Penultimate Draft. Forthcoming in *Philosophical Studies*.

Attributionism and Degrees of Praiseworthiness

Daniel J. Miller

Abstract: An increasingly popular theory of moral responsibility, Attributionism, identifies attitudes as the locus of direct responsibility. And yet, two agents with qualitatively identical attitudes may differ in their responsibility due to a difference in whether they act on those attitudes. On the most plausible interpretation of Attributionism, attitude duplicates differ in their responsibility only with respect to the scope of what they're responsible for: one agent is responsible for only their attitudes, while the other is responsible for their attitudes and for acting in a way that reflects those attitudes. Against this, I argue that attitude duplicates may also differ with respect to their degrees of praiseworthiness, and that this is best explained by either the effort or sacrifice instantiated in one's actions—explanations unavailable to Attributionism. If this is correct, then Attributionism fails to provide an adequate account of praiseworthiness, and therefore fails as a theory of moral responsibility.

Keywords: moral responsibility; praiseworthiness; attributionism; effort; sacrifice; degrees of responsibility.

Introduction

A committed Seventh Day Adventist, Desmond Thomas Doss was a pacificist. And after a formative event in his early life (in which his drunken father drew a gun on his uncle during a heated dispute), Doss resolved never to touch a gun. Doss was a shipyard worker in Virginia when the United States joined World War II, and due to his occupation was allowed to defer enlistment in the military (Berman, 2016). Doss enlisted anyway to serve as a medic in the U.S. Army, so that he could "be like Christ: saving life instead of taking life." During training, Doss was verbally and physically harassed by other soldiers for his pacifist convictions; he was treated as a weak link that couldn't be trusted on the battlefield (Blair, 2016).

Some of those soldiers lived to change their minds. In April 1945, Doss's battalion entered the Battle of Okinawa—one of the bloodiest in the Pacific Theater. At the top of a 400-foot cliff dubbed "Hacksaw Ridge", thousands of heavily armed Japanese soldiers were waiting. On one day of the battle, as American troops advanced, the Japanese concentrated fire on them, killing and injuring scores. The remaining American troops were driven back down the ridge. Doss alone remained with the injured. One at a time, Doss dragged them to the edge of the cliff, lowering them to safety. After each one, Doss prayed, "Lord, please help me get one more" (Lange, 2017). Doss is credited with saving the lives of seventy-five men over the course of twelve hours that day (Blair, 2016).

While treating other soldiers on the battlefield some days later, Doss's leg was seriously injured by a grenade. Rather than calling for a medic (and thus diverting them from aiding others), Doss treated his own wounds. Five hours passed before medics arrived to rescue him. While being carried off the battlefield Doss got off of the stretcher, insisting that the medics take a more severely wounded soldier instead. As Doss made his way back on his own, a sniper bullet shattered his arm (Lange, 2017). Doss crawled 300 yards to an aid station. (Leepson, 2015).

Desmond Doss became the first soldier in U.S. history to be awarded the Medal of Honor without ever firing a shot. Due to his injuries, Doss was unable to return to full time work after the war, and spent the remainder of his life working at his local church (Berman, 2016).

Desmond Doss is extraordinarily praiseworthy.[1] But what exactly is it for which he is praiseworthy? One might say that Doss is praiseworthy primarily for his attitudes, and in particular his willingness to perform life-saving actions in the face of much difficulty, and at great cost to himself.[2] Even if this is so, Doss is surely also praiseworthy for his difficult, sacrificial actions themselves. Though most will agree upon this claim, theorists will find themselves divided over the following question: do Doss's actions increase his degree of praiseworthiness over and above his praiseworthiness for his willingness to perform those actions? In other words, do his actions not only make Doss praiseworthy for more things, but also more praiseworthy?[3]

How one answers this question will depend upon one's commitments concerning what agents are directly (i.e., non-derivatively) responsible for. An increasingly popular theory of moral responsibility, Attributionism, identifies attitudes as the locus of direct responsibility. Yet, if the answer to the above question is to be answered in the affirmative, then responsibility cannot fully be accounted for in terms of responsibility for attitudes. In this paper I argue that Attributionism fails to

---

[1] I set aside general skepticism about moral responsibility.

[2] We might understand Doss's willingness to act in terms of an evaluative judgment that the lives of his compatriots provide a sufficient reason to act as he did, in conjunction with a desire to act on the basis of that reason.

[3] Praiseworthiness may include, but also extends beyond, the appropriateness of moral *assessment* (e.g., a judgement that an agent is virtuous, acted virtuously, displayed a good quality of will in their action). On my understanding, an agent *S* is praiseworthy for something *A* only if certain reactive attitudes (e.g., gratitude, esteem) are *pro tanto* appropriate responses to *S* on the basis of *A*. Notably, this understanding of praiseworthiness is consistent with Attributionists' stated views on the appropriateness of the reactive attitudes involved in praising (Smith 2005: 236-237; 2012: 576; Hieronymi 2012: 9-10), as well as the suggestion that Attributionists and their opponents do not differ as much on the sorts of *responses* that are appropriate for responsible agents as they do on the *conditions* of the appropriateness of those responses (Nelkin 2016: 360; Talbert 2017: 18-19). Lastly, reactive attitudes come in degrees: we can be more or less resentful, indignant, grateful, or esteeming. Accordingly, if *S* is more praiseworthy than *R*, then, minimally, it is *pro tanto* appropriate to have the relevant reactive attitudes in a greater degree toward *S* than toward *R* (e.g., it would be *pro tanto* appropriate to be more grateful to *S* than to *R*).

provide an adequate account of moral praiseworthiness, and therefore fails as a theory of moral responsibility.

