn't die. I have also said, however, that Bill is no more to blame in the former case than in the latter, since the degree to which he is to blame in both cases is exhausted by his blameworthiness for his decision. But then the question arises: To what degree is Bill to blame for Jill's death in BILL AND JILL? If he is no more to blame than he would have been if she hadn't died, then shouldn't we say that he is to blame to *no* degree for her death? But that would seem to imply that he is *not* to blame for it after all.

My answer is that we should *not* say that Bill is to blame to no degree for Jill's death. To use numbers in a purely ad hoc fashion, suppose that Bill is to blame to degree 10 for his decision to kill Jill. Then I propose that we also say that he is to blame to degree 10 for her death, for this captures precisely the degree to which he

<sup>24</sup> Cf. Scanlon (2008), pp. 148 ff.; Coleman and Sarch (2012), p. 116, n. 26.

<sup>25</sup> See, e.g., Feinberg (1970), ch. 2; Thomson (1989); Enoch and Marmor (2007); Alexander et al. (2009), ch. 5.

is to blame for what transpired. In order to fend off any suggestion that Bill's being to blame to degree 10 for Jill's death somehow compounds the degree (i.e., degree 10) to which he is already to blame for his decision, we need only note that he is merely *indirectly* to blame to degree 10 for the death, whereas he is *directly* to blame to degree 10 for the decision. The situation here is once again analogous to the situation elsewhere. Consider, for example, the case involving Moe and Curly. Suppose that Curly's pain was bad to degree 10. Then I propose that we also say that Moe's slap was bad to degree 10, since that captures precisely the degree to which what happened was bad.<sup>26</sup> To fend off any suggestion that the slap's being bad somehow compounds the degree to which the train of events was bad, we need only note that the slap was merely *instrumentally* bad to degree 10, whereas the pain was *finally* bad to degree 10.

Notice one distinct advantage of my view regarding the essential emptiness of indirect blameworthiness: the need to sharpen the notion of a "vicinity" of consequences for which one is to blame is far less urgent if blameworthiness for consequences is essentially empty than if it is not; for the failure in any particular case to map the borders of the relevant vicinity accurately need not result in any injustices, as long as the degree to which one is blamed is not thereby affected. Notice, though, one distinct limit to this advantage: if in fact the conjunction of moral with causal responsibility renders certain responses fitting that causal responsibility alone does not, then such a failure could still have significant repercussions, since causal responsibility itself extends to all consequences, both to those that lie outside of the relevant vicinity and to those that lie within its borders.

I don't pretend that these brief remarks prove conventional "wisdom" to be mistaken. It is arguable that the quotation marks should be removed and the wisdom acknowledged as genuine. If so, then, even though the distinction between scope and degree of blameworthiness would remain, nonetheless there would be a positive correlation between them: the greater the scope, the greater the degree. In this way, Bill would be properly declared more blameworthy for what happened in the original case in which he killed Jill than in the case in which he failed to do so because the bird got in the way. Anyone who takes this stand is of course faced with the task of determining just which consequences of that behavior for which one is directly blameworthy are indeed consequences for which one is indirectly blameworthy and which thereby serve to augment the degree to which one is blameworthy, and of explaining why this is so. This task is perhaps not so

<sup>26</sup> Complication: Moe's slap might also have caused Larry anguish, something that was, say, bad to degree 5. Then I suppose we should say that the slap was bad to degree 15, since *this* is what would now reflect how bad what happened was. Further complication: What if the slap had yet more bad consequences? Then we would have to re-calculate once again. (The device of addition that I have used for purposes of this illustration is of course questionable. For discussion, see Zimmerman (2001), ch. 5.)



daunting if the consequences in question are limited to those that are foreseen by the agent, but, for anyone who claims that the occurrence of foreseeable-but-unforeseen consequences also augments the degree to which one is blameworthy, the “problem of the vicinity” looms large.

