

BLAMEWORTHINESS WITHOUT WRONGDOING

Final version published in *Pacific Philosophical Quarterly* 93: 417-437

Abstract: Here I argue that, contrary to a seemingly plausible and widely accepted thesis, it is possible to be blameworthy for doing something that was not objectively morally wrong. If I am right, this would have implications for several debates at the intersection of metaphysics and moral philosophy. I then float an alternative view about which actions can serve as legitimate bases for blame that allows for the possibility of blameworthiness without objective wrongdoing and also suggests an explanation for the initial appeal of the commonly held view that blameworthiness requires objective wrongdoing.

Suppose a person freely did something that, under normal circumstances at least, would be considered morally wrong—she shot and killed a man, for example. Ordinarily we would be inclined to blame this person for what she did. But what if she can justify her behavior? What if she can show that, despite appearances, her actions were not in fact morally impermissible? For instance, what if the man was about to torture and kill her daughter and the only way she could have prevented him from doing this was to shoot and kill him? In that case, her behavior would indeed appear to be morally justified. But would that fact suffice to absolve her of blame for what she did? Does it get her off the hook? In what follows, I will argue that the answer to these questions is *no*; the fact that a person can show that her actions were morally permissible, I contend, does not automatically absolve her of blame for what she did. If I am right, then contrary to a commonly accepted, initially appealing and theoretically influential thesis, the

details of which will be spelled out presently, it is possible for a person to be blameworthy for her behavior even if what she did was not objectively morally wrong.

I

Let us say that it was objectively morally wrong for an agent *S* to perform an action *A* if and only if *S* was morally required all things considered not to *A*.¹ According to what Ishtiyaque Haji has dubbed the Objective View of blameworthiness, necessarily, if *S* is blameworthy for *A*-ing, then it was objectively morally wrong for *S* to *A*.² The Objective View is assumed in much recent work on moral responsibility, and has been put to a variety of different uses by its proponents. Despite its popularity, however, few attempts have been made to defend the thesis or to provide a rationale for it—for the most part, it has simply been taken for granted. But the Objective View is not beyond dispute, and in recent years a handful of arguments against it have emerged. Below I highlight some problems with several of these arguments. Seeing where they go wrong will point the way to an improved argument against the Objective View. I then float an alternative view about which actions can serve as legitimate bases for blame that seems to allow for the possibility of blameworthiness without objective wrongdoing and that also suggests an explanation for, among other things, why the Objective View is initially so appealing. Before doing any of this, though, let me first say a bit more about the Objective View and its role in several disputes at the intersection of metaphysics and the theory of responsibility.

The Objective View has been put to a number of different, and sometimes opposing, uses in recent years. For instance, it figures prominently in this argument for the incompatibility of determinism (the thesis that, at any instant *t*, only one future is compatible with the state of the universe at *t* and the laws of nature) and blameworthiness:

1. Necessarily, if *S* *A*-ed and determinism is true, then *S* could not have avoided *A*-ing (assumption: determinism precludes freedom to do otherwise).
2. Necessarily, if *S* could not have avoided *A*-ing, then *S* was not morally required all things considered not to *A* (from the principle that 'ought' implies 'can').
3. Necessarily, if *S* was not morally required all things considered not to *A*, then it was not objectively wrong for *S* to *A* (true by definition).
4. Necessarily, if it was not objectively wrong for *S* to *A*, then *S* is not blameworthy for *A*-ing (from the Objective View).
5. Therefore, necessarily, if *S* *A*-ed and determinism is true, then *S* is not blameworthy for *A*-ing (from 1-4).³

It has also been used to defend the following version of the much debated principle of alternative possibilities (PAP): *S* is blameworthy for *A*-ing only if *S* could have avoided *A*-ing. Here is the argument:

1. *S* is blameworthy for *A*-ing only if it was objectively morally wrong for *S* to *A* (from the Objective View).
2. It was objectively morally wrong for *S* to *A* only if *S* was morally required all things considered not to *A* (true by definition).
3. *S* was morally required all things considered not to *A* only if *S* could have avoided *A*-ing (from the deontic maxim that 'ought' implies 'can').
4. Therefore, PAP is true (from 1-3).⁴

Some readers may have noticed that both of the preceding arguments appeal to the venerable deontic maxim that ‘ought’ implies ‘can’, a version of which says, necessarily, for any action or omission φ , if S ought to φ , then S can (is able, has it within his power to) φ .⁵ Interestingly, John Martin Fischer appeals to the Objective View as part of an argument *against* this maxim. Fischer’s argument begins with a case in which Sally opts not to raise her hand even though she knows that doing so would alert a lifeguard to a drowning child. Unbeknownst to Sally, however, she is temporarily paralyzed and thus cannot raise her hand. In Fischer’s opinion,

[Sally] is morally responsible (and blameworthy) for not raising her hand, even though she could not have raised her hand. Further, since she is blameworthy for not raising her hand, I would claim that she acted wrongly in failing to raise her hand, and thus that she ought to have raised it. But she could not have raised it. Thus, ought does *not* imply can.⁶

Fischer’s argument here seems to be this:

1. Sally is blameworthy for not raising her hand (this is supposed to be intuitive).
2. Therefore, it was objectively morally wrong for her not to raise it (from 1 and the Objective View).
3. If it was objectively morally wrong for Sally not to raise her hand, then she ought to have raised it (true by definition).
4. Therefore, Sally ought to have raised her hand (from 1-3).
5. Sally couldn’t have raised her hand (true by hypothesis).

6. Therefore, ought does not imply can (from 4 and 5).

If the argument is sound, then not only is the ‘ought’ implies ‘can’ maxim false, but both of the previous arguments, the one for the incompatibility of determinism and blameworthiness and the one in defense of PAP, are unsound, as both arguments depend crucially on that maxim. And if, as I will argue, the Objective View is false, then all three arguments are unsound.

