3
Ignorance

This book is concerned with the relation between ignorance and moral responsibility. According to the Argument from Ignorance that I have presented, it appears that ignorance undermines such responsibility—in particular, it absolves one of blameworthiness—far more often than is commonly thought. This is because the argument culminates in the Origination Thesis, according to which all blameworthiness is rooted in *non*-ignorant, that is, *witting* wrongdoing, and such wrongdoing seems to occur relatively rarely. (Just *how* rarely it occurs is an issue that I will address in this and ensuing chapters.)

The most recent formulation that I have given of the Origination Thesis is this:

> *The Origination Thesis (Draft 2):*
> Every chain of moral blameworthiness is such that at its origin lies a piece of behavior for which the agent is directly morally blameworthy and which he or she knew, at the time at which he or she engaged in it, to be overall morally wrong.

Here, instead of saying that the agent was *not ignorant* of the wrongness of his (or her) behavior, I have said that he *knew* that his behavior was wrong. This way of framing the Origination Thesis may seem quite natural, but in fact caution is needed here, for the relation between (lack of) ignorance, on the one hand, and (lack of) knowledge, on the other, is perhaps not as straightforward as it first appears.

In this chapter I proceed as follows. In §3.1 I explain why it is that, in our inquiry into moral responsibility for ignorant wrongdoing, we should focus in particular on an agent's failure to believe that he or she is doing wrong.

In §3.2 I undertake an extended investigation into what it is to believe, and what it is to fail to believe, something. I begin with an examination of the relation between knowledge and belief, in the course of which I draw a distinction between belief and acceptance. Typically, acceptance is an attitude that one adopts intentionally, whereas belief is not. Nonetheless, it's clear that there is often a close association between belief and intentional behavior. Some philosophers claim that having a belief consists in having a disposition to engage in certain kinds of behavior, but it is notoriously difficult to specify precisely the kinds of behavior at issue. I briefly discuss some proposals. Regardless of the merit of these proposals, almost everyone acknowledges another connection between beliefs and dispositions, and that is that beliefs can themselves be merely dispositional, as

Ignorance and Moral Responsibility. Michael J. Zimmerman, Oxford University Press. © Michael J. Zimmerman 2022.
DOI: 10.1093/oso/9780192859570.003.0003

opposed to being occurrently or consciously held. I investigate what this distinction consists in, paying particular attention to the role played by attention, which comes in degrees, as a result of which the type of consciousness or awareness involved in occurrent belief does, too. It is sometimes claimed that beliefs also come in degrees. This has to do with the degree of confidence that one has in a proposition. In this context, philosophers often talk of credences rather than beliefs. I discuss the nature of credences and examine the relation between them and "outright" beliefs. I then turn to the question of what it is to have beliefs that conflict with one another and with other belief-like attitudes. Finally, I discuss two ways in which one may fail to believe a proposition.

In §3.3 I address the distinction between acting from ignorance and acting merely in ignorance. Many philosophers claim that ignorance provides an excuse for wrongdoing only if one acts from, rather than merely in, ignorance of the fact that one is doing wrong. This is a claim that I think it would be wise to endorse.

Finally, in §3.4 I discuss two possible ways in which there might be blameworthiness without wrongdoing. The first concerns what I call accuses (as opposed to excuses). If one's quality of will can render one blameless even though one has done wrong, the question naturally arises whether one's quality of will can render one blameworthy even though one has done no wrong. I argue for an affirmative answer to this question. The second possibility is that one is to blame, not for doing something wrong, but rather for doing something suberogatory. (The category of suberogation is alleged by some philosophers to be the negative counterpart to that of supererogation.) I am inclined to deny the possibility of suberogation but, if even I am right to do so, that doesn't mean we can ignore it; for someone might believe that she is doing something suberogatory, in which case her quality of will would seem objectionable.

3.1 Knowledge and ignorance

It is often said that ignorance consists in lack of knowledge. Moreover, according to tradition (or legend[1]), knowledge consists in justified true belief. (The kind of knowledge at issue here is propositional knowledge, that is, knowledge *that* something is the case, rather than some other kind of knowledge, such as knowing how to do something or knowing someone by acquaintance, unless these are somehow reducible to propositional knowledge.) This traditional analysis is of course controversial. Each of justification, truth, and belief has on occasion been held *not* to be necessary for propositional knowledge,[2] and Edmund Gettier

[1] See Dutant (2015). [2] For discussion, see Ichikawa and Steup (2017), §1.

famously demonstrated that the justification condition requires refinement if it is to be sufficient (*ceteris paribus*) for such knowledge.[3] Fortunately, we need not concern ourselves with most of these issues here. That is because, contrary to what is often said (and to what I have myself said in the past[4]), it is in fact a mistake to claim that ignorance consists in lack of knowledge. For, whereas it is certainly the case that ignorance of some proposition precludes knowledge of that proposition, lack of knowledge does not suffice for ignorance, for several reasons.

First, objects such as tables and chairs lack knowledge of any and all propositions, yet they are not ignorant of these propositions. To be ignorant of a proposition one must be capable of possessing propositional knowledge. (I won't venture to say what such a capacity itself consists in. Note, though, that I am not claiming that, to be ignorant of some proposition *p*, one must be capable of knowing *p* itself. That would clearly be far too restrictive a condition.)

Second, as Rik Peels has observed,[5] ignorance is ignorance of facts. Facts are (or may be taken in the present context to be) true propositions, and one cannot be ignorant of a proposition that is false. Suppose that Florence believes that the earth is flat. She obviously doesn't know that it is, because it isn't. But also, because it isn't, she's not ignorant of this "fact," either.

Third, as Peels has also observed,[6] it is doubtful that one is ignorant of a proposition that one doesn't know, if one's lack of knowledge is due, not to the proposition's being false or to one's failing to believe it, but rather to one's belief's lacking the requisite justification. Suppose that Harry believes, truly, that the house in which he lives was built in 1928, but suppose that he does so because he believes, falsely, that all the houses in his neighborhood were built then. Then he doesn't know that his house was built in 1928, but he doesn't seem to be ignorant of this fact, either.[7]

Let us say, then, that one is ignorant of a proposition just in case one is capable of possessing propositional knowledge and the proposition in question is true, but one fails to believe it. On the safe assumption that moral agents, agents capable of doing wrong and of being to blame, are also capable of propositional knowledge, it is therefore with such agents' *failure to believe* of some true proposition that it is true that we are at bottom concerned, and it is worth reformulating the Argument from Ignorance and the Origination Thesis to reflect this fact. Thus:

[3] Gettier (1963). [4] In various places, including Zimmerman (1988), p. 75.
[5] Peels (2017), p. 166. [6] Peels (2017), p. 167.
[7] Some people to whom I've proposed this example are inclined to say that Harry *is* ignorant of the fact in question. There's no need for present purposes to take a definitive stand on this issue. For, as Peels again has observed (2017, pp. 166 f.), "true belief does *not* provide an excuse, not even a partial one." He gives the example of someone who sets fire to a barn, believing truly, but without sufficient warrant for knowledge, that there is someone still inside. Cf. Rosen (2008), p. 596.

The Argument from Ignorance (Draft 3):
Suppose that

(1) (a) *P* committed *A* and (b) *A* was overall morally wrong, but (c) when *P* committed *A, P* **failed to believe** that *A* was overall morally wrong.

In general it's true that

(2) if one committed some act or omission that was overall morally wrong while **failing 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**.

Thus

(3) *P* is morally to blame for *A*, and thereby morally to blame for any of its consequences, only if *P* is morally to blame for *P*'s **failure to believe** (call it ***F***) that *A* was overall morally wrong.

It's also in general true that

(4) one is morally to blame for something only if one was in control of that thing.

Thus

(5) *P* is morally to blame for ***F*** only if *P* was in control of ***F***.

Three further general truths are that

(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.

Thus

(9) *P* is morally to blame for *A*, and thereby to blame for any of its consequences, only if there was something else (call it *X*) for which *P* is directly morally to blame and of which ***F*** was a consequence.

But

(10) whatever *X* was, it cannot have been an act or omission such that (a) *P* committed it and (b) it was overall morally wrong and (c) when *P* committed it, *P* **failed to believe** that it was overall morally wrong, since otherwise the foregoing argument regarding *A* would apply all over again to *X*.

Thus

(11) whatever *X* was, it was either (a) not some act or omission that *P* committed or (b) not overall morally wrong or (c) something such that, when *P* committed it, *P* **believed** that it was overall morally wrong.

Two further general truths are that

(12) one has direct control over something only if that thing is an act or omission, and

(13) one is morally to blame for an act or omission only if that act or omission was overall morally wrong.

Thus

(14) whatever *X* was, it was an act or omission such that, when *P* committed it, *P* **believed** that it was overall morally wrong.

The Origination Thesis (Draft 3):
Every chain of moral blameworthiness is such that at its origin lies a piece of behavior for which the agent is directly blameworthy and which he or she **believed**, at the time at which he or she engaged in it, to be overall morally wrong.

For convenience, I will continue to characterize the Origination Thesis as the thesis that all blameworthiness is rooted in *witting* wrongdoing, but bear in mind from this point on that I do *not* mean by this that the agent *knows* that the behavior in question is wrong, since, in keeping with the foregoing remarks, I am not presupposing that the agent's belief that the behavior is wrong must meet the standard of epistemic justification required for knowledge.

There is a distinction between knowing *that* some proposition is true and knowing *whether* that proposition is true. The latter consists in either knowing that the proposition in question is true or knowing that it is false. So, too, there is a distinction between being ignorant *of the truth* of some proposition and being ignorant *of whether* that proposition is true. The latter consists in either being ignorant of the fact that the proposition in question is true or being ignorant of the fact that it is false. Notice that, even though it is impossible to know with respect to some proposition that one is ignorant of its truth, it is perfectly possible to know with respect to some proposition that one is ignorant of whether it is true. If some proposition is such that one neither believes nor disbelieves it, and one knows this, and one also knows that the proposition in question is either true or false, then, if one is rational, one will know that either the proposition is true but one fails to believe it or it is false but one fails to disbelieve it. This is all one needs to know in order to know that one is ignorant of whether the proposition is true. I will return to this point in §7.2.1.

3.2 Believing and failing to believe

Neither believing nor failing to believe is a simple matter. A proper assessment of the Argument from Ignorance and the Origination Thesis requires that we attend to each of these matters.