1. Attributionism and Attitude Duplicates

Theorists have traditionally maintained that agents can be directly responsible only for actions (and perhaps omissions), and only derivatively responsible for attitudes (e.g., beliefs, desires, emotions). On this traditional view, the scope of direct responsibility is often explained in terms of that over which we can exercise a certain kind of control.[4]

According to its main rival, Attributionism, an agent is morally responsible for some item insofar as it expresses or reflects the agent's underlying attitudes, evaluative judgments, or quality of will (Smith, 2005, 2008; Hieronymi, 2008; 2014; Scanlon, 2008, 2015; Talbert, 2012, 2019).[5] A distinctive feature of this view is the commitment that agents can be directly responsible for attitudes, despite not having voluntary control over them.

But Attributionism in fact goes further, flipping the traditional view on its head entirely. Attributionism doesn't simply maintain that agents can be directly responsible for their attitudes; it also implies that agents are directly responsible *only* for their attitudes. On Attributionism, agents are responsible for actions (or omissions) only derivatively, and insofar as they reflect underlying non-voluntary attitudes (judgments, quality of will). As T. M. Scanlon puts it, "attitudes are the only things a person is responsible for in the most fundamental way. Other things … can be the grounds for reactive attitudes only derivatively, insofar as they reflect the person's normative attitudes" (2015: 98).

Other Attributionists express similar commitments. Angela Smith, understanding responsibility as answerability, offers the following illustrative example: "A cruel person . . . is someone who judges that the fact that something will cause pain or suffering to another is no reason to avoid it . . . *It is that judgment, as reflected in her actions and attitudes, for which we consider her answerable. . .*" (2008: 389-390, emphasis added). Pamela Hieronymi maintains that "[w]e are fundamentally responsible for a thing … because it reveals our take on the world … what we find true or valuable or important. But we cannot enjoy discretion with respect to whether we find something true or valuable or important".

---

[4] Randolph Clarke calls this the "Action View" (n.d.). Related views are sometimes framed in terms of items over which we exercise voluntary control, though there are various conceptions of what this amounts to (Fritz, 2018).

[5] These theorists' views are sometimes referred to under different names. Angela Smith (2005), for example, refers to her own view as the "Rational Relations" view. Neil Levy uses the term "quality of will" views to refer to views like these (2011: 158). Some theorists have (due to the influence of T. M. Scanlon's work on these views) described them as "Scanlonian" (Miller, 2014; Shoemaker, 2015: 133). However, see Talbert (n.d.) for an argument that these theorists' views are all properly referred to as versions of Attributionism.

Hieronymi concludes that "we cannot enjoy discretion with respect to those things for which we are most fundamentally responsible" (Hieronymi, 2014: 19-20). Lastly, as Matthew Talbert succinctly states, "[t]hings over which we *lack* control are in fact fundamental to responsibility on the attributionist approach" (2019: 25, n. 3, emphasis added).

The disagreement between these two views is, at root, a disagreement about what we can be directly (or fundamentally) responsible for, and in virtue of which we are responsible for all else.[6] In what follows I argue that responsibility for attitudes cannot adequately explain degrees of praiseworthiness, and thus that Attributionism is false. We can begin by identifying the central principles that attributionists seem committed to concerning an agent, *S*, and something, *A*, for which *S* is responsible:

(1) *S*'s responsibility for *A* is entirely explained by the judgments and attitudes that *A* reflects.[7]

If (1) is true, another principle plausibly follows:

(2) *S*'s *degree* of responsibility for *A* is entirely explained by the judgments and attitudes that *A* reflects.[8]

If (2) is true, then it also seems to follow that:

(3) If there is no difference between two agents with respect to their judgments and attitudes then there is no difference in their degree of responsibility.

It's this final principle, (3), that most clearly runs up against our judgments concerning praiseworthiness. Principle (3), in short, states that any two agents that are "attitude duplicates" (i.e., have all and only the same attitudes) are equally blameworthy or praiseworthy.[9] If (3) is false, and if it

[6] Some might contend that these two views are in fact about different things altogether, since they are sometimes said to operate with different notions of responsibility. Although there may be something to the idea that there are different notions of responsibility at work in these two views, there is nevertheless a significant overlap about the appropriateness of the reactive attitudes, as I explain in note 3. Importantly, the appropriateness of the reactive attitudes involved in praise is the central focus of this paper.
[7] I intend "explained by" to track what grounds responsibility, such that it could be substituted with "in virtue of".
[8] Here (and in (3)) I intend "degree of responsibility" to mean degree of praiseworthiness or blameworthiness.
[9] The term "attitude duplicates" is shorthand, since I intend the term to include not only what might narrowly be referred to as attitudes (e.g., desires, emotions), but also other non-voluntary items or states such as judgments,

follows from each of the previous principles, then it follows that they are also false. Thus, if *any* of these three principles are essential to Attributionism, then Attributionism is false.

Matthew Talbert (2019) expresses a commitment to Principle (3) as a clear implication of Attributionism concerning cases of circumstantial luck, and offers an example that forcefully illustrates this commitment. During World War II, the British Royal Air Force bombed the German city of Dresden, resulting in the deaths of 25,000 people. Talbert writes:

> [I]magine a would-be Dresden pilot, grounded because of illness, who fully shared the intentions and attitudes of the pilots who actually participated in the mission. Suppose that an attributionist regards the actual Dresden pilots as blameworthy because of their normative judgments regarding the prospect of dropping incendiary bombs on civilians. If our attributionist finds that the would-be bomber fully shared these judgments, it is hard to see how she can avoid regarding him as blameworthy *to the same degree* as the pilots who actually participated in the bombing raid (2019: 34-35, emphasis added).