A successful argument against the Objective View would also undermine a popular view about the epistemic dimension of blameworthiness. According to the view in question, barring culpable ignorance, *S* is blameworthy for *A*-ing only if *S* knew, or at least was aware of the fact, that he was morally required not to *A*. This is a version, *V*, of what George Sher has recently dubbed the Searchlight View of moral responsibility.⁷ Now, because knowing or being aware that *p* entails *p*, it follows from *V* that a person is blameworthy for what he did only if the action in question was objectively morally wrong. In other words, *V* entails the Objective View. So if the Objective View is false, then contrary to what *V* asserts, it is possible, even in cases in which culpable ignorance is not an issue, for *S* to be blameworthy for *A*-ing even though *S* didn’t know, and was not aware of the fact, that *A*-ing was morally impermissible.

Evidently, then, a lot hinges on whether the Objective View is true.

The Objective View is initially quite appealing. No doubt this is due in part to the fact that in most ordinary cases of blameworthiness the action for which the agent is blameworthy is in fact objectively wrong. Blameworthiness and objective wrongdoing thus typically go hand in hand. Later I offer an explanation for why this is so, one that is consistent with the possibility of being blameworthy for an action that was not objectively wrong. For now, I would simply like to point out that if we focus only on the ordinary cases, the Objective View will seem quite

attractive indeed. But we should not ignore the possibility that there are atypical cases in which the agent is blameworthy for a particular action of hers despite the fact that that action was not objectively wrong. In what follows, I examine several cases that allegedly fit this bill.

II

Imagine an insidious agent who does something in an effort to harm another person but inadvertently ends up doing something tremendously beneficial for the person instead. Cases like this, it is sometimes claimed, are counterexamples to the Objective View. Surely, one might think, such a person deserves blame for what she did, even if what she did was not, strictly speaking, morally impermissible. If this is right, then the Objective View should be jettisoned.

An argument along these lines is advanced by Haji.⁸ Central to it is a case involving a doctor—Dr. Deadly—who tries to kill her patient by giving him drug C, which she mistakenly believes to be a deadly poison. In fact, however, drug C is actually the cure for her patient’s malady. Following Haji, let us refer to this case as ‘Deadly’s Defeat’. Haji claims that Deadly is blameworthy for injecting her patient with drug C, but that, contrary to what the Objective View implies, it was not objectively wrong for her to do so. After all, doctors are morally obliged to cure their patients; hence, by giving the patient drug C and thereby curing him, it seems that Deadly unintentionally did exactly what she was required to do.

For the sake of argument, let us grant that it was morally permissible, and thus not objectively wrong, for Deadly to administer drug C to her patient. The question then becomes whether she is really blameworthy for doing so, as Haji claims. Later in the paper, as I mentioned, I will propose an alternative view about which actions can be legitimate bases for blame, one that seemingly allows for the possibility of an agent’s being blameworthy for *A*-ing

even if it was not objectively wrong for the agent to *A*. Depending on how the details of Haji's example are spelled out, this alternative view implies that Deadly may indeed be blameworthy for giving the patient drug C, regardless of whether doing so was objectively wrong. Nevertheless, in the absence of that view or any additional argument in support of the claim that Deadly is blameworthy for giving the patient drug C, I do not think cases of this sort should compel proponents of the Objective View to abandon their position. My aim in the remainder of the section is to explain why I think this. Identifying the shortcomings in Haji's argument will also suggest ways in which the case against the Objective View might be improved.

An obvious reply to Haji by defenders of the Objective View (who grant that it was not objectively wrong for Deadly to give the patient drug C) would be to insist that what Deadly is really blameworthy for in this case is *trying* to kill the patient. But arguably she shouldn't have tried to do that. It is plausible that Deadly was morally required not to try to kill her patient and, consequently, that it was objectively wrong for her to try to kill him. If so, a proponent of the Objective View who grants that it was morally permissible for Deadly to inject the patient with drug C can seemingly account for the intuition that Deadly is blameworthy by pointing out that Deadly, although not blameworthy for giving the patient drug C, as it was not wrong for her to do that, is blameworthy for trying to kill the patient, as it was wrong for her to do that. If this is right, then Deadly's Defeat would not be a counterexample to the Objective View.

Haji considers this sort of response, but finds it unpersuasive. According to him, it "appears to rest on the dubious principle that it is always the case that if one attempts to do something, and that thing is wrong, then one's attempt to do that thing is also wrong." But Haji maintains that Deadly's Defeat is a counterexample to that principle as well, for although it was wrong for Deadly to kill the patient, "Deadly's attempt to kill the patient," he says, "(where the

attempt consists, at least partly, in Deadly's injecting the patient with medicine C) does not seem to be wrong."⁹ And, of course, if attempting to kill the patient was not wrong, then proponents of the Objective View could not consistently claim that Deadly is blameworthy for attempting to kill him without abandoning their View.

Haji is probably right to reject the principle he mentions. However, we needn't explore the merits (or lack thereof) of that principle further, for Haji is mistaken to suppose that the present objection turns on it. Proponents of the Objective View need not invoke the principle in defense of their claim that it was objectively wrong for Deadly to try to kill the patient; their claim is plausible independently of the principle. Showing that the principle is false therefore would not successfully defuse the objection at issue. But what about Haji's contention that Deadly's attempt to kill the patient, insofar as it consists at least partly in her giving him drug C, is not objectively wrong? Brief attention to some basic issues in action theory will help facilitate an evaluation of that claim.

Consider an example of Donald Davidson's.¹⁰ Don flips a switch, turns on the light, and illuminates the room. Unbeknownst to him, he also alerts a prowler to the fact that he is home. How many actions did Don perform? Well, that depends on who you ask. Those like Davidson who favor a coarse-grained theory of action individuation would say that Don performed one action which can be given four different descriptions, whereas those who favor a fine-grained theory, according to which *A* and *B* are distinct actions if and only if they exemplify different act-properties, would say that Don has performed at least four actions, as *flipping a switch*, *turning on a light*, *illuminating a room* and *alerting a prowler* are distinct act-properties. Those who prefer the coarse-grained view should henceforth read the action variable '*A*' as a variable

for an action under a particular description, and those who prefer the fine-grained approach should read it as a variable for actions themselves.