3.2.1 Knowledge and belief

I remarked above that it is sometimes held that (propositional) knowledge does not require belief. It is certainly true that people sometimes say that they don't believe something, even when it's clear that they know what they claim not to believe. Ichikawa and Steup consider the example of Walter, who comes home after work to find that his house has burned down. Walter says, "I don't believe it." But the "it" in question, namely, the fact that his house has burned down, is something that Walter knows, indeed knows all too well, to be true.[8]

This is not a convincing counterexample to the claim that knowledge requires belief. Walter has misspoken. It's obvious that he doesn't believe that his house is still standing, since otherwise he would not have reacted as he did. That he doesn't believe that his house is still standing does not of course entail that he does believe that it's not still standing, but, again, unless he had the latter belief, he would not have reacted as he did. Rather than his lacking this belief, what his remark indicates, as Ichikawa and Steup themselves note, is simply that he finds it hard to come to terms with the fact that his house has burned down. Indeed, in this regard, what he says is quite mild. Someone else in his position might well have said, "I *can't* believe it," thereby indicating that he finds it *very* hard, perhaps impossible, to come to terms with what has happened.

Suppose that Walter in fact cannot reconcile himself to what has happened. He is distraught, overwhelmed, unable to "move on." His friend, Nora, might say, "Poor fellow, he just can't accept the fact that his house has burned down." Unlike Walter, Nora has *not* misspoken. Some philosophers hold that belief and acceptance are one and the same,[9] but, as several others have recently emphasized,[10] they are not. (Or, to put matters less forcefully: there are common and important senses of the terms "believe" and "accept" that are distinct from one another. So, too, for the terms "disbelieve" and "reject.") As I am using the term, to accept a proposition is to take it as given, in order to achieve some desired outcome. It is an attitude that one typically adopts intentionally. One might accept a proposition for the sake of argument, for example, or in order to make plans for the future. In

[8] Ichikawa and Steup (2017), §1.2.

[9] See, e.g., Chisholm (1976), p. 27.

[10] See especially Cohen (1989) and Bratman (1992). See also B. Williams (1973), p. 140; Audi (2008), p. 413; Schwitzgebel (2015), §2.5.

contrast, in typical cases one cannot intentionally believe a proposition (although one can intentionally engage in some activity—undertake an inquiry, say—with the purpose of effecting a change in one's beliefs; I will discuss this issue more fully in §4.6). Note that, given this difference between acceptance and belief, it often happens that one accepts a proposition *because* one believes it. However, it can also happen that one fails to accept some proposition that one believes and fails to believe some proposition that one accepts. The case of Walter serves to illustrate the first of these possibilities. A case in which, whether compulsively or out of an abundance of caution, one double- or triple-checks the lock on a door that one knows to be secure, serves as another illustration of this possibility. An argument that proceeds by way of *reductio ad absurdum* serves as an illustration of the second possibility, as does a case in which I make plans on the basis of the assumption that you will be punctual, even though, given your habitual tardiness, I have no expectation that you will in fact turn up on time.

Colin Radford has offered another, much-discussed counterexample to the claim that knowledge requires belief.[11] Jean, a French-Canadian, professes ignorance of English history. Tom decides to test him on this and asks him a series of questions about the dates of certain significant events—when William the Conqueror landed in England, when Queen Elizabeth I died, and so on. To Jean's surprise, he answers several of these questions correctly, albeit very hesitantly, and he infers that he must at some time, many years before, have learned the relevant information, although he has no firm recollection of ever having done so. Radford claims that, when Jean gave the correct answers, he knew the relevant information but didn't believe it.

If we assume, as Radford clearly intends, that Jean's answers were not lucky guesses and do indeed indicate that he once learned the relevant information, we should surely agree that Jean did at some time know this information. Whether he knew it at the time that Tom quizzed him is perhaps less clear, although I suspect that many would say, along with Radford, that he did. But why say that, at that time, Jean knew it but didn't believe it? Jean's lack of confidence in his own answers indicates that he didn't believe *that he knew* the information in question, but that's another matter.[12] Radford himself allows for, indeed insists on, Jean's knowing this information nonetheless. I submit that there is just as much reason to insist on Jean's believing this information.

What it is perhaps most important to note, though, is that there is no need, for present purposes, to reach a definitive conclusion about whether knowledge requires belief. That is because, for the reasons given above, our present concern is not with whether one does something in the *knowledge* that it is wrong—that is, overall morally wrong—to do it, but simply with whether one does it in the *belief*

[11] Radford (1966). [12] Cf. Cohen (1966), p. 11.

that it is wrong to do it. It was in order to emphasize this very point that I moved to the most recent formulations of the Argument from Ignorance and the Origination Thesis. If, contrary to what I believe, it were possible to do something knowing it to be wrong but *not* believing it to be so, this would not affect my central question, which is whether the *failure to believe* that what one is doing is wrong affords one an excuse for one's conduct. (It's true that, under such circumstances, the label "Argument from *Ignorance*" would be misleading, but, since I reject the possibility of such circumstances, I will stick with that label.)

3.2.2 Beliefs and dispositions

The question still remains, of course, what it is to believe something. This is a complicated and controversial question, on which there is an enormous literature. In the present context, I can only offer a few observations.

Philosophers have long noted a close association between belief and behavior. If you believe that it will rain later, you may well take an umbrella with you as you leave your house, whereas, if you believe that the sun will shine all day, you probably won't take an umbrella with you. This is a mundane, uncontroversial, but nonetheless helpful observation. How to improve on it in order to tighten one's grasp on the concept of belief, however, is highly controversial. The terms "may well" and "probably" that I just used are crucial, since clearly you might leave your umbrella behind even though you expect it to rain, and you might bring it with you even though you expect it not to rain. There is obviously no one-to-one correlation between belief and behavior; the link between the two is far more complex than that. Several philosophers have proposed that beliefs are *dispositions* to behave in certain ways, dispositions that may on occasion fail to be manifested, even under "favorable" conditions, due to the presence of some countervailing conditions. (Thus, even if you expect it to rain, you may leave your umbrella behind, if you also expect the wind to be so strong as to render the umbrella unusable, or if your umbrella is broken, or if you don't mind getting wet.) But, in light of such exceptions to the manifestation of the dispositions in question, it is extraordinarily difficult to identify precisely what the dispositions are (supposed to be) dispositions *to*. Perhaps the best that can be said in general along these lines is that one's believing some proposition *p* consists in being disposed to behave *as if* *p* were the case,[13] but just how this proposal is to be understood is far from clear. Suppose that today, having taken a look at the sky and noticed the gathering clouds, you took your umbrella with you when you left your house. Does this count as your acting "as if" it might rain? It may seem plausible to say so. But

[13] Cf. Marcus (1990).

suppose that your neighbor, Ned, likes to get wet. When today he took a look at that same sky and noticed those same clouds, he left his umbrella behind. Does this also count as his acting "as if" it might rain? It may seem plausible to say this, too. But don't you and Ned have the *same* belief, even though you are disposed to behave, and do behave, very *differently*? Unless the concept of behaving as if *p* were the case is clarified, the present proposal is difficult to assess.[14] Moreover, it seems plausible to think that, however the concept of behaving as if *p* were the case might be analyzed, it would remain possible that one have a disposition to behave in this way, not in virtue of believing *p*, but in virtue of something else—in virtue, say, of wanting to please one's friend, who very much likes it when one behaves as if *p* were the case.

In light of considerations of the sort just mentioned, many philosophers deny the "dispositionalist" view that beliefs are to be identified with dispositions to behave in certain ways. Several alternative accounts, which I will not discuss here, have been offered.[15] Despite being opposed to dispositionalism, almost all these accounts nonetheless endorse the idea that many, indeed most, of our beliefs are (merely) dispositional, although just what *this* idea comes to is, once again, controversial. The basic point, though, one that is highly relevant to my project in this book, is that, to believe that *p* is the case, one need *not* be currently considering (entertaining, adverting to, thinking about) *p*. For example, I expect that you believe that Paris is the capital of France, and that you believed this five minutes ago, too. Because I have just brought this fact to your attention, you may well be currently entertaining the proposition that Paris is the capital of France, but I bet you weren't thinking about it five minutes ago. Contrast this case with one that I raised in Chapter 1. Prior to consulting Wikipedia, I certainly did not believe that Astana was the capital of Kazakhstan (nor, of course, did I believe that it wasn't). If, prior to reading Chapter 1, you had this belief, then you were better informed than I was.

One account of the nature of belief, both occurrent and dispositional, that I find promising has recently been proposed by Rik Peels.[16] Peels notes that often, when we consider (entertain, advert to, think about) some proposition *p*, we think that *p* is indeed the case. (Note the difference between thinking *about* and thinking *that*; the former, though necessary, is not sufficient for the latter.) For example, when I think about Paris being the capital of France, I think *that* it is. This is *occurrent* belief. But, as just noted, not all belief is occurrent. Five minutes ago I believed that Paris was the capital of France, even though I wasn't considering this proposition. Peels distinguishes between two types of non-occurrent belief (neither of which he calls dispositional, although both of them involve having a disposition). Roughly, Peels holds that one *dormantly* believes *p* just in case one does not now

[14] Cf. Schwitzgebel (2015), §1.2.
[15] A useful overview may be found in Schwitzgebel (2015).
[16] Peels (2017), pp. 28 ff.

occurrently believe *p* but did occurrently believe it the last time one considered it and would now occurrently believe it if one were to consider it. (Let me stress that this is indeed rough, since, as Peels takes pains to acknowledge, there can be exceptional cases in which the counterfactual is false but one nonetheless has the relevant dormant belief. One might be distracted, for example, or under great stress, in which case, despite believing *p*, one might *not* think that *p* was the case even if one were to consider it.) Peels claims further (and again I am putting his proposal roughly) that one *tacitly* believes *p* just in case one neither occurrently nor dormantly believes it but it obviously follows, from one's own perspective, from propositions that one does occurrently or dormantly believe and one would now occurrently believe it if one were to consider it; for example (an example I borrow from Ingmar Persson[17]), *p* might be the proposition that birds do not wear boots.