### 5.1.5 Tracing culpability for ignorant wrongdoing

So far, I have used the term “trace” liberally. I have said that, if ever one is indirectly to blame for some consequence of one’s behavior, one’s blameworthiness for the consequence is to be traced to one’s blameworthiness for the behavior. Philosophers often use the term more narrowly, restricting it to cases in which the consequence is *itself* some form of behavior. It is, of course, precisely such cases with which the Argument from Ignorance is concerned. In her classic paper, “Culpable Ignorance,” Holly Smith introduced terminology in this context that has since become standard. In cases in which one is to blame for ignorant wrongdoing, the ignorant wrongdoing is something that Smith calls an “unwitting wrongful act,”<sup>27</sup> one’s blameworthiness for which is to be traced to one’s blameworthiness for one’s ignorance; and one’s blameworthiness for one’s ignorance is in turn to be traced to one’s blameworthiness for some prior act or omission which Smith calls a “benighting act.”<sup>28</sup> This account accords perfectly with the Argument from Ignorance, although it is notable that Smith doesn’t draw the conclusion that the benighting act must itself be a *witting* wrongful act.

Smith’s paper is concerned with the *epistemic* condition of moral responsibility; she investigates how blameworthiness for *ignorant* wrongdoing is to be traced to blameworthiness for some prior piece of behavior. Those who favor the narrow use of the term “trace” do not restrict it to such cases, however. They are often, indeed usually, concerned with the *agential* condition of moral responsibility and invoke tracing in order to account for how blameworthiness for *out-of-control* wrongdoing is possible. Such wrongdoing might be unwitting, but it needn’t be. The stock example is that of the drunk driver who cannot control his driving but is nonetheless to blame for it, since he is to blame for the drinking that led to it. This example is problematic, however, since, as many have remarked,<sup>29</sup> it’s not clear whether such cases do typically involve behavior that, at the time at which it occurs, is strictly out of the agent’s control. So, instead, let us consider a cleaner case, the case of Alice that I first presented in Chapter 1. She declined to take her medication for Tourette’s and, as a result, hurled invective at Ben. Let me now simply stipulate that her decision not to take her medication was something that

<sup>27</sup> Here the term “act” is to be construed broadly, as encompassing omissions as well as acts in the strict sense.

<sup>28</sup> H. M. Smith (1983), pp. 547 f.

<sup>29</sup> See, e.g., Husak (2012).

was in Alice's direct control and which she knew to be wrong, whereas her hurling invective at Ben was, at the time it occurred, wholly outside of her control. Then, even if Alice was fully aware at the time that hurling invective at Ben was wrong, nonetheless her blameworthiness for this behavior can only be indirect; it is to be traced to her being to blame for her decision not to take her medication.

It's worth noting that, even if one is directly responsible for some later piece of behavior, that would not preclude tracing one's responsibility for it to some earlier piece of behavior for which one is directly responsible, so that one would *also* be indirectly responsible for one's later behavior. This point is often overlooked. Consider Beatrice who, like Alice, is subject to strong impulses to sling insults at others. In her case, though, the impulses are not irresistible. Like Alice, Beatrice has medication that helps eliminate these impulses; but it has a nasty taste, and so she is reluctant to take it. This morning, Beatrice decided to skip her medication. This afternoon, as Derek walked by, she had a strong impulse to insult him—strong, but resistible, though unfortunately not resisted. Let us assume that Beatrice is directly to blame for her decision to insult Derek. If that decision was a consequence of her earlier decision to skip her medication, then she may also be indirectly to blame for her later decision.

It might be objected that declaring Beatrice both directly and indirectly to blame for her decision to insult Derek involves an illegitimate form of double-counting, but that is not so, precisely because there were two occasions on which she incurred blameworthiness—those occasions of course being the time at which she made the decision to insult Derek and the earlier time at which she decided to skip her medication. Whether her being *twice* to blame for her later decision *increases* her blameworthiness for this decision is an interesting question. Those who, like myself, take indirect responsibility to be essentially empty will answer in the negative. Others may respond differently.