It is important to notice that when actions are individuated coarsely, it is actions under descriptions that are the targets of various act-evaluations. For instance, whether an action is intentional or unintentional and whether it is done for a reason or not will depend on how it is described. Under the description *turning on the light*, Don's action is both intentional and, in ordinary cases at least, would be done for a reason, but under the description *alerting the prowler*, the action is neither intentional nor is it done for a reason. Similarly, whether an action is objectively wrong and whether the agent is blameworthy for performing it will depend, among other things, on how the action is described.

Bearing these things in mind, return to Haji's claim that "Deadly's attempt to kill the patient (where the attempt consists, at least partly, in Deadly's injecting the patient with medicine C) does not seem to be wrong." Now if that claim is true, it will not do for proponents of the Objective View to insist that Deadly is blameworthy for attempting to kill her patient. However, it seems to me that Haji's claim is false and, accordingly, that defenders of the Objective View can plausibly claim that Deadly's attempt to kill the patient was both objectively wrong and something for which she may be blameworthy. Their claim can be defended on both the coarse and fine-grained approaches to action individuation adumbrated above.

On the coarse-grained approach, whether Deadly is blameworthy for what she did will depend, in part, on how her action is described. Under some descriptions, her action was not objectively wrong and so, according to the Objective View, is not something for which she is blameworthy. Under other descriptions, however, her action was objectively wrong and therefore is something for which she may be to blame. For instance, under the description *giving the*

patient drug C, her action was (we are supposing) morally permissible and so, according to the Objective View, is not something for which she is blameworthy. But under the description *attempting to kill the patient*, her action does seem to be objectively wrong, contrary to what Haji claims, and therefore is something for which she may be to blame.

On the fine-grained approach, wrongness and blameworthiness—like intentional and done for a reason—attach to particular actions rather than to actions under descriptions. But notice that, on this approach, attempting to kill the patient and giving him drug C will turn out to be two different actions, as *attempting to kill a patient* and *giving a patient a drug* are distinct act-properties. A doctor can attempt to kill a patient without giving the patient a drug, and can give the patient a drug without attempting to kill him. It seems, then, that proponents of the Objective View who favor the fine-grained theory can plausibly maintain that injecting the patient with drug C was not objectively wrong and so is not something for which she is blameworthy, but that the distinct action of attempting to kill the patient was objectively wrong and so is something for which she may be to blame.

Haji offers yet another argument in defense of his claim that Deadly's attempt to kill the patient was not objectively wrong that, if successful, would effectively undermine the objection to Haji's argument presently being explored. The argument is based, in part, on the following deontic principle: if *S* cannot *A* without *B*-ing, and *S* can refrain from *B*-ing, then if *S* ought to *A*, then *S* ought to *B*. He claims that "if we grant that injecting the drug is obligatory for Deadly, then, if Deadly's attempt to kill the patient consists (at least partly) in Deadly's injecting the patient with C, the attempt itself...must...be obligatory," given the principle just identified.¹¹ But again, if attempting to kill the patient was morally obligatory for Deadly, as Haji claims, then proponents of the Objective View cannot say that she is blameworthy for attempting to kill him.

Let us grant that it was obligatory, and thus not objectively wrong, for Deadly to give the patient drug C, and let us also grant the deontic principle to which Haji adverts. Does it follow from that principle that trying to kill the patient is morally obligatory for Deadly? It's not clear that it does, for why suppose the antecedent of the principle is satisfied in this case? That is, why suppose that Deadly cannot inject the patient with drug C without also trying to kill him? Haji says: "it seems that Deadly cannot [intentionally] inject the patient without attempting to inject the patient."¹² That's right, but all that follows is that attempting to inject the patient with drug C is obligatory. It does not follow that attempting to kill the patient is obligatory. Again, to reach that conclusion Haji would need the following claim to be true: Deadly cannot (intentionally) inject the patient with C without trying to kill him. However, I see no reason whatsoever to accept that claim and, moreover, it seems straightforwardly false. It won't help to say, as Haji does in several places, that Deadly's attempt to kill the patient consists at least partly in his injecting her with drug C. That too is true, but it does not provide us with any reason to suppose that Deadly cannot (intentionally) inject the patient with C without trying to kill him.

It is open to proponents of the Objective View, I have suggested, to claim that Deadly is not blameworthy for injecting the patient with drug C (on the assumption that it was not objectively wrong for her to do that), though she may be blameworthy for *trying* to kill the patient, as it was arguably objectively wrong for her to do that. Of course, it is entirely possible for a person to be blameworthy for more than one event or state of affairs at the same time. So it is possible that Deadly is to blame both for giving the patient drug C and for trying to kill him. However, in the absence of any further reasons to think that Deadly is blameworthy for both actions (or for the action under both descriptions), Haji's example is unlikely to persuade proponents the Objective View to jettison their thesis, as they can seemingly account for the

intuition that Deadly is blameworthy without having to concede that she is blameworthy for an action that was not objectively wrong.

Deadly, let us suppose, freely gave the patient drug C and did this despite believing that she was morally required not to. Given these additional details, perhaps Haji would attempt to defend his claim that Deadly is blameworthy for giving the patient drug C despite the fact (let us continue to assume it is a fact) that it was not objectively wrong for her to do so by appealing to the following thesis (or something like it): in the absence of any serious monkey business (e.g., manipulation by nefarious neurosurgeons and the like), *S* is blameworthy for *A*-ing if *S* freely *A*-ed despite believing that it was objectively wrong to *A*. But then Haji's argument would not differ importantly from another argument against the Objective View to be discussed presently. That would not be particularly problematic if the argument in question were sound, but it isn't.

III

According to Michael Zimmerman, cases in which a person freely does something she mistakenly believes to be objectively wrong are counterexamples to the Objective View. This is because, on his view, "acting freely in the belief that one is doing objective moral wrong is sufficient for being morally blameworthy for one's behavior," but not for doing something objectively wrong.¹³ If this is right, then contrary to what the Objective View asserts, there can indeed be cases of blameworthiness without objective wrongdoing.¹⁴

Zimmerman's argument against the Objective View relies on two assumptions. The first is that freely performing an action in light of a belief that the action is objectively wrong does not suffice for objective wrongdoing. The second is that freely doing something you believe to be

objectively wrong is sufficient for being blameworthy for your behavior. The first assumption is plausible and Zimmerman's defense of it is persuasive, but the second merits further attention.