There may be questions to be raised about the details of Peels's account, questions that there is no need to discuss here,[18] but one question that should be addressed is this: In allowing for the possibility that a dispositional belief be not dormant but tacit, is the account too expansive? Your believing five minutes ago that Paris was the capital of France will presumably qualify as a dormant belief, precisely because this is a proposition that you once considered and continue to endorse. But Peels's account attributes to you, and me, and everyone else beliefs in propositions that we have never considered. (His is not the only account to do so.[19]) There will presumably be *indefinitely*, quite possibly *infinitely*, many such propositions. Do we really have so many beliefs? Peels gives the following argument for the claim that indeed we do:

> People believe that they are less than two miles tall, that they are not bats, and that 2 is a smaller number than 999, even if they have never thought about these propositions. If you are not sure whether people believe these things, ask whether they *know* these things. There, it seems, the answer is surely positive. My neighbor knows that he is not a bat, even if he has never considered that proposition. But knowledge . . . entails belief. Hence, people *believe* these things, even if they have never entertained the propositions in question . . .[20]

One might reject this argument on the basis of denying the premise that knowledge entails belief. As I have said, though, I accept this premise. Nonetheless, I am inclined to resist Peels's argument. I am inclined to think that, in general, people do *not* know that they are less than two miles tall, that they are not bats, etc. It may

[17] Persson (2019), p. 26.

[18] Questions such as whether a dormant belief that *p* is the case can be formed only by way of considering *p* (regarding which see n. 23 below), what the criterion is for one proposition's "obviously following" from another, and the like.

[19] Again, see Persson (2019), pp. 26 f. Cf. Schwitzgebel (2015), §2.2, on "implicit" belief.

[20] Peels (2017), pp. 32 f.

well be, of course, that, were you to ask them whether they were less than two miles tall, they would immediately assent to this proposition (unless they suspected some sort of trick), but that, it seems to me, only shows that they would immediately form the belief upon being asked, rather than that they already have this belief—they have a *disposition to believe* the proposition but do not *dispositionally believe* it.

The distinction between a dispositional belief and a (mere) disposition to believe has been examined in detail by Robert Audi.[21] He claims that the distinction consists, at least in part, in the distinction between the accessibility of a proposition by a retrieval process that draws on memory and its accessibility only through a belief-formation process. When you assent to the proposition that Paris is the capital of France, you are accessing information that you already possess, information that is stored in your memory. When you assent to the proposition that 2 is a smaller number than 9999 (I have changed Peels's example slightly in order to fix on a proposition that, I assume, you had not already entertained and affirmed), you are not drawing on your memory. Or, to put the point more precisely, you are not drawing on your memory of the fact that 2 is a smaller number than 9999; you are, instead, drawing on your memory of *other* facts from which you immediately infer this proposition.

But, while I am inclined to deny that there can be tacit beliefs (as defined by Peels), I recognize that the case is not closed. Peels himself acknowledges the distinction between a disposition to believe and a dispositional belief.[22] He doesn't claim that every disposition to believe something qualifies as a tacit belief; he holds that one tacitly believes a proposition only if it obviously follows, from one's own perspective, from other propositions that one believes. This is what allows him to say that most people know, and hence believe, that they are not bats, even if they don't believe that they are not mammals of the order Chiroptera, and even though they would presumably assent to the latter proposition once they learned that that is how bats are classified zoologically. Moreover, I think it must be admitted that it does seem quite natural to say that most people do indeed know that they are not bats (and that birds don't wear boots, and that 2 is a smaller number than both 999 and 9999, etc.). To the objection, perhaps implicit in Audi's account of belief, that one cannot have a belief that one has not formed, and there is no occasion on which one forms the tacit beliefs that Peels identifies, Peels could, I think, respond as follows. First, it is perhaps somewhat misleading to say that people form beliefs at all, since this might be taken to suggest some voluntary activity on their part, and no one voluntarily forms beliefs, at least not directly. (See Premise 6 of the Argument from Ignorance. This premise is admittedly controversial; I will discuss it further in §4.6. I do *not* mean to suggest that Audi himself holds that beliefs are

[21] Audi (1994). [22] Peels (2017), p. 36.

formed voluntarily.) It would be better to say that people acquire beliefs, or come to have them, or that beliefs form (or form themselves) in people. Second, Audi appears to presuppose that, in many cases at least,[23] one can come to have and retain a belief that *p* is the case only if one considers *p* and then stores it in one's memory. This presupposition immediately rules out the possibility of tacit beliefs in such cases, of course, but it is open to question. As Peels understands tacit beliefs, they constitute a proper subset of dispositions to believe. Such dispositions must presumably be acquired and retained in some way, but of course they need not be acquired by way of considering some proposition or retained by storing some proposition in one's memory. This observation also helps to dispel the worry that possession of a very large or infinite number of beliefs requires some vast storage capacity, inasmuch as, even if there are strict limits on what and how much we can remember, there seems to be no corresponding limit on what dispositions we can have. There seems to be no doubt, for example, that people in general have a disposition to agree that 2 is a smaller number than 999, a disposition (whether another disposition or the same disposition is a matter of debate) to agree that 2 is a smaller number than 1000, a disposition to agree that 2 is a smaller number than 1001, etc. ad infinitum.

In sum, it's clear that we have not only occurrent but also dormant beliefs. Whether we also have tacit beliefs is not so clear. I am inclined to think that we do not, but I acknowledge that there is reason to think otherwise.[24]

3.2.3 Occurrent beliefs

Dormant beliefs and, if there are any, tacit beliefs involve the disposition to have an occurrent belief. I have endorsed Peels's claim that occurrently believing some proposition *p* involves, in turn, considering (or entertaining, or adverting to, or thinking about) *p*. But what does *this* involve? This is yet another complicated and controversial question, on which I can again offer only a few observations. But offer them I must, since just how subversive of our everyday ascriptions of blameworthiness the Origination Thesis is may seem to depend in large part on whether the relevant beliefs must be occurrent—an issue that I will address in §6.2.1.

[23] Audi explicitly notes in his (1994), pp. 420 f., that it seems to be possible to acquire some beliefs (e.g., the belief that an ambulance passed by) through perception rather than consideration.

[24] There is a related, but different debate in the philosophy of mind about whether we have knowledge of the rules of grammar that we follow when we engage in linguistic behavior. Some theorists hold that we do have knowledge of these rules but that our knowledge is "tacit," in that what we know is inaccessible to us and, not being inferentially integrated with any of our beliefs, is itself not something that we believe but rather something toward which we have only a "subdoxastic" attitude. For discussion, see Macdonald (1995), Davies (1995), and Searle (1995). I will forgo any comment on this debate, except to note in passing that I have observed, on more than one occasion, small children misusing irregular verbs, treating them as if they were regular (saying, for example, "I hitted the ball," "He holded the bat," and so on). Such misuse seems to me to be evidence of a mistaken inference.

Consider again your belief that Paris is the capital of France. Five minutes ago (or, by now, ten or fifteen minutes ago), before I brought this belief to your attention, it was dormant. Now, I assume, it is occurrent. Earlier, you had the relevant information, but you were not thinking about it; now you are. It was, if I may put it this way, in your mind but not on your mind. What precisely is the difference?

Instead of saying that your belief that Paris is the capital of France is now *occurrent*, I might have said that you are now *conscious* or *aware* of this fact. This suggests an equivalence that some would deny, however. In a probing discussion of the difference in law between recklessness and negligence, Douglas Husak has the following to say:

> [*C*]*onsciousness* or *awareness* of a substantial and unjustifiable risk is the single factor that distinguishes recklessness from negligence. If the defendant *is* aware of the risk, he is reckless. If he is *not* aware, he is negligent as long as he *should* have been aware…
>
> When *is* a person aware of something? In particular, when is a person aware of a risk?… [T]he difference between the reckless and negligent defendant is simply a function of their beliefs: the reckless defendant, unlike the negligent defendant, *believes* that he is creating a substantial and unjustifiable risk…
>
> [However,] awareness (or belief) need not be *occurrent*… [I]t is clear that defendants need not have the explicit thought about the risk before their conscious minds in order to be reckless. In other words, reckless persons need not be saying to themselves "this is risky" when they act. Few of us would ever be reckless if we needed to rehearse such thoughts… Indeed, the number of beliefs we hold at any particular time would be very few if this test were adopted. For example, I would not believe the true proposition "My name is Douglas Husak" when concentrating on my writing. Phenomenologically, we rarely articulate propositions to ourselves in order to qualify as believing them. And for good reason. Persons who are keenly aware of a risk and seek to minimize it quickly learn not to be distracted from their task by mentally entertaining such propositions. For example, the experienced baseball player who bats with two strikes does not think to himself "I might miss the next pitch and strike out." Such thoughts would sap his focus from the job at hand: hitting the baseball if it is thrown over the plate. Yet I assume that no one would infer that this batter is unaware of and does not believe he is at risk of striking out… *Of course* he is aware of the risk…[25]

[25] Husak (2011), pp. 207 ff.

This rich passage raises a couple of issues that warrant explicit consideration.

First, Husak writes as if consciousness and awareness are one and the same.[26] But this is not the case. Awareness is veridical, whereas consciousness need not be. Crawling across the hot desert sand, you are conscious of water on the horizon. You are not aware of it, however, because what you see is a mirage. Also, awareness is an intentional attitude, but it is controversial whether consciousness must be intentional. (An attitude is intentional, in the present, technical sense, just in case it is directed toward some object.) To be aware is to be aware *of* something. Often, of course, when one is conscious, one is conscious of something. The question is whether it is possible to be conscious *without* being conscious of something. If, as many claim, to be conscious is to be the subject of *experiences*, that is, to be in a state that has a certain *phenomenal* character, then it seems plausible to say that consciousness need not be intentional; for it seems plausible to say that not all experiences need be experiences that are directed toward some object. (For example, you may be depressed but not depressed *at* anything.) However, if, as many others propose, consciousness is to be understood in terms of having *access to information*, then it must be intentional; the information in question will be the object toward which it is directed. These two conceptions of consciousness are not mutually exclusive, but they are distinct.[27]

My concern in this book is with whether one believes, at the time one does something, that one is doing wrong. This is a matter of having access to information. For my purposes, then, there will be no need to distinguish between consciousness and awareness on this score, and so I am happy to go along with Husak in this respect. Belief need not be veridical, of course, but insofar as my concern (like Husak's) is with ignorance of the *fact* that one is doing wrong, there will be no need to distinguish between consciousness and awareness on this score, either (for the time being, at least; I will return to this question in §3.4 below).