Talbert recognizes that we're naturally inclined to blame the actual bomber more than the would-be bomber. But Talbert maintains that this inclination can be explained away: we're inclined to blame the actual bomber because his actions highlight underlying judgments that we find morally objectionable. If we were to discover that the would-be bomber *would* have done the same things because he shared the same attitudes (i.e., they are attitude-duplicates), then we should find him equally blameworthy. Talbert's remarks reveal the heart of Attributionism: agents are not fundamentally responsible for their actions, but rather for their attitudes. On Attributionism, actions serve only as epistemic grounds for our responsibility judgments: they are merely indicators of what agents are fundamentally responsible for; viz., our underlying attitudes, judgments, or quality of will.

It is because of this that, on Attributionism, an agent's actions can never increase an agent's overall degree of blameworthiness or praiseworthiness past the degree to which the agent is responsible simply in virtue of their underlying attitudes. And crucially, it is this fact that prevents the attributionist from being able to capture everything that matters for praiseworthiness.

2. Three Desmonds

Reflection upon the case of Desmond Doss suggests that an agent's actions *can* increase his overall degree of praiseworthiness past the degree to which he might be praiseworthy in virtue of his

---

dispositions, and anything that might be included under the general heading of "quality of will". Accordingly, when I use the term "attitudes" in this broader sense, I also mean to include these other kinds of non-voluntary items and states.

underlying attitudes. Perhaps the clearest way of motivating this claim is to consider a comparison between two agents who are attitude duplicates. One way to do this is to suppose that one of the two agents performs a given action and the other doesn't, and then ask whether the performance of that action makes the first agent more praiseworthy than his attitude duplicate.

However, if we're focusing only on whether the performance of a single action can make a difference in degrees of praiseworthiness between two attitude-duplicates, then our intuition about the relative praiseworthiness of those agents may be less than clear. Even if the performance of the action does increase the first agent's degree praiseworthiness over the other's, that difference may not stand out if the increase in degree of praiseworthiness (due to the performance of the single action) is not very great. But there is a way of sharpening our intuitions on this point: if the performance of an action can make even a small difference to an agent's degree of praiseworthiness, then we should expect to see that difference more clearly when the degree of praiseworthiness of that agent accumulates across a *series* of actions. My strategy, then, is to focus on a comparison among two or more agents, not with respect to the performance of a single action, but rather across a series of actions.

To this end, let's return to Desmond Doss and ask the question: what if Desmond never had the opportunity to stand alongside his compatriots in the Battle of Okinawa? Call the actual Desmond Doss *Desmond 1*. Now consider a counterpart, *Desmond 2*:

> DESMOND 2: Desmond 2 is just like Desmond 1 except that, before beginning basic training, he was injured while working at the shipyard and consequently prevented from joining the war. If Desmond 2 had joined the war, then he would have acted just as Desmond 1 actually did at the Battle of Okinawa.

Needless to say, it would be quite counterintuitive to suggest that Desmond 2—who never saved anyone's life, nor risked his own to do so—is as praiseworthy as Desmond 1. Now, recall that attributionists are committed (says Talbert) to parity concerning the blameworthiness of the actual and would-be Dresden bombers. Are attributionists then also committed to saying that Desmond 1 and Desmond 2 are equally praiseworthy? Fortunately for them, they're not. What grounds an agent's responsibility, on Attributionism, are the agent's *actual* attitudes, and not merely the truth of certain counterfactuals about how the agent would behave under certain circumstances. While an agent's actual attitudes may ground the truth of certain counterfactuals (e.g., the would-be Dresden bomber's actual attitudes may ground the truth of counterfactuals about how he would have acted that day were

he not ill), the truth of counterfactuals are not always indicative of an agent's actual attitudes. For example, it may be true that I would have behaved in particularly generous or self-sacrificial ways had I chosen to devote my life to humanitarian work. And it's plausible that in such counterfactual scenarios my actions would reflect underlying virtuous attitudes. But, of course, none of this implies that I *actually* possess those attitudes. It only implies that I would have had such attitudes had I been in those circumstances.

Accordingly, attributionists have a principled way of avoiding the counterintuitive claim that Desmond 1 and Desmond 2 are equally praiseworthy, since it's highly implausible that Desmond 2 (sitting at home in Virginia) would have all and only the same attitudes as Desmond 1 had during the battle of Hacksaw Ridge. Though Desmond 2 may have had some similar attitudes as Desmond 1 (i.e., attitudes had prior to the war), Desmond 1 plausibly developed attitudes over his course of training and time with his compatriots leading up to the Battle of Okinawa (e.g., concern for *those* particular men, a willingness to die for them, etc.). So, although the counterfactual is true— *if* Desmond 2 had joined the war, then he would have acted just as Desmond 1 actually did at the Battle of Okinawa—this counterfactual isn't true in virtue of the actual set of attitudes that Desmond 2 has, but rather in virtue of those attitudes that he *would* have come to develop by the time he would have entered the Battle of Okinawa. Thus, Desmond 2 isn't an attitude duplicate of Desmond 1 during Hacksaw Ridge, and so attributionists needn't say he is just as praiseworthy as Desmond 1.[10] But now consider another, nearer counterpart of Desmond 1, *Desmond 3*:

> DESMOND 3: Desmond 3 is just like Desmond 1 except that, as he and his platoon were scaling Hacksaw Ridge, Desmond 3 slipped and broke his leg. Thus, Desmond 3 was prevented from joining the Battle of Okinawa. If Desmond 3 hadn't broken his leg, then he would have acted just as Desmond 1 actually acted at the Battle of Okinawa.