Consider the following story.¹⁵ It's 1943, and Corrie is hiding Jewish refugees in her home in Haarlem, Holland. One day a group of German soldiers ask her if she is harboring any Jews. Now, Corrie takes the Ten Commandments very seriously and, like Kant, believes that lying is always morally impermissible, no matter what. However, she also believes that it would be objectively wrong to tell the soldiers the truth, for by doing so she would effectively be handing over the refugees to be tortured and killed, which, of course, would be wrong. She doesn't regard remaining silent as morally permissible either, for if she keeps silent, she believes that the officers will get suspicious, raid the house, find the refugees and whisk them all, herself and her family included, off to concentration camps where they would almost certainly perish. In short, no matter what Corrie does, she will do something she believes to be objectively morally wrong. In the end, Corrie freely lies to the soldiers, despite believing that she is morally obliged not to lie to them. She does this, however, solely to prevent herself, her family and those she is hiding from being unjustly imprisoned and killed. But if there had been a course of action that she regarded as morally permissible, she most certainly would have opted for it.

Is Corrie blameworthy for lying to the soldiers? Because she freely lied to them, despite believing that it was objectively morally wrong for her to do so, Zimmerman's second assumption—*viz.*, that freely doing something in light of a belief that it is objectively wrong is sufficient for blameworthiness—implies that she is. However, given the details of the case, the claim that Corrie is blameworthy seems extremely implausible.

According to Zimmerman's preferred view of blameworthiness, if a person is blameworthy for her conduct, the behavior in question reflects ill on her and serves as a basis for

a certain kind of negative moral judgment or appraisal of the person, at least with respect to that particular episode of her life.¹⁶ As he puts it elsewhere, it's as if there is debit in the person's "moral ledger," a black mark on her moral record.¹⁷ But Corrie's behavior in this case arguably does not reflect ill on her, nor, it seems, does she merit any kind of negative moral appraisal on the basis of what she did. No bad marks should be entered on her moral report card. If this is right, then contrary to what Zimmerman's second assumption implies, Corrie is not blameworthy for lying.

This conclusion is further supported by reflection on the relationship between blameworthiness and certain negative reactive attitudes. Someone who is blameworthy is, in principle at least, the appropriate object of reactive attitudes such as resentment, indignation and guilt. However, Corrie is arguably not the appropriate object of these 'blaming emotions'. No one has cause to resent Corrie or be indignant toward her for what she did, nor should she feel guilty about her behavior.¹⁸ This isn't to say that it would be inappropriate for Corrie to feel bad about what happened. Perhaps she might reasonably regret or find unsavory having to lie in order to protect innocent lives, but given the details of her situation, she surely has no reason to reprimand, censure or blame herself, nor, it seems, does anyone else.

Evidently, then, Corrie is not blameworthy for lying to the soldiers. Since Zimmerman's second assumption implies that she is, that assumption should be rejected; freely performing an action in light of a belief that one is thereby doing something objectively morally wrong does not suffice for blameworthiness. Zimmerman has therefore failed to provide sufficient conditions for blameworthiness, and so, *a fortiori*, has failed to provide sufficient conditions for blameworthiness that are not also sufficient for objective moral wrongdoing. Zimmerman's argument against the Objective View is therefore unsuccessful.

A closely related argument is advanced by Haji. Central to it is the idea of *suberogatory* action. An action is said to be suberogatory provided that it is not objectively wrong but is nevertheless morally untoward in some respect.¹⁹ According to Haji, people can be blameworthy for performing suberogatory actions. If he is right, then contrary to what the Objective View implies, a person may indeed be blameworthy for what he did even if he didn't do anything objectively wrong.²⁰ Let's take a closer look at Haji's argument.

Commenting on Julia Driver's example of a man who, being first in line, takes a more convenient seat on a train and thereby deliberately prevents two lovers from sitting together when he could have easily taken a less convenient seat, Haji says that if we assume that the man "believes he is doing wrong by preventing the lovers from sitting together despite the availability of another seat," the man "may be blameworthy for not giving up his seat even though, let's assume, his not giving up his seat is suberogatory and hence morally optional."²¹ Suppose we grant Haji all this. Even so, the fact that the man's action is suberogatory doesn't seem to play a crucial role in Haji's judgment that the man is blameworthy. Rather, the reason Haji thinks the man is to blame for what he did seems to be that, by refusing to give up his seat thereby preventing the lovers from sitting together, the man deliberately did something he mistakenly believed to be morally wrong. The fact—if it is indeed a fact—that the man did something suberogatory only seems to be relevant insofar as it shows that the man was mistaken to think that his action was wrong. But whether he was mistaken because the action was in fact suberogatory, as Haji claims, or because it was merely permissible but not suberogatory seems to be entirely beside the point. What is essential to Haji's argument, it seems, is that the man was mistaken in thinking the action was morally wrong but deliberately performed that action anyway.

Haji provides a similar analysis about a case in which “Joy is aware that others are waiting for her table in the crowded café, mistakenly believes she is doing wrong by lingering on, intentionally lingers on *in light of the belief that she is doing wrong*, but really suberogates by lingering on.”²² Let us grant (though it is perhaps debatable) that Joy did not do anything objectively wrong by lingering on. Haji insists that she is blameworthy for doing so nonetheless, and again his reason seems to be that she deliberately did something she (mistakenly) believed to be wrong. The fact—if it is indeed a fact—that lingering was suberogatory does not appear to be essential to the argument. What matters is that Joy was mistaken in thinking that she was doing something wrong by lingering but is blameworthy for lingering nonetheless. Precisely why she is mistaken, whether because lingering was indeed suberogatory, as Haji claims, or because doing so was merely permissible but not suberogatory, appears to be irrelevant.

So in the end it’s not clear that Haji’s appeal to suberogation is really doing any heavy lifting in his argument against the Objective View. The key thought seems to be, rather, that if someone deliberately performed an action in light of a belief that that action was objectively morally wrong, the person may be blameworthy for what she did, even if her belief that the action in question was morally impermissible is false. Insofar as this is the main thrust of Haji’s argument, it is quite similar to Zimmerman’s and fails for much the same reasons. It may be that the agents in the examples Haji adduces are blameworthy for what they did, and it may even be true that they did not do anything objectively wrong (though, again, this last claim is questionable). However, if they are blameworthy, this is not simply because they freely did something they believed to be wrong. Freely doing something one believes to be wrong, we have seen, does not suffice for blameworthiness.²³ Evidently, something further is required.