The second issue, regarding which I do differ with Husak, is this. He stresses what he takes to be an important distinction between consciousness or awareness, on the one hand, and occurrent belief, on the other. He appears to think of entertaining a proposition (something that is necessary for occurrently believing it) as a matter of focusing one's attention on it, of having it "at the forefront of one's mind."[28] As he observes, he does not have his own name at the forefront of his mind when he concentrates on his writing (or, at least, not usually; presumably he did do so when he gave his example). And, as he also observes, the batter knows that he had better not have the fact that he is at risk of striking out at the forefront of his mind if he is to succeed in not striking out. Nonetheless, Husak claims, the batter is clearly aware of this risk, and, I assume, Husak would say that he is clearly

[26] He is not alone in this. Cf. Sher (2009), p. 127; Levy (2014), p. 29.
[27] For discussion, see Siewert (2017) and Van Gulick (2017).
[28] See Husak (2016), pp. 195 f. Cf. Schwitzgebel (2015), §2.1.

aware of his own name. Another context in which, it might be claimed, one can be aware of something without attending to it is one in which one is listening to a complex piece of music.[29] Imagine that you are listening with your eyes closed to a recording of Beethoven's seventh symphony. The whole orchestra is engaged: strings, winds, etc. As you listen, you shift your attention from the strings to the winds and back again; then you listen even more closely to the strings, focusing on the cellos while bracketing the violins; and so on. It may seem plausible to say that, when you are attending to one set of instruments, you are not attending to the others, and yet all the while you are conscious or aware of them all.

I am not persuaded by such examples that one can be conscious or aware of something without attending to it.[30] Consider first what Husak says. In saying "…no one would infer that this batter is unaware of and does not believe he is at risk of striking out," he appears to be identifying awareness with (true) belief. But, given that beliefs can be dormant, this seems a mistake. No doubt Husak knows his own name even when he's not thinking about it, but is he really *aware* of it when he concentrates on his writing? Were you *aware* five (or ten, or fifteen, or twenty) minutes ago that Paris was the capital of France? It seems odd to say so. (Of course, had I asked you this question then, or had I interrupted Husak while he was writing and posed the relevant question to him, each of you would presumably have answered, correctly, that you were indeed aware of the matter at hand. But that answer would have been correct precisely because I had brought the matter to your attention.) It seems odder still to say that each of you was *conscious* of the relevant information. But then what about Husak's point that, if the batter had concentrated on not striking out, his doing so would have been self-defeating? We can grant this, while avoiding the view that awareness is simply a matter of (true) belief, if we make what I take to be a common and plausible observation: attention comes in *degrees*, and hence consciousness or awareness does, too. It seems plausible to say that, when he concentrates on his writing, Husak (usually) pays no attention at all to his own name. Likewise, before I raised the issue, you were paying no attention at all to the fact that Paris was the capital

[29] Cf. Watzl (2011), pp. 146 ff.

[30] For further discussion, see Jennings (2015). None of the (alleged) examples of consciousness without attention discussed in this article seem applicable to the particular case of being conscious of wrongdoing.

There is the converse question of whether someone can attend to something without being conscious of it. A dramatic, often-cited example of this alleged possibility is that of blindsight, which is the ability to detect and respond to visual stimuli despite the fact that, due to damage to one's primary visual cortex, one is not aware that one is seeing anything. There are cases in which people do indeed display this ability. (See Weiskrantz (1986) for discussion.) We can surely agree that in these cases the subjects are not aware of the stimuli. Whether they nonetheless attend to these stimuli is controversial, however; some (e.g., Wu (2014), pp. 110 ff.) contend that they do, while others (e.g., Prinz (2011)) claim that they don't. Unless there is some analogous phenomenon when it comes to attending to *non*-visual information, however (and I know of no suggestion that there is), I think we may safely ignore this possible instance of attention without consciousness.

of France. The case of the batter seems different, however. Although he is concentrating on the job at hand, he is nonetheless also conscious of the fact that he might strike out. (Consider Husak's own apparent admission that someone may be *more or less* keenly aware of a risk.) This observation also helps deal with the music example. Contrary to the claim that you are conscious of the entire orchestra and yet are attending to only one section of it, I suggest instead that no section is receiving your *full* attention precisely because you are conscious of, and thus *are* attending to, all the other sections, too, if only partially.

It is difficult to know how to flesh out this common observation about degrees of attention. One way to try to do so is to make use of certain metaphors. I have already employed a couple. I have talked of having certain information at the "*forefront*" of one's mind. This suggests, aptly I think, that certain other information may be on one's mind but lie somewhat further "back." When information is "on," as opposed to merely "in," one's mind, it is not just information to which one has access, it is information that one has accessed. It is something about which one is currently thinking, *even if* it is not "right up front." Another metaphor, employed by Husak himself and equally apt, is that of *focusing*. The batter focuses on the job at hand, but, as in any visual field, whether literal or metaphorical, there is not just a focal point but also a periphery, with some items lying closer to the former and others, of which one is perhaps only "dimly" aware, closer to the latter. Another, very closely related metaphor is that of a *center* of attention, a metaphor that might be embellished by saying that this center is surrounded by concentric circles.[31] (Husak himself talks of "concentration.") The "area" of attention is contained within the circumference of the outermost circle.[32]

[31] Cf. Watzl (2011).

[32] It is perhaps worth mentioning support for the thesis that awareness requires attention that is provided by the phenomenon of inattentional blindness. A dramatic, often-cited experiment that suggests that inattentive consciousness is *not* possible involved subjects watching a video in which six people, three wearing black shirts and three wearing white, moved from one spot to another, passing a pair of basketballs among themselves. The subjects were asked to count the number of passes made by those wearing white shirts. Doing so required careful attention. As the balls were tossed back and forth, a gorilla (or, rather, a person in a gorilla suit) entered the scene stage left, paused at center stage, turned to face the camera, thumped its chest, and then exited stage right. After having watched the video, half of the subjects, when asked whether they had seen the gorilla, replied that they had not. (See www.theinvisiblegorilla.com/gorilla_experiment.html. For discussion, see Simons and Chabris (1999).) Of course, no such single experiment can establish the truth of the *general* thesis that awareness requires attention. We can surely agree that these subjects did not attend to the gorilla. Whether they were aware of it, despite being unable to report on it afterwards, is controversial, however. Some contend that they may have been (e.g., Smithies (2011), pp. 254 ff.), others that they were not (e.g., Prinz (2011)). But even if they weren't, that still leaves the possibility that other people, in the same or another setting, might display awareness without attention. Nonetheless, the general thesis certainly seems plausible. It seems reasonable to say that those who both noticed the gorilla and managed to count the number of passes correctly succeeded in attending to both matters—not equally, perhaps, but that's precisely my point. (Whether we should say that in such a case one's attention is divided or multiplied is an interesting question.) Likewise, it seems reasonable to say that the batter in Husak's example attended primarily to doing the job at hand while also taking note of the fact that he might strike out.

Another complication here has to do with *self*-awareness. One can be aware of something without being aware that one is aware of it. (The term "conscious state" is often used in this context, but it is treacherous, I think, because ambiguous. There is a distinction to be drawn between a state *in* which one is conscious and a state *of* which one is conscious.[33]) The batter, for example, may not be aware that he is aware of the risk of striking out, while being quite aware that he is focusing, or attempting to focus, on hitting the ball if it is thrown over the plate. (Perhaps focusing "purely" on the job at hand—presumably the optimal state under the circumstances—would preclude such self-awareness, however.)

There is, it should be acknowledged, another way (consistent with the thesis that awareness requires some degree of attention, but also consistent with the denial of this thesis) to construe what happens in Husak's baseball scenario, and that is to say that the batter attends *first* to the risk of striking out and *then* to doing the job at hand. If the degree of focus needed to accomplish the job precludes devoting any degree of attention at that *same* time to the risk (or to anything else, including the fact that he is focusing on the job at hand), it would of course remain true that, at *some* time during his session at bat, the batter attended to the risk. Perhaps this construal of the scenario better accommodates Husak's allusion to the batter's being *keenly* aware of the risk of striking out, since a normal person cannot maintain keen awareness of one matter while focusing on another. In any event, notice how odd it would be if, when commenting later on his turn at bat, the player were to say, "You know, the thought of striking out never occurred to me." If that had indeed been the case, I think we should say that, far from being aware that he might strike out, the batter was oblivious to this possibility.

I am suggesting, then, that occurrently believing a proposition p consists in being conscious of or, in the case of true propositions, aware of the truth of p. This requires attending to p to some degree, but the degree may be minimal; for that is all that considering (entertaining, adverting to, thinking about) p requires.

3.2.4 Degrees of belief

An occurrent belief that p is the case involves attending to p. I have suggested that attention comes in degrees. It is often claimed that belief *also* comes in degrees. This is an entirely different matter, one that cuts across the issue that I have just been discussing. One can hold a dormant belief (such as the belief that Paris is the capital of France) with great confidence while holding an occurrent belief (such as the belief that one has just spotted an oasis in the desert) with considerable diffidence. In the latter case, one might even disavow having a belief altogether, saying only that one *suspects* that p is the case.

[33] Cf. Finkelstein (1999), p. 81.

Instead of degrees of belief or confidence, philosophers often talk of *credences*.[34] One common way of trying to regiment the discussion is to treat credences as falling within a numerical range from 0 to 1. A credence of 1 in *p* represents complete confidence, absolute certainty, that *p* is the case, while a credence of 0 in *p* represents complete confidence that *p* is *not* the case; credences between 1 and 0 thus represent uncertainty whether *p* is the case. This is a useful device, since it provides an intuitively appealing picture of how one can have, and how it can be reasonable to have, some degree of credence in a proposition while also having some degree of credence in its negation. But I would caution against too strong or uncritical a reliance on it, for two reasons.

First, it seems that in many, perhaps all cases there will be no fact of the matter as to *just* how strong one's credence in a proposition is. One's credence in a proposition may fall within some range—between 0.8 and 0.9, say—and yet not have any precise value within that range. It smacks of absurdity to claim, for example, that someone has a credence of precisely 0.83256 in some proposition. True, this absurdity might on occasion reside in the fact that no one can know that some credence has this precise strength, but on other occasions it seems to reside in the very idea of a credence having any precise strength. (Compare the claim that, regardless of whether there is a useful metric available to gauge strength of desire, one person has a desire for cake that is *precisely* as strong as another person's desire for ice cream.) The fact, as I see it, that credences can be imprecise raises some interesting questions.