On the present approach, then, tracing in the narrow sense (tracing blameworthiness for later behavior to blameworthiness for earlier behavior) is simply a particular species of tracing in the broad sense (tracing blameworthiness for consequences in general to blameworthiness for behavior that led to these consequences). Given the possibility of indirect responsibility for consequences in general, there would seem to be nothing notable or surprising about the idea that one can be indirectly responsible for one's own behavior in particular. And yet this idea has recently been the target of several objections.

One objection, raised by Matt King in a pair of articles, is that tracing responsibility (in the narrow sense of "trace"—the only sense in which he uses the term) is theoretically dispensable, both in cases of out-of-control wrongful action and in cases of unwitting wrongful action.<sup>30</sup> He claims that an agent's blameworthiness

<sup>30</sup> King (2011) and (2017).

for such wrongful action can be fully accounted for by registering his or her blameworthiness for having behaved either recklessly or negligently in committing the benighting act that led to the later wrongful act, and hence that appealing to (narrow) tracing in this context is to make use of a special explanatory tool that serves no purpose. This objection misfires, though, precisely because appealing to (narrow) tracing is not to make use of some special explanatory tool; on the contrary, doing so simply involves applying the ordinary explanatory tool of tracing blameworthiness for consequences in general to that particular kind of case in which the consequence at issue consists in the agent's own wrongful behavior. (If some of those who invoke tracing in the narrow sense were to suggest otherwise, then I think King would be quite right to complain. But I'm not aware of anyone's making this suggestion.) I would add, moreover, that King's claim that all such cases involve either recklessness or negligence in the commission of the benighting act is mistaken. It is easy to imagine cases in which one commits a benighting act with the knowledge, and perhaps even the intention, that doing so will lead to one's committing a later wrongful act that is either unwitting or out of one's control (or both).

In objecting to resorting to the device of tracing (in the narrow sense), King does not claim that doing so is mistaken, but only that it is unnecessary, when accounting for an agent's blameworthiness for wrongful acts that were either unwitting or out of the agent's control. Although I just objected to his argument for this claim (on the grounds that he mistakenly takes proponents of tracing in the narrow sense to conceive of it as a special explanatory tool), the claim is nonetheless one with which I sympathize; for if, as I have argued, indirect responsibility is essentially empty, then one's being blameworthy for an unwitting wrongful act does not augment the degree to which one is already to blame for the benighting act that led to it. There are other philosophers, however, who have argued that resorting to the device of tracing (in the narrow sense—again, the only sense in which they tend to use the term) is not simply unnecessary but can in certain cases lead to mistaken assessments of responsibility.

One objection to this effect has been raised by Douglas Husak.<sup>31</sup> He is concerned with the degree of blameworthiness that advocates of narrow tracing assign to agents for unwitting wrongful acts committed as a result of some prior benighting act. (Husak focuses on cases in which the benighting "act" is a failure to inquire into the lawfulness of the behavior in which one eventually engages, but his concern has general application.) Husak asks us to compare two agents, whom I will call Wendy and Nathan. They both commit wrongful acts of the same kind, but, whereas Wendy does so wittingly, Nathan does so unwittingly as a result of some prior benighting act. Suppose that Wendy is blameworthy to degree 10 for

<sup>31</sup> Husak (2016), pp. 184 ff.

her wrongful act. What guarantee is there, Husak asks, that Nathan will likewise be blameworthy to degree 10 for *his* unwrongful act? If his blameworthiness for this act is to be traced to his blameworthiness for the prior benighting act, why assume that he will be blameworthy to degree 10 for that act? Isn't it on the contrary likely that he will only be blameworthy to a lesser degree for it, in virtue of the fact that in committing it he at best merely *risked* committing the later wrongful act? Suppose, then, that he was merely blameworthy to degree 5 for his benighting act. What reason could there then be to declare him blameworthy to degree 10 for his unwitting wrongful act?