IV

At this point, a natural question to ask is this: if freely performing an action in light of a belief that the action is objectively morally wrong does not suffice for being blameworthy for performing that action, what else is required? Here is a suggestion: the action must manifest a *morally objectionable quality of will* on the part of its agent. My aim in this section and the next is to develop this suggestion further and to highlight various uses that can be made of it. In this section, after a few clarificatory remarks, I offer some support for the suggestion and employ it in an improved argument against the Objective View. In the next, I contend that it points to an explanation for, among other things, why the Objective View is initially so attractive.

In talking about the quality of agents' *wills*, I am not referring to any mystical power or faculty of mind with which agents are imbued, but rather to the moral quality of various mental items that are implicated in the production of voluntary, intentional actions. Thus, let us say that an agent has a morally objectionable quality of will just in case her beliefs, desires, values, intentions, plans, etc. exhibit a morally unjustifiable degree of ill will, indifference or lack of due regard for others or for the moral considerations that bear upon her situation. The idea of an action *manifesting*—or, as I shall sometimes put it, *expressing*—a certain quality of will has an essential causal element. An action *A* manifests (or expresses) a morally objectionable quality of will only if the attitudes in which the objectionable quality of will consists are among the non-deviant causes of *A*.

Stated a bit more precisely, my suggestion, again, is this: an agent who freely *A*-ed despite believing at the time that it was objectively morally wrong to *A* is blameworthy for *A*-ing

if and only if *A* expresses a morally objectionable quality of will on the part of the agent. The plausibility of this suggestion can be illustrated with an example.¹

Beatrix freely shot and killed Bill. She did this despite believing that it was objectively wrong to kill Bill, that it was within her power to avoid killing him and, indeed, that it was within her power to avoid wrongdoing altogether. Why, then, did Beatrix kill Bill? What motivated her to perform this action? A significant role was played by her hatred of Bill and her (no doubt morally unjustified) desire to rid the world of him. (The two used to be an item before Beatrix caught Bill with another woman. It was this unfortunate event that spawned Beatrix's antipathy for Bill, which she has subsequently nursed over the years.) Unbeknownst to Beatrix, however, Bill was just about to torture and kill her daughter, and the only way she could have prevented him from doing this was to shoot and kill him. Call this case 'Kill Bill'.²⁴

Compare Beatrix with Corrie, who we encountered in a previous example. Both freely did something they believed to be objectively wrong. There is, however, an important difference between them and the actions they performed, *viz.*, Beatrix's action manifested a morally objectionable quality of will, whereas Corrie's action did not. When Corrie lied to the soldiers, she did this freely despite believing that it was objectively wrong, but only to save herself, her family and the Jewish refugees she was harboring from unjust imprisonment and death. She did not express any ill will or lack of due regard for anyone, nor did she show disdain for any of the moral considerations that were relevant to her decision. She wanted to do the right thing but, from her standpoint at least, she couldn't help doing something she regarded as wrong, for through no fault of her own she believed that all her options were morally off limits. If she had believed that there was a morally permissible alternative available to her, she would have opted for it. Corrie's behavior thus does not seem to express an objectionable quality of will, which,

1

one might plausibly claim, is precisely why she is not to blame for what she did. Not so with Beatrix, however. When she shot and killed Bill she did so out of ill will for him, despite believing that it was objectively wrong to kill him, all the while believing that there was a morally permissible course of action available to her. Her behavior therefore does appear to express an objectionable quality of will and, accordingly, seems to be something for which she may be to blame.

Recall that according to the view of moral blameworthiness advocated by Zimmerman, a person is blameworthy for her conduct just in case the behavior in question serves as a basis for a certain kind of negative moral judgment or appraisal of the person, at least with respect to that particular episode of her life. As he and others sometimes like to put it, when a person is blameworthy for *A*-ing, it's as if there is debit in the person's moral ledger, a black mark on her moral record, in virtue of her having *A*-ed. Now, it certainly seems as if Beatrix merits a negative moral appraisal for killing Bill. This action of hers expressed her ill will for Bill and her lack of due regard for his wellbeing, and she performed it freely despite believing at the time that the action was objectively wrong and that it was within her power to avoid wrongdoing. Someone who freely *A*-ed as a result of her ill-will and lack of due regard for the wellbeing of another, despite believing that it was objectively wrong to *A* and that it was within her power to avoid wrongdoing entirely, arguably merits moral criticism for *A*-ing. Such a person deserves a bad mark on her moral report card for having *A*-ed. If meriting such a black mark suffices for blameworthiness, as Zimmerman and others insist, then Beatrix is blameworthy for killing Bill.

That Beatrix is blameworthy for killing Bill is further supported by reflection on the relationship between blameworthiness and the blaming emotions. It seems that Beatrix is the appropriate object of at least some of those emotions for having freely killed Bill, given her own

attitudes and beliefs about that action at the time. For instance, it seems that it would be appropriate, at least in principle, though not necessarily all things considered, for us to be angry with and feel indignation toward Beatrix for killing Bill, insofar as she acted out of unjustified ill will for him, thereby displaying a lack of due regard for his wellbeing, despite believing that it was wrong to kill him and that she could have avoided wrongdoing altogether. For similar reasons, it would be appropriate for Beatrix to feel guilty for killing Bill. But being the appropriate object of these blaming attitudes for having *A*-ed arguably suffices for blameworthiness for *A*-ing. Hence, it seems that Beatrix is indeed blameworthy for killing Bill. No doubt we might be glad that she did what she did, for if she hadn't, her innocent daughter would have been tortured and killed. Similarly, if Beatrix were to discover that, by killing Bill, she had unwittingly saved her daughter's life, she too might be glad that she killed him. But all this is compatible with her being the appropriate target of the blaming emotions and thus with her being blameworthy.²⁵

It seems, then, that we have good reason to say that Beatrix is blameworthy for killing Bill. Contrary to what the Objective View implies, however, it was arguably not objectively morally wrong for her to kill him. This follows from a plausible principle governing killing in self-defense and in defense of the innocent. Define an *unjust aggressor* as someone who is trying, without good reason, to kill you or those, such as your children, whom you have a duty to protect, and is blameworthy for doing so. Now consider the following Principle of Self-Defense (PSD): it is morally permissible to kill an unjust aggressor, if doing so is the only way to save your own life or the lives of those you have a duty to protect, provided that you can do so without doing something else that you are morally required not to do. This principle, or

something like it, is intuitively appealing and widely accepted, and given a pair of further assumptions, implies that it was not objectively wrong for Beatrix to kill Bill.