One question, perhaps not too pressing, concerns the relation between credence and confidence. A precise credence of, say, 0.9 or a moderately imprecise credence between, say, 0.8 and 0.9 represents a fairly high degree of confidence in the proposition in question. Likewise, a precise credence of, say, 0.1 or a moderately imprecise credence between, say, 0.1 and 0.2 represents a fairly low degree of confidence in a proposition. But what degree of confidence should be associated with a massively imprecise credence between, say, 0.1 and 0.9? It would be misleading to say either that it is high or that it is low. It would also be misleading to say that it is middling, since that suggests a credence of, or close to, 0.5. Perhaps the best that can be said is that the degree of confidence is "broad," but I'm not sure whether that makes any sense. I'm inclined to think, therefore, that in this context talk of confidence may be misplaced. Some may take that to be a reason to deny that such massively imprecise credences are possible, but it seems plain to me that they are. (I give an example in the next paragraph.)

Another, far more pressing question concerns just how we should conceive of imprecise credences. According to one popular proposal, imprecise credences are

[34] Note that how *confidently* one holds a belief should be distinguished from how *firmly* one holds it. The latter has to do with how reluctant one is to give up the belief. For discussion, see Pojman (1986), pp. 154 f. and Moon (2017), p. 768.

representable, not (as precise credences are) by a single probability within the range of 0 to 1, but by a closed interval that comprises a set of probability functions that span that range.[35] But this convention seems to me to mask, and sometimes to misrepresent, the various ways in which credences can be imprecise. One may have an imprecise credence due to the fact that one's evidence is incomplete, or conflicted, or involves considerations that are in some respect incomparable, and the type of imprecision can vary accordingly and in a manner that is not accurately reflected in the proposed model. Consider, for example, a case in which you have good but incomplete evidence regarding a coin: you know that it is biased in such a way that, when tossed, it either lands heads 75% of the time or tails 75% of the time, but, whether because of lack of evidence or because of conflicting evidence, you don't know which of these biases it has. As a result, your credence regarding the proposition that the coin will land heads the next time it is tossed is imprecise. Is this credence properly represented by the interval [0.25, 0.75]? Presumably not, since, on the present approach, this represents a kind of credal indifference on your part across the range of probabilities contained in the interval. A better representation would seem simply to be the pair of numbers {0.25, 0.75}.[36] Or consider a case in which your evidence is very meager indeed: you know that the coin has some bias, but you have no idea what that bias is. Is your credence that it will land heads the next time it is tossed properly represented by the interval [0, 1]? Presumably not, for two reasons. First, this interval contains 0.5, which represents no bias at all. Second, this interval also contains both 0 and 1, which fall outside the range of relevant probabilities.

The second reason why I would caution against too strong a reliance on thinking of credences as being in some way representable as falling between 0 and 1 is that it suggests that one's credences regarding a proposition must sum to 1. Although this is arguably the case for someone who is fully rational,[37] there can be no guarantee that the credences of someone who is not fully rational will be so well behaved.

The expression "degrees of belief" strongly suggests that credences are degrees of *belief*, but in fact the relation between credence and what I, along with many others, will call *outright* belief is not at all straightforward.[38] One simple proposal about this relation is that one has an outright belief that p is the case if and only if

[35] For a helpful discussion of this idea, and of imprecise credences generally, see S. Bradley (2019).

[36] Both the example and the proposed representation are drawn from Mayo-Wilson and Wheeler (2016), p. 61.

[37] Note that this claim is compatible with the claim that not only precise but also imprecise credences can be rational, if the latter are representable by closed intervals. For example, rationality may require not only that, if one has a credence of, say, 0.9 in p, then one have a corresponding credence of 0.1 in $\sim p$, but also that, if one has a credence, say, of [0.8, 0.9] in p, then one have a corresponding credence of [0.1, 0.2] in $\sim p$.

[38] Alternative but, I think, less apt expressions are: "full belief," "plain belief," "belief flat out," "belief period," and "belief full stop."

one's credence in p is 1. This proposal seems clearly unacceptable, since it rules out the possibility of holding a belief with less than full confidence, a possibility that would appear to be frequently realized. Another proposal is that one has an outright belief that p is the case if and only if one's credence in p is greater than 0.5. But this proposal is unacceptable, too. Suppose, to borrow a well-known example, that you have just bought a lottery ticket. In all, 100 tickets have been sold. You believe (that is, you have an outright belief) that the lottery is a fair one, and so your credence in the proposition that you will not win the lottery is 0.99. Do you therefore believe that you will not win? Not necessarily. Given your belief that the lottery is fair, you also have a credence of 0.99 in each of the following propositions: Ticket 1 is not the winning ticket, Ticket 2 is not the winning ticket, . . . , and Ticket 100 is not the winning ticket. Call these propositions P_1, P_2, . . . , and P_{100}, respectively. Now, suppose that your beliefs happen to conform to a norm, one that many philosophers endorse,[39] that requires that, if one believes some proposition p and also believes some proposition q, then one believes $p\&q$. Then, if you believed each of P_1, P_2, . . . , and P_{100}, you would also believe P_1 & P_2 & . . . & P_{100}. But you *don't* believe this long conjunctive proposition. On the contrary, you are *certain* that it is *false* (that is, your credence in it is 0), since you are certain that one of the 100 tickets is the winning ticket. Hence you don't believe *any* of P_1, P_2, . . . , and P_{100}, even though your credence in each of them is very high (0.99, which of course is much higher than 0.5).[40]

Credences are often said to consist in assignments of "subjective probability."[41] This might be taken to suggest that we say, not that one believes that p *is* the case, but rather that one believes that p is *probably* the case, if and only if one's credence in p is greater than 0.5. But this is problematic, too. I think that, on one understanding, this proposal is probably *correct* (that is, my credence in it is relatively high), but only if "believes" is understood as *not* referring to outright belief.[42] I say this for two reasons. First, an outright belief that p is the case requires that one grasp or understand p, and it seems likely that there are many individuals (small children, for example, and perhaps also some animals) that have credences but do not have any grasp of the concept of probability. Second, there are different kinds of probability. A commonly-acknowledged, even if somewhat obscure, threefold distinction is that between objective, epistemic, and subjective probability. It is clear that one's credences needn't align with one's beliefs about either objective or epistemic probabilities (whatever it is that these consist in, precisely). For example, one might acknowledge that only 10% of the people with the same medical condition from which one is suffering survive for more than a few weeks

[39] Though not all. See Foley (2009).

[40] Perhaps it is also possible for one to have an outright belief that p is the case even though one's credence in p is less than 0.5. Such a belief would no doubt be irrational, but that doesn't make it impossible.

[41] See, e.g., Huber (2009), pp. 4 ff.

[42] Contrast Moon and Jackson (2020).

(a matter of objective probability) and also acknowledge that there is no good reason to think that one's own case is exceptional (a matter of epistemic probability), while nonetheless being optimistic about—that is, having a high credence in—one's own long-term survival. Perhaps such optimism is irrational, but, as long as it is possible, it shows that the present proposal cannot be accepted, if it is either objective or epistemic probability that is at issue. That leaves subjective probability. Given that the subjective probability of a proposition is the measure of one's credence in it, this interpretation of the present proposal amounts to the claim that one's credence in *p* is greater than 0.5 if and only if one believes that one's credence in *p* is greater than 0.5. This, again, is clearly unacceptable, since one may be mistaken about how confident one is that *p* is the case.

In light of considerations such as the foregoing, I am inclined to believe that it is not possible to provide a satisfactory explication of credence in terms of outright belief or vice versa. Having a belief in a proposition of course does not preclude having a credence in it—on the contrary, the former presumably requires the latter; nonetheless, as we have seen, the latter does not require the former.

[Brief digression: It is perhaps worth taking note here of a somewhat mysterious phenomenon, and that is that, in certain contexts, making explicit mention of the fact that one believes some proposition serves to indicate that one's credence in it is not as high as it might otherwise seem. Suppose that you have just come from a meeting with Bob. A colleague asks you, "Where's Bob?" You reply, "He's in Room 216." Notice two things.[43] You have asserted that Bob is in Room 216; you have *not* asserted that you believe that he is in Room 216, although this belief is what is commonly known as a conversational implication of your assertion (that is, it would be reasonable, under the circumstances, for your colleague to infer that you do indeed have this belief). Now suppose that 15 minutes have passed, and another colleague comes along and asks, "Where's Bob?" This time you reply, "I *believe* he's still in Room 216." Notice two more things. This time, you have explicitly asserted that you believe that Bob is in Room 216, and your doing so serves paradoxically to indicate that you hold your present belief with *less* confidence than the belief you had 15 minutes ago. Of course, you could have given an even more cautious reply, such as "I'm *inclined to think* that he's still in Room 216," or "I *suspect he may* still be in Room 216," and the like.[44]]

The apparent independence of credence from belief raises an important question about my project in this book, a project that I have characterized as an

[43] Cf. G. E. Moore (2005), pp. 63 f.; B. Williams (1973), p. 138.

[44] Such hedging is common in many recent philosophical works, including this one (for confirmation, see the opening sentence of the previous paragraph). I recall Nicholas Rescher decrying this practice in private correspondence; he also published an editorial lamenting it (see Rescher (1984)). Although the practice can be irritating, it is, I believe (note again!), a mistake to call for its elimination. It signals the tentativeness with which one makes a claim, and it invites discussion. Feigning confidence is misleading and counterproductive.

examination of the relation between moral responsibility and the failure to *believe* that what one is doing is (overall morally) wrong. The question is whether this project, or my characterization of it, is too narrow, insofar as it doesn't take *credence* into account. This question may seem especially pressing when it comes to giving a satisfactory account of responsibility for *reckless* behavior, since, as Husak notes in the passage quoted above, such behavior is, by definition, behavior that involves running some kind of *risk*, and risk is to be understood in part in terms of probabilities. I will take up this important question in §7.1.4. For the time being, though, I will continue to portray my project as one that is concerned with belief about, rather than credence in, wrongdoing.

3.2.5 Conflicting beliefs

It is possible for someone to have conflicting beliefs. By this I mean that it is possible for one and the same person at one and the same time both to believe some proposition *p* and to believe some proposition *q*, even though it is impossible for it to be the case that *p*&*q* is true. (An analogous account of conflicts in intention and conflicts in desire can be given.) The impossibility in question might be merely empirical, or it might be something stronger. George has a conflict of the former kind if he believes that Len was in London yesterday while also believing that Len is in Paris today, when in fact all travel between the two cities has been closed off for over a week (the explanation of the conflict being that George doesn't know of the travel ban). Rhonda has a conflict of the latter kind if she believes that all squares are rectangles while also believing that no rhombuses are rectangles (the explanation of the conflict being that Rhonda isn't aware that rhombuses needn't have acute angles).