Husak raises his objection in the context of a discussion of ignorance of law, and he directs it toward those who hold the view that agents such as Wendy and Nathan should be punished equally for their equally wrongful acts. He is surely right to question this view, but it's important to note that advocates of narrow tracing aren't committed to it. Certainly there is no reason to think that, if Wendy is blameworthy to degree 10 for her wrongful act, then Nathan must also be blameworthy to degree 10 for his. If, indeed, Nathan is only blameworthy to degree 5 for his benighting act, then, given my position on moral luck, I would say that he is also only blameworthy to degree 5 for the wrongful act that he committed as a result. Of course, a proponent of moral luck might want to say that Nathan is blameworthy to some higher degree for what transpired, but once again there is no need to think that his degree of blameworthiness must match Wendy's.

A second objection to narrow tracing has been raised by Craig Agule. It has to do with certain instances of what he calls Odysseus cases.<sup>32</sup> As the name implies, an Odysseus case is a case in which one incapacitates oneself in order to achieve some goal. The special sort of Odysseus case on which Agule focuses has the following features: first, in incapacitating oneself, one knowingly runs the risk of later committing some wrongdoing while one is incapacitated; second, one is not to blame for running this risk; and third, one commits the later wrongdoing. For example, one might justifiably consent to being anesthetized in order to undergo some necessary surgery, knowing that there is a risk of later committing some wrong (such as mistreating the nursing staff) while slowly emerging from the grip of the anesthetic, a risk that unfortunately turns out to be realized. Agule claims that in such a case one is not to blame for the wrongdoing that one commits while incapacitated, and yet the proponent of narrow tracing is committed to saying that one is to blame for this wrongdoing, since it was a foreseeable upshot of earlier behavior for which one is responsible.

This objection is mistaken. For the sake of argument, let's grant that one is morally responsible for incapacitating oneself. Then the proponent of tracing (whether narrow or broad) presumably is committed to saying that one is also

<sup>32</sup> Agule (2016).

morally responsible for one's later wrongdoing, since it was a foreseeable upshot of earlier behavior for which one is morally responsible. But recall from §2.1.1 that moral responsibility needn't take the particular form of blameworthiness. Even if the proponent of tracing agrees that one is responsible for one's later wrongdoing, he or she is surely not committed to saying that one is to *blame* for that wrongdoing, precisely because one is not to blame for the earlier behavior that led to it. On the contrary, an Odysseus case of this sort is just another case in which one has an excuse for doing something wrong. Indirect blameworthiness presupposes direct *blameworthiness*. If there is no direct blameworthiness for one's deciding to incapacitate oneself, then the proponent of narrow tracing can agree that there is indeed no indirect blameworthiness for any consequence of this decision (unless this consequence is also a consequence of something else for which one is to blame), whether or not such a consequence is itself an instance of wrongdoing.

Agule holds (incorrectly, as I have just argued) that invoking narrow tracing can lead to the mistake of declaring blameless agents blameworthy. Others have held that it can lead to the mistake of declaring blameworthy agents blameless. Manuel Vargas, in an article entitled "The Trouble with Tracing," offers a number of cases in support of this claim.<sup>33</sup> One such case involves a jerk called Jeff, whom Vargas describes as rude and inconsiderate about the feelings of others, but also as being unreflective about this moral failing, one that originated in a crucial formative period during Jeff's adolescence. When he was 15 years old, Jeff came to the realization that girls found him unattractive. In the belief that girls were only attracted to boys who exhibited what is in fact (although he didn't understand it to be) jerk-like behavior, Jeff deliberately undertook to become the sort of person who behaves in this way, and he succeeded in doing so with surprisingly little effort. Many years later it transpired that, as a middle-aged manager at a mid-sized company and with his jerk-like attitudes and habits deeply ingrained in his character, Jeff was faced with the task of laying off workers in order for the company to remain solvent. He went about doing so in his usual manner—rudely, inconsiderately, but unreflectively. Vargas claims that Jeff is to blame for firing the employees in the way he did, despite the fact that, having acted unreflectively, he was unaware at the time that it was wrong of him to do so. Moreover, Vargas says, Jeff's blameworthiness for this behavior cannot be traced to his blameworthiness for some prior piece of behavior at some suitable prior time, since the only time at which he exercised control over his developing into a jerk was a time at which he could not have foreseen that he would or might well treat his employees as he did.