Desmond 3 is plausibly an attitude duplicate of Desmond 1. Unlike Desmond 2, the relevant counterfactual is true in virtue of the actual set of attitudes that Desmond 3 has (and not merely the attitudes that he *would* have if he hadn't broken his leg).[11] Thus, Attributionism's commitment to

[10] The availability of this response drives a wedge in between Attributionism and the view espoused by Michael Zimmerman (2002) concerning moral luck: while Zimmerman maintains that the truth of certain counterfactuals alone is enough to ground blameworthiness and praiseworthiness, Attributionists needn't follow him there (Talbert, 2019: 35-36).

[11] We can also simply stipulate this.

Principle (3) implies that, although he may be praiseworthy for *fewer things*, he is no less praiseworthy than Desmond 1. But it's highly implausible that Desmond 3 (lounging in his tent eating lunch) is as praiseworthy as Desmond 1. We needn't deny that Desmond 3 is praiseworthy for his virtuous dispositions, or his willingness to perform life-saving actions. But surely Desmond 1, who actually did so—who through great effort mustered up the courage to remain on Hacksaw Ridge after every other able-bodied soldier had retreated back down the cliff, who risked his life time and again to save the lives of his compatriots, who painstakingly crawled through blood and mire, achingly dragged each soldier to the edge of the cliff and lowered them to safety, and who on a later day suffered serious injuries while seeking to aid the injured—is not merely praiseworthy for more things, but more praiseworthy.

3. Accommodationist Strategies

There are two general avenues of response available to the attributionist. The first is to accept the implication (bite the bullet?) that Desmond 3 is just as praiseworthy as Desmond 1, even though he didn't act heroically. But conversations with attributionists lead me to believe that this won't be the attributionist's preferred route—at least not initially.[12] Instead, I believe that the attributionist will be tempted to a second avenue of response that involves seeking a way to accommodate the intuitive claim that Desmond 1 is more praiseworthy than Desmond 3. In what follows I consider two accommodationist strategies. Unfortunately for the attributionist, each strategy is either in tension with their own commitments or *ad hoc*.

The first accommodationist strategy is to maintain that agents are praiseworthy *each time* that they act in a way that reflects underlying good or virtuous attitudes. On this strategy, the more such actions an agent performs, the more praiseworthiness the agent accrues. Thus, Desmond 1 (who performs many more actions, each of which reveal morally commendable attitudes) turns out to be significantly more praiseworthy than Desmond 3.

We can see the problem with this strategy by focusing on what, exactly, an agent is supposed to be directly praiseworthy for. It is not, according to Attributionism, the agent's *actions*. Nor is it particular *reflections* of underlying attitudes in those actions. Instead, according to Attributionism, an agent is fundamentally praiseworthy only for their attitudes themselves. But the number of actions an

[12] I am indebted to Angela Smith and Matthew Talbert for conversations on this and related concerns for Attributionism.

agent performs doesn't, on its own, multiply the number of attitudes those actions reflect. Instead, for any agent there will be certain number of attitudes that can be reflected in any number of actions. If some attitude (e.g., that one's compatriots are worth dying for) is reflected in multiple actions (e.g., saving the life of soldier X, soldier Y, soldier Z, etc.), there is still only *one* attitude that's being reflected. In other words, the attributionist can't double-count (or triple-count, etc.) attitudes in an attempt to accommodate the intuition of increased praiseworthiness.[13] Call this the *no double-counting rule*.

The second accommodationist strategy maintains the possibility that an agent might form a new judgment or attitude just prior to each action. If this is so, it won't run afoul of the *no double-counting rule*, as the first strategy does. Consider, then, how this second strategy might apply to the case at hand. One might plausibly stipulate that both Desmond 1 and Desmond 3 share the general attitude that their compatriots are worth dying for. In the thick of things, though, Desmond 1 forms new, more particular judgments or attitudes—ones directed toward particular compatriots—prior to each life-saving act (e.g., that *Adam* is worth dying for, *Ben* is worth dying for, *Carl* is worth dying for, etc.).[14] If this is so, then Desmond 1 is praiseworthy for a number of things that Desmond 3 is not, namely, these more particular attitudes. Consequently, the attributionist might argue, Desmond 1 is more praiseworthy than Desmond 3.

The most apparent problem with this strategy is that it accommodates the relevant intuition only by resisting the stipulation that Desmond 1 and Desmond 3 are, in fact, attitude duplicates. But there's no reason why it can't plausibly be stipulated that Desmond 3 also has the more particular attitudes mentioned above, and this is especially so if we understand these attitudes *dispositionally*. It's generally acknowledged that one may have attitudes that one has never consciously attended to but that nevertheless would become occurrent under certain conditions. As Angela Smith explains, the attitudes that are attributable to us are not limited to ones that we are conscious of:

> I may not realize, until I am faced with a choice, that I value the intellectual freedom and autonomy associated with a career in academia more highly than the economic rewards and benefits associated with a career in law. Or I may discover in some situation that I care more about being liked by others than I do about standing up for my moral principles (2005: 252).

[13] I am indebted to Matthew Talbert for a helpful correspondence on this point.
[14] Although I've construed these as judgements, they can just as well be construed in terms of other attitudes (e.g., desires).