Might the conflict be strictly logical? That is, to put the question schematically, is it possible for there to be cases in which both *Bp* and *B*~*p* are true for one and the same person at one and the same time? In a well-known article, Saul Kripke has argued that there is reason to think that the answer to this question is "Yes."[45] He gives the example of a Frenchman, Pierre, who speaks only French and, as a result of what others have told him, believes the proposition that he expresses by saying, "Londres est jolie." Subsequently, Pierre travels to England, learns some

[45] Kripke (1979). The question at issue should be carefully distinguished from the following questions:

(1) Is it possible for one and the same person at one and the same time to have both a non-zero credence in *p* and a non-zero credence in ~*p*?
(2) Is it possible for *B(p*&~*p)* to be true?
(3) Is it possible for *Bp* and ~*Bp* both to be true?

The answer to the first question is clearly "Yes" and to the third clearly "No." I will leave the second question open.

rudimentary English, and takes up lodgings in an unattractive part of a city that, through conversation with the local inhabitants, he comes to know as "London." As a result, he sincerely assents to the proposition that London is not pretty, all the while apparently retaining his earlier belief that it is pretty. Is this a genuine case in which both *Bp* and *B~p* are true? The matter is controversial. It's worth noting that, if it is, the explanation of the conflict in belief rests once again on the fact that the person who has the conflicting beliefs (in this case, Pierre) is ignorant of some crucial pertinent fact (in this case, that "Londres" and "London" refer to the same city).

Are there cases in which both *Bp* and *B~p* are true whose explanation is *not* to be found in the person's being ignorant of some such underlying fact? Tamar Szabó Gendler has presented several kinds of cases that may seem to fit this description.[46] One that she discusses at length involves a person—call her Patty—who is near the edge of a precipice. She believes—she knows—that she is quite safe, since there is a high fence that prevents her from falling; nonetheless, she pulls anxiously away from the edge, thereby showing, it might be claimed, that she also believes that she is not safe. Such a situation is not at all uncommon. Other examples that Gendler gives include the common reluctance to eat fudge that is shaped like dog feces, even when one knows it to be fudge and not feces; the common experience of being terrified by what one sees on a movie screen; and the like. Perhaps we might also include the common tendency to speak back to the GPS in one's car (mine is called Emily).

Some commentators do indeed diagnose a conflict in belief in such cases.[47] Others offer a different diagnosis. Some, appealing to the thesis that "believes" is a vague predicate, claim that such cases involve a conflict between attitudes that, though akin to belief, do not fully qualify as such.[48] Gendler herself holds that such cases do *not* involve conflicting beliefs. On the contrary, she claims that they involve a conflict between a belief and another kind of attitude, one that she calls an "alief." As Gendler conceives of it, an alief is a mental state that is in several respects quite different from a belief. Among the differences are these: beliefs typically change in response to changes in evidence, whereas aliefs do not; and aliefs typically drive behavior "automatically," without the intervention of conscious thought and without the mediation of desires, whereas beliefs do not. Thus, although Patty believes that she is safe, she nonetheless alieves that she is not; she would revise her belief if she were to discover that the fence was damaged, whereas her sense that she is in danger persists despite her evidence. So too, although I believe that Emily is not a real person, I nonetheless alieve that she is a real person (especially when I get annoyed with her when she keeps saying

[46] Gendler (2008a) and (2008b).
[47] See, e.g., Borgoni (2015).
[48] See, e.g., Schwitzgebel (2010).

"Recalculating," the equanimity with which she does so serving only to add to my vexation). And so on for the other cases.

A particularly interesting kind of case in which Gendler claims that there is a conflict between a belief and an alief, one to which I will return briefly in §9.4, is that of someone—I'll call her Veronica—who is an avowed anti-racist. A white woman, she endorses racial equality with conviction, and frequently speaks out against those who hold racist views, and yet her behavior betrays the fact that she herself has strong racist tendencies. She automatically associates blacks with violence, tends to assume that they are intellectually inferior to her, and so on right down the line of wretched stereotypes, and she is usually quite unaware that she is doing so. On those rare occasions on which she catches herself out, she reacts to her own behavior with shame and bewilderment. According to Gendler, Veronica (whose case I have embellished a little) has a non-racist belief coupled with a contrary racist alief.

Not everyone accepts Gendler's assessment of cases such as Veronica's. Some critics point out that beliefs are not always responsive to evidence (a fact that Gendler herself acknowledges with the wry remark, "[T]hink about flat earthers, Roswellians, or your political opponents."[49]), whereas some automatic responses are (think of slamming on the brakes when a dog runs out in front of your car).[50] Some say that it can be appropriate to say of people like Veronica that they don't "really" believe "deep down" in racial equality.[51] Others say that such people do believe in racial equality but also "endorse" racial inequality in a "patchy" manner.[52] Still others say that such people do indeed have conflicting beliefs, one conscious and open to rational assessment and control, the other unconscious and rationally inaccessible.[53] It might seem important that this issue be settled, and indeed for some purposes (e.g., that of providing an accurate map of the mind) it surely is. For my purposes in this book, however, I don't think it matters one way or the other whether we say that Veronica's racist behavior is to be attributed to a belief of hers, to an alief, or to something else. This claim may surprise you. I will offer an explanation for it in §8.2, one that I will then go on to qualify in §9.4.

3.2.6 Failing to believe

Let me turn, finally, from believing to failing to believe. One fails to believe some proposition *p* just in case one does not believe it; that is, just in case $\sim Bp$.[54] There

[49] Gendler (2008b), p. 566, n. 26. Cf. Archer (2018). [50] Cf. Schwitzgebel (2010), p. 539.
[51] Cf. Schwitzgebel (2010), p. 539. [52] Cf. Levy (2017a), p. 13.
[53] Cf. Borgoni (2015). Cf. also Kornblith (2017), pp. 2586 f.; McHugh (2017), p. 2752.
[54] In saying this, I am presupposing that one currently exists. Notice that, for any proposition, it's true that Aristotle doesn't believe that proposition, but it would be odd to say that he fails to believe it.

are two basic ways in which it might be true that $\sim Bp$. First, one might *disbelieve* p; that is, $B\sim p$. Of course, as has just been discussed, it may be that its being true that $B\sim p$ is compatible with its also being true that Bp, but it seems safe to say that this is not the norm. On the contrary, the fact that it is true that $B\sim p$ is often what explains why it is the case that $\sim Bp$. Second, one might be *doxastically uncommitted* regarding p; that is, $\sim Bp \ \& \ \sim B\sim p$.[55] There are three types of such uncommittedness.[56] First, one may have some credence in the proposition and yet fail to believe it outright. In such a case one may be said to *suspect* that the proposition is true. (If one is rational, this suspicion will be accompanied by a complementary suspicion that the proposition is false.[57]) Second, one may have no credence whatsoever in the proposition, in virtue of never having considered it and thus being wholly *oblivious* to it. Third, one may have no credence in the proposition, despite having considered it, in virtue of being at a complete loss what to think about it—in which case I will say that one is *utterly baffled* by it.[58,59]

The most recent version of the Argument from Ignorance given in §3.1.1 is couched simply in terms of "failing to believe." No distinction is drawn between cases of disbelief and cases of doxastic uncommittedness, let alone between the different types of such uncommittedness. Do these distinctions matter? It may seem so. It may seem that whether one is blameworthy for ignorant wrongdoing can turn on whether or not one believes that one is doing the right thing, or merely suspects it, or is oblivious to or utterly baffled by the issue. Similarly, the most recent version of the Argument draws no distinction between the failure to have an occurrent belief that one is doing wrong, the failure to have a dormant belief that this is so, and the failure to have a tacit belief (if such a belief is possible) that this is so, and again it may well seem that this can matter as far as blameworthiness is concerned. I will take up these issues in Chapters 6 and 7.

[55] Cf. Chisholm (1976), p. 27; Peels (2017), pp. 165 ff.

[56] "Neutrality" would be a less awkward term than "uncommittedness," but it suggests a credence of 0.5, which of course is not necessary for one's neither believing nor disbelieving a proposition.

[57] In this context, I am taking "suspect" to mean the same as "have a (or some) suspicion," no matter how weak one's suspicion may be. In everyday usage, "suspect" is typically restricted to cases in which one's suspicion consists in having a credence greater than 0.5; where one's credence in a proposition is less than that, it would normally be said that one suspects that the proposition *may* be true. Also, it would not normally be said that one suspects a proposition to be true if one believes it outright.

[58] I suspect that utter bafflement (as I am using the term) occurs only rarely. A "credence" of [0, 1] may perhaps qualify as utter bafflement, being equivalent to one's having no credence at all, but a credence with a narrower range will not qualify as such.

[59] As to *suspension of judgment*: *if* we construe this as a way of failing to believe a proposition, I take it to consist in either suspecting that the proposition in question is true or being utterly baffled by it. I think, though, that suspension of judgment is typically construed as a way of failing to *accept* a proposition rather than of failing to believe it.

3.3 Acting from ignorance

Whenever one unwittingly does wrong, one commits the wrongdoing *in* ignorance of the fact that one is doing so. It is common to distinguish two ways in which this may happen. One acts *from* ignorance of the fact that what one is doing is wrong when one's ignorance *accounts* for one's wrongdoing; that is, the wrongdoing is to be *attributed* to one's ignorance; that is, one does what one does *because* of one's ignorance; that is, roughly, one would *not* commit the wrongdoing if one were not ignorant of the fact that it is wrong. In contrast, one acts *merely in* ignorance of the fact that what one is doing is wrong when one's ignorance does *not* account for one's wrongdoing; that is, the wrongdoing is *not* to be attributed to one's ignorance; that is, it is *not* the case that one does what one does because of one's ignorance; that is, roughly, one *would* still commit the wrongdoing even if one were not ignorant of the fact that it is wrong. This distinction can be traced to Aristotle, although it's not clear that he understood it precisely as I have just articulated it.[60] It's a distinction that many find significant, claiming that ignorance provides an excuse for wrongdoing only when the wrongdoing is committed from, and not merely in, ignorance.[61]

This view is plausible. Why should ignorance be thought to provide an excuse for wrongdoing if it is merely idle, supplying no explanation of one's behavior? Still, some remain unconvinced, claiming in turn that the fact that one would have engaged in the same behavior under counterfactual circumstances (circumstances in which one is not ignorant of the fact that the behavior is wrong) cannot be the grounds of one's culpability in the actual circumstances.[62] I think this is a mistake. As I see it, counterfactuals are highly relevant to the determination of actual culpability,[63] but the present view need not be cast in terms of counterfactuals anyway. All that's being claimed is that one's *actual* mental state of ignorance regarding one's wrongdoing furnishes an excuse only if it plays a certain role in the ancestry of that wrongdoing. If this view is correct, then the Argument from Ignorance requires further revision. (If it is not correct, then of course no revision is called for. But I'm inclined to think that it is correct, and caution dictates assuming that it is.) The following is perhaps a little rough (because "because" is a tricky term), but I hope it will do:

The Argument from Ignorance (Draft 4):
Suppose that

(1) (a) *P* committed *A* and (b) *A* was overall morally wrong, but (c) when *P* committed *A*, ***P* did so because** *P* failed to believe that *A* was overall morally wrong.