John Martin Fischer and Neal Tognazzini, who are proponents of narrow tracing, agree with Vargas that Jeff is to blame for his later behavior, but they deny that tracing cannot account for this.<sup>34</sup> They point out that there are various

<sup>33</sup> Vargas (2005).

<sup>34</sup> Fischer and Tognazzini (2009).

ways in which the regrettable outcome of Jeff's hapless attempt to render himself attractive to girls might be characterized, among which are:

- (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.

And they claim that, the broader the characterization, the more plausible it is to think that Jeff could, even at 15 years old, have foreseen that he would or might well behave in the manner described. They conclude:

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. 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... 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!<sup>35</sup>

This response seems to me quite right—as far as it goes; I would of course want to apply the Argument from Ignorance to Jeff's case as to any other, and if Jeff was not only unaware that he was acting wrongly when firing his employees as he did but also unaware that he was acting wrongly when undertaking to reform his character as he did, then I would conclude that Jeff is *not* to blame for either bits of behavior (unless some other relevant witting wrongdoing occurred yet further back in his history). The approach adopted by Fischer and Tognazzini is the same as that which I mentioned earlier when discussing the possibility of some of

<sup>35</sup> Fischer and Tognazzini (2009), p. 538.

Grünenthal's employees being to blame for the condition of Sally's child, even though they could not have known anything about Sally and her child in particular. (It's worth noting, though, that, in their response to Vargas, Fischer and Tognazzini pay no heed to the complication posed by the "problem of the vicinity.")

In an article entitled "More Trouble with Tracing," Seth Shabo has argued that, even if the response by Fischer and Tognazzini succeeds in disposing of Vargas's case about Jeff, there are similar cases that are impervious to this response. He asks us to consider the case of Greg who, like Jeff, is insensitive to the needs and interests of others, but whose insensitivity is the result, not of his ever having decided to become, or even to try to become, this way, but rather of his instinctively emulating the mannerisms of his socially dominant classmates, in virtue of which he came to regard others as falling into one of two classes, "big shots" and "nobodies." When asked to compile and submit a report to his company's board on which employees' positions should be eliminated, Greg does so in a manner that is typical, showing profound insensitivity to the "nobodies" involved. Shabo states:

We... 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.<sup>36</sup>

Shabo recognizes that in Greg's case there is an outcome similar to Outcome 3 in Jeff's case, namely:

(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>37</sup>

But Shabo denies that this outcome was foreseeable by Greg at any point in his history, since, contrary to Jeff's case, there was no particular time at which the risk of this outcome should have been salient to Greg. Shabo maintains that Greg is nonetheless to blame for this outcome, even though it was unforeseeable, since it resulted from his failure to cultivate greater sensitivity to the needs and interests of others, something that he should have recognized he had an obligation to do.

In my view, Shabo's claim that Greg is to blame for Outcome 4 is unfounded. Even if it's true that Greg ought to have recognized the need to cultivate greater sensitivity to others rather than emulate the mannerisms of his classmates, he is not to blame for failing to do this, if, as is implicit in Shabo's description of the

<sup>36</sup> Shabo (2015), p. 997.

<sup>37</sup> Shabo (2015), p. 997.