Assume that Bill was an unjust aggressor; he was trying, without good reason, to torture and kill Beatrix's daughter and, moreover, was blameworthy for doing so. Assume also that killing Bill did not require Beatrix to do anything else that she was morally required not to do—killing him did not require that she also kill an innocent bystander, for example. Given these two additional assumptions, PSD implies that it was morally permissible, and thus not objectively morally wrong, for Beatrix to kill Bill, since doing so was, by hypothesis, the only way to prevent him from torturing and killing her child.²⁶

It appears, then, that we have a genuine case of blameworthiness without objective wrongdoing; Beatrix is blameworthy for killing Bill even though it was not objectively morally wrong for her to do so.

How might proponents of the Objective View handle a case like Kill Bill? Notice that the sort of objection discussed earlier in connection with Haji's example Deadly's Defeat will not work here. Proponents of the Objective View cannot plausibly claim that what Beatrix is really blameworthy for in this case is *trying* to kill Bill. To be sure, Beatrix tried to kill Bill, and (unlike Dr. Deadly) she succeeded in doing what she tried to do. Moreover, she may very well be blameworthy for trying to kill him. However, it was not morally impermissible for Beatrix to try to kill Bill for much the same reasons that it was not impermissible for her to kill him. To see this, consider the following variant of PSD for trying (PSDt): it is morally permissible to *try* to kill an unjust aggressor, if doing so is the only way to save your own life or the lives of those you have a duty to protect, provided that you can do so without doing something else that you are morally required not to do. This principle is no less plausible than the original PSD and, given

the assumptions that Bill was an unjust aggressor and that trying to kill him did not require Beatrix to do anything else she was morally obligated not to do, it entails that it was not objectively wrong for Beatrix to try to kill Bill, as doing so was (let's suppose) necessary to save her daughter's life. So if she is blameworthy for trying to kill him nonetheless, then we have yet another case of an agent being blameworthy for *A*-ing even though it was not objectively wrong for her to *A*.

Readers who are still not persuaded that Beatrix is blameworthy for killing Bill are invited to consider the following modified version of the example and to compare it with the original version. In the modified story, everything is the same as it is in the original version except that Bill was not on the verge of torturing and killing Beatrix's daughter and instead was simply minding his own business. Now, in this version of the case, most people, including most proponents of the Objective View, would readily agree that killing Bill was both objectively wrong and something for which Beatrix is blameworthy. Notice, however, that her behavior in this version of the case manifests no more ill will towards Bill and no more lack of due regard for his wellbeing than it does in the original version. In both versions of the case, Beatrix freely acts contrary to a judgment that she is morally required all things considered not to kill Bill. She does this, moreover, despite believing that it is within her power to avoid wrongdoing and solely because she hates Bill and wants to rid the world of him. Beatrix thus seems no less bad, no less morally criticizable, and no less blameworthy in the modified version of the story than in the original version. Why then should we blame her for killing Bill in the one but not in the other?

If Beatrix is to blame for killing Bill in the modified version of the case, then in the absence of any reason to think otherwise, I should think she is blameworthy for killing him in the original version as well. After all, the *only* difference between the two cases has to do with Bill

and what *he* was planning to do. But how, we might wonder, could *that* difference by itself affect whether Beatrix is blameworthy for her action? Changing that particular aspect of the story alone in no way affects whether Beatrix kills Bill, nor does it affect her morally relevant capacities and powers, nor does it affect her own attitudes towards Bill or the act of killing him. How could something that in no way affects any of these features of the situation be relevant to whether Beatrix is blameworthy for killing Bill? The answer isn't clear.²⁷

Let us proceed on the assumption that Beatrix is blameworthy for killing Bill. I can imagine someone attempting to defend the Objective View by appealing to the following claim: an action is objectively morally wrong if and only if it manifests a morally objectionable quality of will on the part of its agent. On this way of looking at things, anything a person does that reflects an objectionable quality of will on her part is, for her, objectively morally wrong, and any action that is objectively morally wrong for that person will be one that reflects a morally objectionable quality of will. If this is right, then although Beatrix may indeed be blameworthy for killing Bill, this action of hers was morally wrong after all, since it expressed her ill will towards Bill and her lack of due regard for his wellbeing. What are we to make of this objection? An adequate treatment of the theory of wrongdoing on which it turns would require an article of its own. Prima facie, however, there seem to be some fairly powerful reasons to reject it.

Todd accidentally kills a toddler; he backs his car out of his garage, unaware that the child is taking a nap behind the back tire. In an ordinary case of this sort, Todd's action is not indicative of a morally objectionable quality of will. He does not bear any ill will towards the toddler, nor does he fail to show her due regard. It was just a terrible accident. But it hardly follows from any of this that it was not objectively wrong for Todd to kill the toddler. The fact that he is ignorant of the toddler's whereabouts plausibly excuses Todd from blame (assuming,

of course, that he is not culpable for his ignorance), but intuitively it does not make it morally permissible for him to kill her. Evidently, then, a person can do something objectively wrong even though the action in question is not indicative of a morally objectionable quality of will.