[60] Aristotle (1941), 1110b.
[61] See, e.g., Donagan (1977), pp. 128 ff.; Rivera-López (2006), p. 135.
[62] See Sarch (2014), p. 1059. Cf. Rosen (2008), p. 598, n. 14; Husak (2016), p. 224.
[63] See Zimmerman (2002).

In general it's true that

(2) if one committed some act or omission that was overall morally wrong **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.

Thus

(3) *P* is morally to blame for *A*, and thereby morally to blame for any of its consequences, only if *P* is morally to blame for *P*'s failure to believe (call it *F*) that *A* was overall morally wrong.

It's also in general true that

(4) one is morally to blame for something only if one was in control of that thing.

Thus

(5) *P* is morally to blame for *F* only if *P* was in control of *F*.

Three further general truths are that

(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.

Thus

(9) *P* is morally to blame for *A*, and thereby to blame for any of its consequences, only if there was something else (call it *X*) for which *P* is directly morally to blame and of which *F* was a consequence.

But

(10) whatever *X* was, it cannot have been an act or omission such that (a) *P* committed it and (b) it was overall morally wrong and (c) when *P* committed it, ***P* did so because** *P* failed to believe that it was overall morally wrong, since otherwise the foregoing argument regarding *A* would apply all over again to *X*.

Thus

(11) whatever *X* was, it was either (a) not some act or omission that *P* committed or (b) not overall morally wrong or (c) **not** something such

that, when *P* committed it, ***P* did so because *P* failed to believe** that it was overall morally wrong.

Two further general truths are that

(12) one has direct control over something only if that thing is an act or omission, and

(13) one is morally to blame for an act or omission only if that act or omission was overall morally wrong.

Thus

(14) whatever *X* was, it was **not** an act or omission such that, when *P* committed it, ***P* did so because *P* failed to believe** that it was overall morally wrong.

The version of the Origination Thesis that is to be derived from this version of the Argument is rather cumbersome. It goes as follows:

The Origination Thesis (Draft 4):
Every chain of moral blameworthiness is such that at its origin lies a piece of behavior for which the agent is directly blameworthy and which **was such that either** he or she believed, at the time at which he or she engaged in it, **that it was** overall morally wrong **or he or she did not believe this but did not engage in the behavior because he or she did not believe this**.

That's quite a mouthful, but the gist of this version of the Thesis can be captured much more pithily, even if a little roughly. Let us say that wrongdoing that is committed merely in ignorance is wrongdoing that is *quasi-witting*. Then what the Thesis says is, roughly, that all blameworthiness is rooted in wrongdoing that is either witting or quasi-witting.

3.4 Blameworthiness without wrongdoing

As I mentioned above, ignorance is ignorance of facts. The particular form of ignorance with which this book is concerned is ignorance of the fact that one is doing something wrong. But must blameworthiness presuppose some such fact?

3.4.1 Accuses

I noted in the last chapter that it is possible to have an excuse for wrongdoing. Sally, whose child was born with phocomelia as a result of her taking a dose of Distaval, is surely not to blame for her child's condition. Whether she has an

excuse for wrongdoing is perhaps not so clear, since there is some debate about whether behavior such as hers constitutes wrongdoing in the first place. But there are other cases in which it seems that such a verdict is well-nigh irresistible. In §2.1.3 I gave the case of Constance, who constantly tries very hard to ensure that she does the right thing on all occasions. Despite her best efforts, she nonetheless ends up sometimes doing wrong. On such occasions, she is surely not to blame; she has an excuse—a fully exculpatory excuse.

If there can be wrongdoing without blameworthiness, the question naturally arises whether there can be blameworthiness without wrongdoing. I think there can be. As I observed in Chapter 1, philosophers have traditionally identified two conditions, the dual satisfaction of which renders one blameworthy for one's behavior, one having to do with one's mental state, the other having to do with whether one was in control of one's behavior. As I noted in §2.2.1, nowadays the former condition is often couched in terms of one's "quality of will." I expressed reservations about such a use of this term, but let us put these aside for the moment. The question I want to raise is simply this: if it is possible for one to do wrong—something that is not only consistent with but, arguably, *requires* exercising control over one's behavior—and yet *fail* to exhibit the quality of will necessary for blameworthiness, why should it not also be thought possible for one to exercise control over one's behavior but *not* do wrong, while nonetheless *exhibiting* such quality of will? And, if this is possible, why should this not be deemed sufficient for one's being blameworthy for what one has done?

The possibility of being to blame for what one has done, even if one has done no wrong, used rarely to be entertained, let alone endorsed.[64] There is no term in common use to express this possibility. In an earlier work, I coined the term "accuse" (used as a noun, with a hard "s"), which I defined as the reason or grounds for imputing blameworthiness in the absence of wrongdoing,[65] and I argued that accuses are not only possible but indeed arise on occasion. What might count as such an occasion? Peter Graham provides the following case:

> Unbeknownst to Bob, Sue is about to detonate a bomb that will kill a thousand innocent people. The only way to prevent her from doing so is to kill her. Bob, for his own personal reasons, shoots her dead.[66]

According to Graham, Bob is blameworthy despite having done no wrong.

[64] Acknowledgment of this possibility is growing. Cf. G. E. Moore (2005), pp. 95 ff. (first published in 1912); Brandt (1958), pp. 38 f.; Donagan (1977), p. 112; Milo (1984); Parfit (1984), p. 25; Jackson (1986), pp. 362 f.; Thomson (1991), p. 295; Copp (1997), p. 448; Haji (1998), p. 146; Hieronymi (2008), p. 362, n. 12; Scanlon (2008), p. 125; Capes (2012); Khoury (2012), p. 198; Graham (2014), pp. 394 ff.; Levy (2014), p. 36; Shoemaker (2015a), p. 699.

[65] Zimmerman (1997a). [66] Graham (2014), p. 394.

One reaction that you might have to Graham's case is to deny the claim that Bob did nothing wrong. Perhaps an objective view of moral obligation, of the sort I mentioned in §2.1.2, would declare Bob's behavior morally right, indeed morally obligatory, but, you might think, that just shows why we should reject such a view. Bob *had no reason to believe* that, in killing Sue, he was thereby saving a thousand innocent people from death, let alone that this was the only way to do so. You might think that, in light of this fact, what he did was therefore morally wrong after all, and so his also being to blame for what he did comes as no surprise.

To this, though, Graham could respond as follows.[67] All we need do is tweak the case a little so that Bob *is* aware of the threat that Sue poses to the innocent people, while letting it remain the case that he kills her for his own personal reasons. (We may imagine that he doesn't care at all about the innocent people. He simply sees an opportunity to get rid of Sue, whom he has hated ever since she was promoted over him.)

You might respond in turn by saying that we must take care to specify just what it is that Bob is to blame for. We shouldn't say that he is to blame for his action of *killing Sue*, precisely because this action wasn't wrong. But we can and should say that he is to blame for *killing Sue for his own personal reasons*, and *this* action was indeed wrong.[68] But I don't think this will do. Even if we agree that it was not wrong for Bob to kill Sue but it was wrong for him to kill her for his own personal reasons (a pair of claims that it would take some rather delicate theorizing to reconcile with one another), there seem to be cases that cannot be handled in this way. Suppose (this may take some stretch of the imagination) that Bob was so constituted that he *could not* kill Sue unless he did so for his own personal reasons. Then, on the present approach, he would be in a dilemma; he would be obligated to kill Sue and he would be obligated to refrain from doing so for his own personal reasons, and yet he could not do both of these things. I submit, but (as before[69]) I will not argue here, that this is an untenable result.

But then, you may well ask, if Bob isn't to blame either for killing Sue or for killing her for his own personal reasons, what *is* he to blame for? There are several possible answers. One is to say that he is to blame for doing what he *believed* was wrong. I don't think this will do, since Bob might not have had the belief in question. (This point is compatible with the Argument from Ignorance, as long as Bob is to blame for lacking the belief.) A second possibility is that what Bob is to blame for is not any *action* of his but an *attitude*. This is the approach recommended by Graham.[70] A problem with this approach is that it is difficult to see how it can be reconciled with the agential condition of control. Bob's hatred for Sue may be despicable, but how can he be to blame for having this attitude if it wasn't in his control? (This is not a purely rhetorical question. I will take it up in

[67] See Graham (2014), p. 395, including n. 13.
[68] Cf. Rosen (2015), p. 76. Cf. also Nelkin (2011), p. 107.
[69] See the very end of §2.1.2.
[70] Graham (2014), pp. 396 ff. Cf. Shoemaker (2015a), p. 699.

Chapter 9.) A third possibility is to draw on the distinction that I discussed in §2.2.4 between that for which one is to blame and that in virtue of which one is to blame. We might say that, in the case in question, Bob is not to blame *for* any action that he committed (let alone any attitude that he had), but he nonetheless is to blame *in virtue of* the quality of will that underlay his behavior. On this approach it is a mistake to think that, if one is blameworthy, one must be blameworthy for something, although it remains true that one must be blameworthy in virtue of something. I think that this approach has much to recommend it, but the idea is jarring and I won't press it here. I suggest instead that we embrace a final possible answer, which is to deny the presupposition and say that Bob is indeed to blame for killing Sue, even though that was the right thing for him to do. This no doubt sounds odd, but I think that it is nonetheless correct. Notice that accepting this answer does *not* require that we say that Bob is to blame for doing the right thing, an expression that may be taken to suggest, erroneously, that he is to blame for what he did *because* it was right.