case, he was unaware of the wrongness of this failure (and this failure cannot itself be traced to some prior witting wrongdoing). This is the verdict of the Argument from Ignorance, and I see no reason to reject it. (That is perhaps stated too boldly. For further discussion, see Chapter 9.) More to the present point, however, it would be a mistake to think that either Vargas or Shabo has raised a problem with the strategy of tracing responsibility, despite what the titles of their articles might be thought to imply. What they have raised doubts about is the view that blameworthiness for some outcome requires the foreseeability of that outcome. Vargas doesn't draw a distinction between direct and indirect blameworthiness; he simply declares Jeff blameworthy for the way in which he treated his company's employees. Since he does not draw this distinction, he has done nothing to impugn it. He has at best shown it to be a mistake to invoke it when trying to account for Jeff's blameworthiness. Shabo does draw this distinction; he explicitly claims that Greg is indirectly blameworthy (Shabo says "derivatively blameworthy") for the way in which he treated his company's employees. In saying this, he is presumably committed to claiming that Greg's blameworthiness for this behavior *is* to be traced to some earlier event for which he is blameworthy, namely, his failure to recognize the need to become more sensitive to the needs and interests of others. Shabo seems to deny this; he claims that cases like Greg's involve "*untraceable* responsibility transmissions."<sup>38</sup> He uses this phrase because he denies that in such cases there is any particular time to which responsibility can be traced back. But this seems to me highly misleading. If Greg's alleged blameworthiness for the way in which he treated his company's employees is indeed indirect (derivative), then it *must* be traceable to his direct blameworthiness for something else. This, I take it, is a *conceptual* truth.<sup>39</sup> Shabo himself identifies the something else in question: Greg's omission to cultivate greater sensitivity to the needs and interests of others. Now, it can admittedly be very difficult to specify just when it is that an omission occurs (although it is of course the case that, if an omission occurs, it occurs at some time or times). In Greg's case, there is likely no single time that can be identified as *the* time at which it occurred. Even so, if this omission is to play the role of that for which Greg is directly blameworthy and in virtue of which he is indirectly blameworthy for his treatment of his company's employees, then, far from being refuted, the propriety of appealing to tracing is confirmed.

## 5.2 Exculpatory culpable ignorance

According to the Argument from Ignorance, and in the terminology introduced by Smith, one is to blame for an unwitting wrongful act (and its

<sup>38</sup> Shabo (2015), p. 1009.

<sup>39</sup> As I said at the outset of §5.1.1, Premise 8 of the Argument from Ignorance is *definitionally* true.



consequences) *only if* one is to blame for the ignorance that underlay it, and one is to blame for this ignorance only if one is to blame for the benighting act that led to it. This proposition only tells us when one *isn't* to blame for an unwitting wrongful act—one isn't to blame for it when the ignorance that underlay it was excusable; it doesn't tell us when one *is* to blame for an unwitting wrongful act. I think many would find it plausible to supplement this proposition with the further proposition that one is to blame for one's unwitting wrongful act *if* it was a foreseeable consequence of the ignorance that underlay it, and this ignorance was a foreseeable consequence of the benighting act that led to it, and one is to blame for that benighting act.<sup>40</sup> Combined, these propositions entail the thesis that excusable ignorance provides an excuse for an unwitting wrongful act, but inexcusable ignorance may not. (I say "may not" rather than "does not," since in some perhaps hard-to-imagine cases the unwitting wrongful act may not be a foreseeable consequence of the ignorance in question.) In this section I will distinguish between two interpretations of this thesis and argue that, although it is acceptable on one of these interpretations, it is unacceptable on the other.

### 5.2.1 Does ignorance of wrongdoing ever provide an excuse for wrongdoing?

I have repeatedly characterized the Argument from Ignorance as implying that ignorance of wrongdoing provides an excuse for wrongdoing far more often than is commonly thought. Yet if only *excusable* ignorance provides such an excuse, isn't it strictly false to say simply that ignorance *itself* can do so?

In response, let me make two observations.

First, even if it is the case that an excuse stemming from ignorance is in the offing only when the ignorance in question is itself excusable, it seems to me not seriously misleading to say that it is the ignorance that, under those conditions, provides the excuse. The situation here is in some respects similar to a situation in which we say that a short circuit caused a fire, even though we understand that there would have been no fire if the short circuit had occurred under different conditions (lack of oxygen, lack of flammable material, etc.).