Accordingly, Smith maintains that the relevant attitudes (in this case, evaluative judgments) are "not necessarily consciously held propositional beliefs, but rather tendencies to regard things as having evaluative significance" (2005: 251). Notably, other prominent attributionists agree on this (Scanlon, 1998: 272, Talbert, n.d.). So, even if Desmond 1 is more likely to become consciously aware of his more particular attitudes while saving particular compatriots, it's plausible that these attitudes are not *formed* at that time, but rather attitudes Desmond 1 previously held that were then *manifested* in the heat of the moment. Thus, if Desmond 1 and Desmond 3 are attitude duplicates *prior* to the battle, then Desmond 1's activities during the battle needn't alter this similarity. And again, at the very least, there's no reason why this can't be plausibly stipulated.

Perhaps, however, the attributionist can modify the second accommodationist strategy. The attributionist might claim that even if the propositional content of an attitude is the same, the particular attitudes manifested in the moment might take on a qualitatively different character. Perhaps, for example, the attitude has a greater felt intensity when it is manifested in action. Suppose that something like this is true of Desmond 1. Even so, it's unclear why the fact that they are more intensely *felt* in the moment suffices to make them *different* attitudes. Moreover, an insistence that it does departs from the dispositional conception of evaluative judgments (discussed above) that attributionists operate with. While the *manifestation* of a disposition may be occasioned by a different phenomenal quality, this seems beside the point concerning the disposition itself. And, in any case, we should pause to ask ourselves whether our initial intuition of Demond 1's exceptional degree of praiseworthiness really does depend upon an underlying belief we hold that he formed these qualitatively different attitudes in the process of performing his heroic acts (and thus that he is not, in fact, an attitude duplicate of Desmond 3). It instead seems more plausible that this modification of the second accommodationist strategy (according to which the agent might develop more particular attitudes of some sort or another in the moment, and that this is what explains increased praiseworthiness) is instead an *ad hoc* response to the argument I have offered against Attributionism. Furthermore, I think we would intuitively take Desmond 1 to be exceptionally praiseworthy—and indeed, *more* praiseworthy than Desmond 3—regardless of whether we are operating with some underlying (and heretofore unrecognized) assumption that Desmond 1 formed such particular and distinctive attitudes in the moment.

As stated earlier, I expect that most will be inclined to agree upon the judgment that agents like Desmond 1 are more praiseworthy than agents like Desmond 3. As we have seen in this section, though, the most obvious strategies the attributionist can use to accommodate this judgment fail for

one reason or another. Moreover, these strategies have never previously figured into attributionist theories, and appealing to them in the face of the present challenge would seem unduly *ad hoc*. Perhaps there are more promising strategies available. I leave their development to the attributionist.[15, 16]

4. Effort and Sacrifice

If Attributionist strategies fall short in explaining the increased praiseworthiness of agents like Desmond 1, we need alternatives. In this section I put forward what I believe to be two of the most plausible explanations. The first is that (i) agents like Desmond 1 exert *effort* in performing the relevant actions that their attitude duplicates don't. The second is that (ii) agents like Desmond 1 make *sacrifices* in performing the relevant actions that their attitude duplicates don't. These two explanations are not mutually exclusive. While my sense is that each explanation will be true for at least some scenarios, I don't take a stance on this here. Instead, I argue for the disjunctive claim that (i) or (ii) is true.[17] Since each of these factors can be instantiated only through action, Attributionist theories cannot appeal to either. Because of this, Attributionism cannot make use of the most plausible explanations of increased praiseworthiness.

As Dana Nelkin observes, we often take an agent's degree of responsibility (praiseworthiness or blameworthiness) to depend upon the degree of effort and sacrifice (2016: 357).[18] Nelkin offers a gloss of both effort and sacrifice as distinct species under the broader heading of difficulty:

> [D]ifficulty can be understood in at least two ways: on the one hand, it can be understood as requiring a great deal of effort, and, on the other, it can be understood as requiring a great sacrifice of one's interests. These often go together, but they might come apart. For some, it might not require a great effort to do

[15] One additional accommodationist strategy that attributionists might pursue is that Desmond 1 is more praiseworthy than Desmond 3 the former retained his commendable attitudes in the face of danger and Desmond 3 did not. I am grateful to Randy Clarke for suggesting this potential strategy for the attributionist.

[16] There is one potential strategy that is (unlike the ones discussed here) *non*-accommodationist in nature and that might be developed along the lines of Scanlon's approach to outcome luck in cases of blameworthiness. Scanlon's strategy there is to deny that outcomes make a difference to degrees of blameworthiness but may increase the *significance* of a fault for those negatively affected (Scanlon, 2008: 147-150; Scanlon, 2015: 105; Talbert, 2019: 29-33). Along these lines, one might argue that the significance of Desmond 1's actions is greater for those positively affected by them, and that this explains our inclinations to respond differently to the two Desmonds, while denying that this difference is due to any difference in degrees of praiseworthiness. I am grateful to Matthew Talbert for suggesting this potential strategy.

[17] I leave it open whether there may be further factors that also contribute to increased praiseworthiness, such as the value of outcomes, which I discuss toward the end of this section. Notably, this factor also cannot be accounted for by Attributionism.

[18] Nelkin maintains that this common assumption is correct: "[t]he more difficult to act well, all else equal, the more praiseworthy for succeeding, and the less blameworthy for failing" (2016: 372).