There also seem to be situations in which a person's behavior manifests an objectionable quality of will but is not objectively wrong. I have argued that Kill Bill is a case in point. Here is another.²⁸ Fran's leg is gangrenous, and she will die if it is not amputated soon. Years ago Gina promised to amputate the leg if doing so ever became necessary, but upon realizing that Fran will die if the leg is not removed, Gina refuses to perform the operation. Just then the nurse informs Gina that there is no more analgesic and thus no way to numb Fran's pain if Gina decides to cut off the leg. Upon hearing this, Gina immediately changes her mind, decides to cut off Fran's leg as promised, and acts accordingly. However, she does so not to spare Fran's life, but solely to cause Fran excruciating pain. In this case, Gina cuts off Fran's leg, and this action of hers clearly manifests a morally objectionable quality of will, as it was motivated solely by her desire to cause Fran tremendous amounts of pain. But given the circumstances, would we really want to say that it was objectively wrong for Gina to amputate Fran's leg? I should think not. Indeed, in cutting off the leg, it seems that Gina did precisely what she was morally required to do. After all, she promised to amputate the leg if it became necessary to do so, and in the circumstances amputation was necessary, for otherwise Fran would have died.

Having an objectionable quality of will thus seems to be independent of objective wrongdoing; as the preceding examples illustrate, it seems that a person can do something objectively wrong without the action in question manifesting an objectionable quality of will, and that a person can do something that manifests an objectionable quality of will even though the action in question is, in fact, morally permissible.

A few loose ends need tying up.

I began the preceding section with the following question: if freely *A*-ing in light of a belief that it is objectively morally wrong to *A* does not suffice for blameworthiness for *A*-ing, what else is required? My answer, once more, was that *A* must express a morally objectionable quality of will on the part of its agent. This answer entails the following thesis about which actions can serve as legitimate bases for moral blame: *S* is blameworthy for *A*-ing only if *A* manifests a morally objectionable quality of will on the part of *S*. Call this the Quality of Will thesis. According to it, the only actions for which a person is to blame are those that express a morally objectionable degree of ill will, indifference or lack of due regard for others or for relevant moral considerations that bear on the situation.

It is important to keep in mind that the Quality of Will thesis only states a necessary condition for blameworthiness. Taking their cue from P.F. Strawson's justly celebrated essay 'Freedom and Resentment'²⁹, a number of theorists have stressed the importance of the quality of an agent's will when it comes to assessing the agent's moral responsibility for her behavior. It is open to these theorists to make the stronger claim that whether a morally competent individual is blameworthy for something she did is to be settled *entirely* by whether the action in question reflects a morally objectionable quality of will. That is not the claim being made here, however. One can stress the importance of the quality of an agent's will in assessing her moral responsibility without making the stronger claim.

In discussing Haji's example Deadly's Defeat I mentioned that, depending on how the details of that example are spelled out, it may very well be the case that Deadly is blameworthy for giving her patient drug C, as Haji claims, even though it may not have been objectively

wrong for her to do so. We are now in a position to see why. Recall that Deadly mistakenly believed that drug C would kill her patient when, in fact, it was the cure for the patient's malady, and that Deadly injected the patient with drug C in an attempt to kill the patient. Deadly's action was evidently a manifestation of her ill will for the patient and of her lack of due regard for his wellbeing and thus, according to the Quality of Will thesis being advanced here, is something for which she may legitimately be to blame, provided of course that she satisfies any further requirements for blameworthiness. And this is so, it seems, even if, as Haji believes, it was not objectively wrong for Deadly to give the patient drug C.

If the Objective View is indeed false, as I and others have argued, why then does it initially seem so appealing? An answer is suggested by the Quality of Will thesis. Earlier I said that the Objective View is appealing in part because "in most ordinary cases of blameworthiness the action for which the agent is blameworthy is indeed objectively wrong." The Quality of Will thesis points to an explanation for why this is so. In most ordinary cases of objective wrongdoing, the action in question typically manifests a morally objectionable quality of will on the part of the agent, whereas actions that are not objectively wrong typically do not reflect any such objectionable quality of will. Hence, actions that are objectively wrong are ordinarily legitimate bases for blaming someone, according to the Quality of Will thesis, whereas actions that are not objectively wrong typically are not. Now, as I indicated earlier, if we focus only on these ordinary cases, the Objective View can seem quite appealing. However, as cases like Kill Bill seem to illustrate, because an action can be morally permissible and yet still manifest an objectionable quality of will on the part of its agent, that action can be something for which the agent is to blame despite its not being objectively morally wrong.

Recall, finally, Zimmerman's claim that *S* is blameworthy for *A*-ing if *S* freely *A*-ed in light of a belief that *A*-ing is objectively morally wrong. That claim, we have seen, is false. Why, then, are Zimmerman and others tempted to endorse it? The Quality of Will thesis also suggests an answer to this question. Typically when someone does something she believes to be objectively wrong, her action is an expression of her ill will, indifference or lack of due regard for others or for the relevant moral considerations that bear upon her situation. Hence, an action performed freely in light of the belief that it is objectively wrong will, according to the Quality of Will thesis, ordinarily be something for which its agent is to blame. And, if we focus only on these ordinary cases, Zimmerman's claim will seem quite appealing. But again, there can be cases in which a person freely does something in light of a belief that the action in question is objectively wrong and yet the person is not blameworthy for performing that action, which means that Zimmerman's thesis is to be rejected, despite whatever initial appeal it may possess.

VI

I have argued that a person can be blameworthy for *A*-ing even if it was not objectively morally wrong for her to *A*.³⁰ If I am right, then the Objective View is false. I also offered an alternative view about which actions are legitimate bases for blame—the Quality of Will thesis—according to which a person is blameworthy only for actions that manifest an objectionable quality of will. This view allows for the possibility of blameworthiness without objective wrongdoing and also provides an explanation for the initial appeal of both the Objective View and Zimmerman's claim that freely *A*-ing in the belief that it is objectively wrong to *A* suffices for blameworthiness.

The Objective View is certainly alluring. Contrary to what some theorists would have us believe, however, it is hardly an unquestionable axiom in the theory of responsibility. Rather, as

we have seen, it is a substantive assumption that, at the very least, is open to serious prima facie challenges and thus requires further explication and defense by its proponents.³¹

NOTES

¹ Objective wrongdoing is sometimes contrasted with *subjective* wrongdoing. Let us say that it was *subjectively* morally wrong for *S* to *A* if and only if *S* *believed* at the time of action that she was morally required all things considered not to *A*. If, as seems plausible, it is possible for a person to be mistaken about what is morally required of her, then objective and subjective wrongdoing can be pried apart. In what follows, several cases are discussed that seem to illustrate this possibility. The sort of wrongdoing at issue in the main text is objective moral wrongdoing.