Second, as I have just noted and will shortly argue, although the claim that only excusable ignorance provides an excuse for wrongdoing is indeed true on one interpretation, it is false on another. That is, on one interpretation, even inexcusable ignorance of wrongdoing provides an excuse for wrongdoing. On this

<sup>40</sup> This further proposition presumably needs qualification in light of the remarks made in §5.1.3 about some consequences encompassing others and about a vicinity of consequences. Consider such qualification implicit in what follows.

interpretation, the claim that ignorance of wrongdoing provides such an excuse is true without qualification.

### 5.2.2 The shadow cast by the benighting act

Return to the case of DOT AND PAT, and now imagine that the medication she gave him—call it *M2*—was, though quite effective in treating his condition, nonetheless inferior to another medication, *M1*, that could have been used instead. Then, given certain background assumptions, Dot's giving Pat *M2* instead of *M1* constitutes wrongdoing on her part. But suppose that she was unaware that this was the case, because she was unaware that *M1* was superior to *M2*. Does she have an excuse for treating Pat as she did?

On the view that only excusable ignorance excuses, the answer to this question will be: Perhaps, but only if Dot has an excuse for her ignorance. Suppose now that her ignorance stemmed from her failure to read a medical journal that she had not only a professional but also a moral obligation to read. Had she read this journal as she ought to have done, she would have learned that *M1* was superior to *M2* in cases such as Pat's and she would then not have committed the unwitting wrongdoing of giving him *M2*. Her decision not to read the journal thus constituted a benighting act, for which she might well be to blame. If she is to blame for it, and if her subsequent ignorance and unwitting wrongful act were foreseeable consequences of it, then, it is plausible to think, Dot is indeed to blame for giving Pat *M2* rather than *M1*.

It's important to note that there are other ways in which Dot's failure to read the journal could have led to her giving Pat *M2* rather than *M1*. Suppose that delivering *M1* to one's patient requires mastering a certain technique, which Dot of course didn't master but could and would have mastered if she had read the journal. Then her giving *M1* to Pat wasn't something that was, at the time at which she treated him, within her intentional control, and perhaps it wasn't even in her simple control. (On the distinction between simple and intentional control, see §4.1.6.) In this case, there is strong pressure to say that Dot's giving Pat *M2* rather than *M1* *wasn't* wrong; since she hadn't mastered the requisite technique, giving him *M2* was the best that she could have done under the circumstances. If so, describing her behavior as an unwitting *wrongful* act is a mistake.<sup>41</sup> Even so, if this behavior was a foreseeable consequence of her benighting act, then it is something for which she would seem to be to blame. It might even have happened that, shortly before her treatment of Pat, Dot had learned of the superiority of *M1* but

<sup>41</sup> Cf. H. M. Smith (1983), p. 552, and (2017), p. 97. I would want to qualify this claim and say that Dot's giving Pat *M2* wasn't *immediately* wrong, though it may well have been *remotely* wrong. This is a subtlety that needn't detain us here. For discussion, see Zimmerman (1996), pp. 96 ff., and (2008), pp. 151 ff.

hadn't had time to master the technique required for its delivery. Then describing her giving Pat *M2* rather than *M1* as *unwitting* is also a mistake. Even so, if her doing so was a foreseeable consequence of her earlier benighting act, then her being blameworthy for this benighting act would seem to render her blameworthy for giving Pat *M2* rather than *M1*.

To hold Dot blameworthy for her treatment of Pat in virtue of her being blameworthy for her decision not to read the medical journal is of course to hold her *indirectly* blameworthy for this treatment. It is a separate question whether she is *directly* to blame for it (or, more particularly, for her decision to treat Pat in this way). There is surely very good reason to hold that Pat is *not* directly blameworthy for this. After all, she didn't believe that it was wrong to treat Pat in this way; on the contrary, she believed that it would have been wrong *not* to treat him in this way. The quality of will that informed her decision was thus above reproach.<sup>42</sup> There is therefore no reason to hold that there was a *fresh* incurrence of blameworthiness by Dot at the time she treated Pat beyond the blameworthiness that she had *already* incurred at the time she decided not to read the journal. Indeed, there is very good reason to deny that this is so, for notice that the case could be developed in such a way that Dot was positively *praiseworthy* for her treatment of Pat. Imagine, for example, that Pat was impoverished and that *M2* was very expensive; Dot knew that Pat couldn't afford the medication and, going well beyond the call of duty, decided to pay for it out of her own pocket.