> something that results in great sacrifice. It might be "easy" in the sense of requiring little in the way of either physical or mental effort for a particular soldier to jump on a grenade to save her fellow soldiers, for example. She doesn't have to try hard at all, but the sacrifice is very large (Nelkin 2016: 357).[19]

I will understand the distinction between effort and sacrifice roughly along these lines, while also fleshing out each concept with a focus on its plausible contribution to increased praiseworthiness.

Before proceeding, it is worth keeping in mind the distinctions between praiseworthiness for *having* certain character traits or dispositions, praiseworthiness for *developing* certain traits or dispositions, and praiseworthiness for *acting* virtuously. This is not to say that these are unrelated. Perhaps, for example, the third of these depends partly upon one or both of the first two. Nevertheless, in what follows I offer reasons to think that the third of these isn't *exhausted* (i.e., fully explained) by the first two, and that it partly depends upon effort or sacrifice. If this is correct, then these factors can help explain the difference in praiseworthiness between agents like Desmond 1 and agents like Desmond 3. Because action is required to instantiate the relevant effort or sacrifice, Attributionism can't capture all that matters for praiseworthiness.

4.1 Effort

I will understand effort as a kind of work or energy, either mental or physical, that is exerted in service of an end.[20] For this reason, effort is instantiated through intentional action: agents *do* things with the aim of bringing about the end at which the effort is directed. It's plausible that Desmond 1 exerts considerable effort in acting with the aim of saving his compatriots. If effort can increase an agent's degree of praiseworthiness, then this offers a promising explanation of why Desmond 1 is more praiseworthy than Desmond 3.

Mental effort is exerted in the performance of mental actions, which include directing or sustaining attention, maintaining concentration, engaging in reasoning or problem-solving, repressing urges or resisting impulses, and directing one's focus on long-range goals (Baumeister and Heatherton, 1996; Shanhav et al., 2017; Andre et al., 2019). Perhaps the most obvious source of mental effort is due to the presence of contrary desires or impulses. To take a trivial example, refraining from eating a pastry this morning requires mental effort on my part because I have a desire to eat a pastry. This

[19] Of course, it may well be that both effort and sacrifice can contribute to an agent's overall degree of praiseworthiness. In this vein, Maslen et al. (2020) argue that an agent's degree of praiseworthiness depends upon multiple factors.
[20] For accounts of mental effort as analogous to work or energy, see Kool & Botvinick (2014) and Boksem & Tops (2008), respectively.

contrary desire presents a psychological obstacle, and overcoming it requires mental effort in the form of self-control. In this case, I may attempt to overcome this obstacle by resisting or "overriding" my urge to eat the pastry, redirecting my attention off of the pastry and onto the task at hand (e.g., grading papers), or reminding myself of my long-range health goals.[21, 22]

However, if mental effort is required to resist contrary desires when performing morally right or good actions, then this fact seems to reflect morally objectionable attitudes or a lack of virtue (Foot, 2002: 11ff; Sorensen, 2009: 91). Perhaps Desmond has selfish desires to prioritize his own life and save himself. Or perhaps Desmond has some residual resentment towards his compatriots for bullying him. If so, then one might argue that the exertion of *this* sort of effort wouldn't contribute to his degree of praiseworthiness. But this is too quick, for two reasons.

First, whether the effort exerted against contrary desires contributes to the agent's degree of praiseworthiness for acting rightly may depend partly upon whether the agent is responsible for having (or developing) those desires in the first place. If the agent is not responsible for the contrary desires, then it would seem that the effort exerted against these desires *increases* the agent's degree of praiseworthiness. To illustrate, suppose that an agent was raised in a racist family and consequently comes to develop racist beliefs and attitudes.[23] The agent, we may suppose, later comes to recognize those racist beliefs and attitudes as false and evil. Of course, this recognition doesn't immediately or entirely eliminate deep-seated inclinations to think and behave in racist ways. Nevertheless, the agent is committed to fighting against these inclinations. This agent is more praiseworthy (than he would have otherwise been) for the effort expended with the aim of acting rightly.

Second, even if an agent *is* responsible for having contrary desires, it's not clear that effort exerted against those desires in acting rightly doesn't increase his praiseworthiness for doing so. Some, for example, may be partly responsible for their own substance addiction (e.g., as a foreseeable result of prior voluntary behavior). Nevertheless, we rightly praise individuals for struggling against their desire for the substance in their efforts to get clean.

I have argued that, although contrary desires sometimes reflect a fault, this needn't preclude the effort expended in resisting them from increasing an agent's praiseworthiness. But even though the presence of contrary desires sometimes reflects an underlying moral fault or deficiency (e.g.,

---

[21] Each of these are controlled exercises of the capacity for "self-regulation" or "effortful control" (Baumeister and Heatherton, 1996; Baumeister, 2002; Andre et al., 2019).

[22] On a very influential account developed in Baumeister et al. (1998), the exercise of mental effort via self-control involves the use of a limited and depletable resource. That view is now contested (Carter et al., 2015).

[23] Alternatively, we can imagine that the agent has these attitudes surreptitiously implanted by a meddling neuroscientist.

insufficient concern for others, resentment), it needn't. For example, suppose that Desmond has a desire to retreat down Hacksaw Ridge (instead of remaining to save his compatriots) so that he can survive and return home safely to his wife. This desire doesn't reflect insufficient concern for the lives of his compatriots, or selfishness, but instead concern for something else that is also morally valuable (i.e., his relationship with his wife). If we suppose that saving the lives of his compatriots was objectively even more important than returning to his wife, the effort expended in overcoming *this* sort of contrary desire certainly seems to increase his praiseworthiness.