No account of blameworthiness is complete unless it takes a stand on the possibility of accuses. If accuses are possible, then blameworthiness is likely incurred more often than is commonly supposed. This is a significant result, one that in a way runs counter to (although it is perfectly compatible with) the apparent implication of the Origination Thesis that blameworthiness is incurred less often than is commonly supposed. It is therefore important to try to determine whether accuses are possible and, if they are, which of the four approaches just outlined is correct. This point notwithstanding, it is in fact not important for *present* purposes to accomplish this task, for the Argument from Ignorance can be formulated in such a way that it remains neutral on the question of whether accuses are possible. The most recent version (Draft 4), which is explicitly couched in terms of one's *failing to believe* that one is doing wrong, lends itself easily to such a formulation. All that needs to be done is to delete any reference to the fact that one's behavior is indeed wrong, thereby leaving it open whether the belief in question, were one to have it, would be veridical. The result is the following simpler argument, the deletions being marked by bracketed ellipses:

The Argument from Ignorance (Draft 5):
Suppose that

(1) (a) *P* committed *A* [. . .] but (**b**) when *P* committed *A, P* did so because *P* failed to believe that *A* was overall morally wrong.

In general it's true that

(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.

Thus

(3) *P* is morally to blame for *A*, and thereby morally to blame for any of its consequences, only if *P* is morally to blame for *P*'s failure to believe (call it *F*) that *A* was overall morally wrong.

It's also in general true that

(4) one is morally to blame for something only if one was in control of that thing.

Thus

(5) *P* is morally to blame for *F* only if *P* was in control of *F*.

Three further general truths are that

(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.

Thus

(9) *P* is morally to blame for *A*, and thereby morally to blame for any of its consequences, only if there was something else (call it *X*) for which *P* is directly morally to blame and of which *F* was a consequence.

But

(10) whatever *X* was, it cannot have been an act or omission such that (a) *P* committed it [. . .] and (**b**) when *P* committed it, *P* did so because *P* failed to believe that it was overall morally wrong, since otherwise the foregoing argument regarding *A* would apply all over again to *X*.

Thus

(11) whatever *X* was, it was either (a) not some act or omission that *P* committed [. . .] or (**b**) not something such that, when *P* committed it, *P* did so because *P* failed to believe that it was overall morally wrong.

One further general **truth is** that

(12) one has direct control over something only if that thing is an act or omission. [. . .]

Thus

(13) whatever *X* was, it was not an act or omission such that, when *P* committed it, *P* did so because *P* failed to believe that it was overall morally wrong.

Given the deletion of any reference to the *fact* that one's behavior is wrong, a more accurate label for this argument would be "The Argument from the Failure to Believe." But I will stick with the original label, which is much more attractive.

This latest revision to the Argument from Ignorance requires no revision to the Origination Thesis, but I will nonetheless restate the latter under a new draft number so as to keep the numbers for Argument and Thesis aligned:

> *The Origination Thesis (Draft 5):*
> Every chain of moral blameworthiness is such that at its origin lies a piece of behavior for which the agent is directly blameworthy and which was such that either he or she believed, at the time at which he or she engaged in it, that it was overall morally wrong or he or she did not believe this but did not engage in the behavior because he or she did not believe this.

As long as "witting" is understood expansively, so that one engages in witting wrongdoing simply in virtue of doing what one believes, even if mistakenly, to be morally wrong, this version of the Thesis can once again be understood, roughly, as the thesis that all blameworthiness is rooted in wrongdoing that is either witting or quasi-witting.

3.4.2 Moral realism

Apart from the fact that it accommodates (but does not entail) the possibility of accuses, it may seem that Draft 5 of the Argument from Ignorance has a further advantage: it doesn't entail that there are any *facts* at all about whether someone has done something morally wrong. This is a feature that a moral non-realist might welcome.

I don't subscribe to such non-realism. I think that there are facts about moral right and wrong—and about moral praise- and blameworthiness—of which people are unfortunately all too often ignorant. If this weren't so, the Argument from Ignorance and the Origination Thesis would be devoid of almost all significance, it seems to me, although perhaps a non-realist might be able to put them to some use.

I won't try to defend moral realism here. That would be too great a digression. I will rest content with the observation that, if you have read the book up to this point, you probably have considerable sympathy for such realism yourself.

3.4.3 Suberogation

Some philosophers have claimed that there is a way in which one can be blameworthy without doing wrong that involves a departure from "common morality"

that is perhaps less radical than that discussed so far, namely, the possibility of one's being to blame for *suberogatory* behavior.[71] Suberogation is to be understood as a kind of mirror-image of the more familiar phenomenon of supererogation. Whereas the latter consists in doing something that is *commendable* but which it is permissible *not* to do, the former consists in doing something that is *objectionable* or *indecent* but which it is nonetheless permissible *to* do.

It is well known that, despite the fact that common morality recognizes the possibility of supererogation (if not by that name, then by the description "going above and beyond the call of duty"), it is remarkably difficult to develop a plausible moral theory that accommodates this possibility. The possibility of suberogation, however, is far less often discussed in either of these contexts. And there may be good reason for this, since, mirror-image notwithstanding, the symmetry between super- and suberogation is not perfect. The mirror is distorted. For consider: the greater the extent to which an act goes beyond duty (on the assumption that this is indeed possible), the more supererogatory it is; its status as being something that it is permissible not to do is not thereby threatened. There is no parallel when it comes to suberogation. That is, it is not the case that, the greater the extent to which an act falls short of decency, the more suberogatory it is; on the contrary, its status as being something that it is permissible to do *is* thereby threatened, in that at some point it will presumably become impermissible or wrong. And it is arguable that this point is reached immediately, as it were, so that the status of being suberogatory is never attained. So why, exactly, should we accept the possibility of suberogatory behavior, let alone the possibility of being blameworthy for such behavior?

Consider some alleged examples of suberogatory behavior: a minor act of discourtesy, such as taking too long in a restaurant when you know that others are waiting for a seat,[72] or taking a seat on a crowded train, thereby preventing a couple from sitting together, when taking another, less convenient seat is an option;[73] or, more seriously, an odious act such as driving a competitor out of business when one's own business is flourishing,[74] or refusing to donate a kidney to one's brother, who will die without it, when doing so would not be particularly onerous.[75] These examples, offered by Roderick Chisholm and Ernest Sosa and by Julia Driver, undoubtedly have some intuitive appeal; nonetheless, I am inclined to think that it is a mistake to classify them as instances of suberogation. It is noteworthy that Chisholm and Sosa, in their article, and Driver, in hers, explicitly say that, in cases of the sort in question, the agent is acting "within his rights," from which they infer that the agent is doing nothing wrong. But that inference is questionable, for it seems plausible to hold that one can do wrong even though

[71] See, e.g., Haji (1998), p. 173; McKenna (2012), pp. 182 ff.
[72] See Chisholm and Sosa (1966), p. 326.
[73] See Driver (1992), pp. 286 f.
[74] See Chisholm and Sosa (1966), p. 326.
[75] See Driver (1992), p. 287.

one is acting within one's rights. One acts within one's (liberty-)rights in performing some act when no one has a (claim-)right against one that one not perform that act, that is, when, in performing that act, one violates no obligation that one *owes to* someone else.[76] But not all obligations, it seems, are obligations that are owed to someone. Consider Peter Singer's famous case of the child drowning in the pond.[77] Singer argues, plausibly, that you would do wrong to pass the child by, but at no point does he rest his argument on the contention that the child has a right to your help. So too, I submit, in the cases offered by Chisholm and Sosa and Driver, the behavior in question is indeed wrong, and not merely suberogatory, even though it involves no violation of anyone's rights.

There are difficulties, of course. Singer uses his case as the basis for an argument to the effect that very many of the acts that we would normally classify as supererogatory (such as donating a large proportion of one's assets to charity) are in fact obligatory, and many people, myself included, balk at this conclusion, although I confess it is not easy to tell just where Singer's argument goes wrong (if, as I hope, it does). One response to his argument that I do not endorse is to deny the initially plausible "datum" that it is wrong to pass the child by, on the basis of the claim that there are indeed no obligations that are not owed to someone, and you don't owe it to the child, or to anyone else, to rescue the child. Note that there may be room for someone who proposes this response to deny that rescuing the child would be supererogatory. In any case, it is certainly consistent to couple this response with the further claim that failing to rescue the child would be suberogatory.

In the absence of further investigation, then, I am in no position to insist that the category of suberogation is empty. What it is perhaps more important to note in the present context, however, is that, even if this category is empty, it of course remains possible for someone to think that it is not. Consider, then, Sabrina, who *believes* that what she is doing is suberogatory. Is she to blame for her behavior? There are three possible cases: first, Sabrina's behavior is in fact perfectly acceptable, that is, it is neither wrong nor suberogatory; second, her behavior is in fact suberogatory (and so not wrong); and third, her behavior is in fact wrong (and so not suberogatory). If accuses are impossible, then Sabrina is not to blame in either of the first two cases. If the latest version of the Argument from Ignorance is sound, then she is not to blame in the third case, either, unless her behavior is a consequence of some earlier episode of witting wrongdoing on her part (and let's assume that it isn't). But doesn't Sabrina, in all three cases, exhibit an objectionable quality of will, and shouldn't this suffice for her being blameworthy for her behavior (given that she was in control of it)?

[76] For a classic discussion of the relation between claim-rights, liberty-rights (and still other kinds of rights), and obligations, see Hohfeld (1919).

[77] Singer (1972), p. 231.

I confess to being unsure how best to respond to this question. One option is to answer "Yes," in which case the Argument from Ignorance needs still further revision. Another option is to answer "No," on the grounds that the quality of will in question, though morally objectionable, is not of the kind to render Sabrina morally blameworthy.[78] I am not prepared to endorse the second response, and so I concede that the Argument may need still further revision. However, since I suspect that the number of actual cases in which an agent believes that she is acting suberogatorily is likely to be very low, the issue strikes me as lacking any practical urgency. For this reason, I won't undertake any full-scale revision of the Argument in order to accommodate cases like Sabrina's. I simply ask you to keep in mind that it may be that every instance of "overall morally wrong" in both the Argument and the Origination Thesis should be replaced by "either overall morally wrong or suberogatory."

[78] Cf. Duggan (2018), pp. 299 ff., where it is argued that one can be "guiltworthy" but not blameworthy for suberogatory behavior.