It's important to note, as Smith herself has recently pointed out,<sup>43</sup> that the denial that Dot is directly to blame for her decision to give Pat *M2* rather than *M1* is *not* grounded in the denial of moral luck. Those who, unlike myself, accept the possibility of moral luck should agree that Dot was not *directly* blameworthy for her decision to treat Pat. What they may claim (and what I deny) is that this decision adds to the degree of blameworthiness incurred by Dot for her *earlier* decision not to read the journal.

Consider now the two interpretations of the thesis that excusable ignorance of wrongdoing provides an excuse for wrongdoing, but inexcusable ignorance may not. If one is not to blame for the ignorance that underlay one's wrongdoing, then one will not be to blame for this wrongdoing, either directly or indirectly (given that the act was not a foreseeable consequence of something else for which one is to blame). Thus excusable ignorance of wrongdoing does indeed provide an excuse for wrongdoing. However, if one is to blame for the ignorance that underlay one's wrongdoing, then, although (first interpretation) one may be *indirectly* to blame for this wrongdoing, nonetheless (second interpretation) one won't be *directly* to blame for it.

<sup>42</sup> Cf. H. M. Smith (1983), p. 559.

<sup>43</sup> H. M. Smith (2017), pp. 108 ff.

### 5.2.3 Inculpatory ignorance

I have just argued that there is a sense in which *all* ignorance exculpates. Yet it is often claimed that ignorance, far from being exculpatory, can *inculpate* an agent. If so, then my argument might seem mistaken.

Consider the case of BILL AND THE BIRD. Bill took a shot at Jill and missed. Clearly he is not to blame for killing Jill, because he didn't kill her. Nonetheless, he may well be to blame for his attempt on her life, and what partly explains his making this attempt is his mistaken belief that nothing would foil it. Some would say that his ignorance was inculpatory rather than exculpatory,<sup>44</sup> *even though* it itself was not something for which he is to blame.

Such cases pose no problem for my argument. I think that it's in fact a mistake to describe Bill's ignorance as inculpatory. *That* is not the basis of his being to blame. On the contrary, on the assumption that he is to blame, what rendered him blameworthy was his awareness of the wrongness of killing Jill, or even attempting to kill her, coupled with his having been in control of his attempt. But even if I'm wrong about this and we should say that Bill's ignorance inculpates him, my position is not threatened. My claim is that it is *ignorance of the fact that one is doing something morally wrong* that provides one with an excuse for wrongdoing. By hypothesis, Bill was *not* ignorant of this fact—his ignorance lay elsewhere—and so I am not committed to saying that he has an excuse for his behavior.

There are other kinds of cases, however, in which I think it comes closer to the truth to say that one's ignorance regarding wrongdoing can indeed be inculpatory rather than exculpatory. In §3.4.1, I argued for the possibility of what I called *accuses*, which consist in one's being to blame, despite not having done anything wrong, because one engaged in behavior that one *believed* to be wrong (and over which one exercised the requisite control). If such cases are indeed possible, then what renders one blameworthy is (in part) one's mistaken belief. Even in such cases, though, I find it misleading to say that it is one's ignorance that renders one blameworthy, since this suggests that it is the *falsity* of one's belief rather than the belief *itself* that is the basis of one's blameworthiness. Of course, if *accuses* are possible, then, as I noted in §3.4.1 when moving from Draft 4 to Draft 5 of the Argument from Ignorance, we shouldn't say, as I just did in the last paragraph, that it is *ignorance of the fact* that one is doing something morally wrong that provides one with an excuse for wrongdoing. Rather, what we should say is that it is *the failure to believe* that one is doing something morally wrong that renders one *blameless for one's behavior*, regardless of whether that behavior is in fact wrong.

<sup>44</sup> Cf. Alexander (1993) and (2017).