Even apart from contrary *desires*, mental effort might be required due to phobia, nerves, or even an appropriate level of fear due to actual danger (Sorensen, 2009: 90). And these sorts of psychological obstacles don't preclude the effort expended in overcoming them from increasing an agent's praiseworthiness.

Lastly, we needn't limit the relevance of mental effort to cases of effort *against* contrary desires or overcoming some other kind of psychological obstacle, since effort (in the relevant sense) is just the exertion of work or energy in service of an end. When the end in question is a morally good one, then the effort expended may reasonably increase praiseworthiness. Desmond's success required carefully navigating the battlefield at Hacksaw Ridge, locating injured soldiers, avoiding enemy fire, and the development of a strategy to transport injured soldiers safely back down the cliff. Each of these tasks require sustained attention, focus on goals, and various kinds of problem-solving, and therefore involve the exertion of mental effort.

Although I have focused primarily on the exertion of mental effort, we may also countenance physical effort. To take a more mundane example, we can suppose that I have two reliable friends who have promised to help me move. On the day of moving, though, one of my friends comes down with a stomach virus, and is unable to help me move. I may have no doubt that the friend who fell ill was quite willing to help me move. But at the end of a long day of strenuous labor, I surely owe a special debt of gratitude to the friend who, sitting next to me sunburnt and dripping sweat, actually did. Notably, although physical effort may be experienced as particularly grueling depending upon one's underlying mental states (e.g., whether one begrudges the work or not), sometimes the physical effort involved in performing actions may remain quite similar, regardless of underlying mental states. The sheer physical exertion involved in Desmond Doss's dragging one body after another through dirt and blood is what it is, regardless of how willing he was to do it.

4.2 Sacrifice

Following Nelkin, I will understand sacrifice in terms of a loss that an agent incurs with respect to their interests. Although one might construe sacrifice broadly, I'm interested in a somewhat narrow notion of sacrifice that's carved out by the following considerations. Taking these considerations into account makes it more plausible that this kind of sacrifice can increase an agent's praiseworthiness.

The first consideration is that a sacrifice is not merely something lost, but something lost *in the pursuit of some end*. The basic structure of sacrifice, then, is that an agent loses X *for the sake of* Y. And an agent sacrifices X for Y in this sense only if the agent intends to bring about (or contribute to bringing about) Y, and performs at least some action toward this end. Clearly, not *any* loss will count as a sacrifice in this sense. For example, if I'm pushed from the top of a bridge to prevent a trolley from running over five innocent people, I certainly incur a loss insofar as I lose whatever value the remainder of my life would have had. But this isn't a sacrifice in the relevant sense, since I don't lose my life with the intention of preventing the deaths of the five other people.

Second, an incurred loss counts as a sacrifice only if the loss falls within the range of *expected outcomes* of the action.[24] Suppose that an agent is driving someone to the emergency room and is just about to arrive when the driver's side of the car is crushed by an oncoming truck. Tragically, the agent loses their life in the pursuit of saving the life of another. But however praiseworthy the agent might be for their intention to save the other's life, it would be odd to say that the agent *sacrificed* himself to save the life of another. After all, the agent wasn't driving to the ER with the awareness (occurrent or otherwise) that doing so might cost him his life.

The above considerations give more precise shape to the sense of sacrifice that is relevant to praiseworthiness.[25] It's plausible that Desmond 1 makes this sort of sacrifice. Desmond was seriously injured (both his arm and leg) while acting to save the lives of others, and as a result was never able to return to full time work after the war. He was also certainly aware (as any soldier would be) that his

---

[24] This is related to the discussion of praiseworthiness in Maslen et al. (2020), who write that "[c]ostliness tracks expected costs—the probability × magnitude of disvalue—rather than absolute costs" (312). Although I agree that the costs or sacrifice must fall into the scope of what is expected by the agent, the actual costs also seem to matter. One might act with a willingness to sacrifice something expected, but end up not having to sacrifice anything. We might praise someone for their willingness to sacrifice, but reasonably praise them more so for actually sacrificing that which they were willing to.

[25] We might add one further, more tentative consideration. In a broad sense, any agent who pursues some end Y with the awareness that they might lose X in doing so is (insofar as they are behaving rationally) *willing* to assume the risk of losing X. But an agent can intend to bring about Y while begrudging the possibility that they might lose X in the process. Take, for example, two people on their way to work who stop to help a stranded motorist change a flat tire. There's a difference, it seems, between the one who grumbles to himself, "this had better not make me late to work", and the other who says to himself, "if I have to be late to work for this, so be it." Perhaps both make praiseworthy sacrifices, but the sacrifice of the second seems to be more praiseworthy. Even so, this consideration seems heavily dependent upon the agent's attitudes that *accompany* the action, and so may be more amenable to attributionist accounts.

actions risked his own safety, and by all accounts displayed a sincere acceptance (rather than a begrudging attitude) toward those risks. Furthermore, it's highly plausible that his sacrifice is at least part of the reason we're inclined to judge that he is more praiseworthy than Desmond 3, and that it is a partial explanation of his increased praiseworthiness.