² Here I follow Haji, I. (1998). *Moral Appraisability: Puzzles, Proposals and Perplexities*. New York: Oxford University Press, p. 141.

³ An argument along these lines seems to be implicit in van Inwagen, P. (1983). *An Essay on Free Will*. New York: Oxford University Press, 1983, pp. 161-162. However, it is unclear whether van Inwagen would endorse the argument presented here.

⁴ Versions of this argument have been defended by Copp, D. (1997) 'Defending the Principle of Alternative Possibilities: Blameworthiness and Moral Responsibility', *Nous* 31, pp. 441-456; Copp, D. (2003). 'Ought' Implies 'Can', Blameworthiness, and the Principle of Alternate Possibilities' in eds. D. Widerker and M. McKenna (eds) *Moral Responsibility and Alternative Possibilities*. Aldershot: Ashgate Press, pp. 265-300, and by Speak, D. (2005). 'PAPistry: Another Defense', *Midwest Studies in Philosophy* 29, pp. 262-268. The argument was first introduced by Widerker, D. (1991). 'Frankfurt on 'Ought implies Can' and Alternative Possibilities', *Analysis* 51, pp. 222-224.

⁵ The 'ought' here is the 'ought' of all things considered moral requirements.

⁶ Fischer, J.M. (2006). *My Way*. New York: Oxford University Press, pp. 24-25.

⁷ See Sher, G. (2009). *Who Knew? Responsibility Without Awareness*. New York: Oxford University Press.

⁸ Haji 1998, p. 146.

⁹ Ibid. Again, the 'ought' at issue here is the 'ought' of all things considered moral requirements.

¹⁰ The example is from Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press, p. 4.

¹¹ Haji 1998, p. 146.

¹² Ibid.

¹³ Zimmerman, M. (1997). 'A Plea for Accuses', *American Philosophical Quarterly* 34, pp. 229-243.

¹⁴ A similar view is held by Brandt, R. (1958). 'Blameworthiness and Obligation' in ed. A. I. Melden (ed) *Essays in Moral Philosophy*. Seattle: University of Washington Press, and by Swinburne, R. (1989). *Responsibility and Atonement*. New York: Clarendon Press.

¹⁵ This story is inspired by an episode in Ten Boom's, C. (1974) *The Hiding Place*. New York: Bantam Books.

¹⁶ Zimmerman 1997, pp. 229-230.

¹⁷ Zimmerman, M. (1988). *An Essay on Moral Responsibility*. Totowa: Rowan and Littlefield, pp. 38-39. A similar view of moral responsibility is held by Feinberg, J. (1970). *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton, NJ: Princeton University Press, p. 30, and by Glover, J. (1970). *Responsibility*. New York: Humanities Press Inc, p. 64.

¹⁸ To be clear, the claim here is not just that it would be inappropriate for someone to give overt expression to the blaming attitudes, but rather that it would be inappropriate in principle for anyone to have the blaming attitudes towards Corrie on the basis of what she did.

¹⁹ Driver, J. (1992). 'The Suberogatory', *Australasian Journal of Philosophy* 70, pp. 286-295.

²⁰ See Haji, 1998, p. 148, and Haji, I. (2002). *Deontic Morality and Control*. New York: Cambridge University Press, pp. 38-39, 167-182.

²¹ Haji 2002, p. 171.

²² Haji 2002, p. 39; Haji's emphasis.

²³ One difference between Haji's position and Zimmerman's is that Haji maintains that in order for a person to be blameworthy for what he did, the person must satisfy a separate condition of autonomy, in addition to acting freely in the belief that she is doing wrong. However, as far as I can tell, this difference in Haji's position does not insulate it from my criticism of Zimmerman's view. We can safely assume that Corrie satisfies Haji's autonomy condition without having to grant that she is blameworthy for lying.

²⁴ Graham, P.A. (2011). 'Fischer on Blameworthiness and "Ought" Implies "Can"', *Social Theory and Practice* 37, pp. 63-80 discusses cases of this sort and he too argues that they are counterexamples to the Objective View. An important difference between Graham's view and my own is that Graham thinks that performing an action on the basis of an evil motive (e.g., chopping someone's leg off merely to cause the person excruciating pain) suffices "all on its own" to warrant resentment and indignation (p. 77). This I deny.

²⁵ Graham 2011, p. 71 makes a similar point.

²⁶ A number of philosophers insist that what a person is morally required to do at any given moment will depend crucially on the person's evidential situation at the time. For example, if all the person's available evidence indicates that the balance of moral reasons speaks against *A*-ing, then the person would be morally required not to *A*. Someone who holds this view about moral requirements might insist that if Beatrix lacked sufficient evidence that Bill was about to torture and kill her daughter, then the balance of evidence available to her at the time would seem to speak against killing Bill, in which case it was objectively wrong for her to kill him after all. To avoid this sort of objection, I assume that Beatrix's belief that it was impermissible for her to kill Bill was not based on the evidence and, indeed, that her available evidence at the time spoke against that belief.

²⁷ The answer cannot be simply that altering whether Bill was about to torture and kill Beatrix's daughter makes killing him objectively wrong, for that answer would simply beg the question. What we want to know is how the wrongness of the act can make a difference when it in no way affects the agent's morally relevant capacities, powers and attitudes about the action in question.

²⁸ The following case is inspired by one of Graham's 2011.

²⁹ Strawson, P.F. (1962). 'Freedom and Resentment', *Proceedings of the British Academy* 48, pp. 1-25.

³⁰ Keep in mind that this claim does not entail that *S* can be blameworthy for *A*-ing even if it was not wrong *in any respect* for *S* to *A*. For instance, the claim is consistent with variants of the following thesis: *S* is blameworthy for *A*-ing only if it was subjectively wrong for *S* to *A*. On subjective wrongdoing, see note 1.

³¹ Thanks to Michael McKenna, Al Mele, Randy Clarke, Stephen Kearns and a pair of anonymous referees for their helpful feedback on earlier versions of this paper.