We may still ask *why* this is the case. Here I offer only a brief and tentative possibility. Perhaps the sort of sacrifice delineated by the conditions I discuss above merits a kind of moral *compensation* for an incurred loss. This thought fits well with our practices of praising. When we praise or honor heroes, we are often painfully aware that we cannot truly *repay* them for their heroism, and instead treat our praise as a substitute for repayment. In some cases this may be the most we can do to appropriately compensate them.[26, 27]

4.3 The Value of Outcomes

I have not addressed an additional factor that may increase an agent's overall degree of praiseworthiness: the value of the outcome(s) that an agent brings about as a result of their action(s) (Maslen et al., 2020). The fact that dozens of men lived that day because of Desmond Doss's heroic actions itself is of great value, and plausibly contributes to his overall degree of praiseworthiness.

Even so, my case needn't rest upon this claim. While it seems abundantly clear that Desmond 1 is more praiseworthy than Desmond 3, the same sort of comparative judgment is nearly as clear even when considering one last variation of the case where, due to bad luck, Desmond fails to save the lives of any of his compatriots. Because this agent is more similar to Desmond 1 than the other two I have introduced, I will call him Desmond 1*:

> DESMOND 1*: Desmond 1* is just like Desmond 1 except that, unbeknownst to him, each soldier that he helps lower down Hacksaw Ridge gets picked off and killed by a nearby Japanese sniper.

---

[26] The tentative explanation that I offer here may have problematic implications. If being praiseworthy is to be deserving of compensation, then it seems to follow that one ceases to be praiseworthy once one is fully compensated. And this would seem to imply that (given how much praise he has received) Desmond Doss is no longer praiseworthy for his actions at Hacksaw Ridge. This would be a counterintuitive implication. For an argument against analogous views of blameworthiness, see Clarke (n.d.). I am grateful to Randy Clarke for pointing out this concern.

[27] Though I have discussed the concepts of effort and sacrifice separately, they may in practice coincide. The most obvious way this might be is that the exertion of effort *itself* is "costly" (Kool and Botvinik, 2014; Kurzban, 2016). And, as Maslen et al. (2020) maintain, we might explain the relevance of effort as part of an agent's "costly commitment" to bringing about a valuable end, where this can involve *successive* choices with the aim of bringing about that end (311). This fits well with the cases discussed here.

Desmond 1 and Desmond 1* are each more overall praiseworthy than Desmond 3. If that's true, then it cannot be the value of outcomes alone that explains why Desmond 1 is more praiseworthy than Desmond 3. Here we can simply look to what Desmond 1 and Desmond 1* have in common. All of the remarks I have made about effort and sacrifice apply just as much to Desmond 1* as they do to Desmond 1. So, even apart from consideration of the value of the outcomes, Desmond 1 is more praiseworthy than Desmond 3. And the best explanation of this fact is (inclusively) either the effort exerted or sacrifice(s) made.

4.4 Attributionism, Effort, and Sacrifice

The forgoing should not be taken to suggest, however, that attributionists leave *no* room at all for effort or sacrifice to play a role in their theory. Effort exerted or sacrifice made can serve as *indicators* of one's underlying attitudes or even the degree of one's moral concern. As Nelkin nicely puts it,

> One might agree that in principle one could do something easy and do something difficult with the same high degree of moral concern and so be equally praiseworthy. But doing something good when it is difficult allows us to see how high one's moral concern is in a way that doing something good when it is easy does not distinguish between low and high degrees of moral concern (Nelkin, 2016: 361).

Nevertheless, the role played by effort and sacrifice in such theories is purely epistemic. In the actual world, Desmond 1's actions provide us with stronger evidence of his underlying attitudes than we would typically have of an attitude duplicate like Desmond 3. But this makes no difference to the foregoing discussion. Since we have stipulated how Desmond 3 would act, we have been in no better an epistemic position with respect to Desmond 1 than Desmond 3. Again, attributionists cannot appeal to effort or sacrifice to explain Desmond 1's increased praiseworthiness, despite these being among the best explanations.

Conclusion

Agents who perform morally good actions are sometimes more praiseworthy than attitude-duplicates who do not. This difference is clearer as the degree of praiseworthiness of an agent accumulates across a series of actions. The best explanations of increased praiseworthiness appeal to factors that are instantiated in *actions* alone, and therefore cannot be explained purely in terms of an agent's underlying attitudes. If this is true, then Attributionist accounts of moral responsibility--accounts that ground responsibility fundamentally in *attitudes* alone—have been found wanting.

Acknowledgements: I am grateful to Matthew Talbert and Randy Clarke for helpful feedback on earlier versions of this paper, and to Angela Smith for a valuable conversation on topics related to the arguments offered in this paper.

References

André, N., Michel A., and Baumeister, R. F. (2019). An Integrative Model of Effortful Control. *Frontiers in Systems Neuroscience*, 13(79), 1-22.

Baumeister, R. F. (2002). Ego Depletion and Self-Control Failure: An Energy Model of the Self's Executive Function. *Self and Identity*, 1(2), 129-136.

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265.

Baumeister, R. F. and Heatherton, T. F. (1996). Self-Regulation Failure: An Overview. *Psychological Inquiry*, 7(1), 1-15.

Berman, E. (2016, November 3). The True Story Behind *Hacksaw Ridge*. *Time*. Retrieved August 29, 2020, from https://time.com/4539373/hacksaw-ridge-movie-true-story/

Blair, E. (2016, November 4). The Real 'Hacksaw Ridge' Soldier Saved 75 Souls Without Ever Carrying A Gun. *NPR*. Retrieved August 29, 2020, from https://www.npr.org/2016/11/04/500548745/the-real-hacksaw-ridge-soldier-saved-75-souls-without-ever-carrying-a-gun

Boksem, M. A. S. and Tops, M. (2008). Mental fatigue: Costs and Benefits. *Brain Research Reviews*, 59(1), 125-139.

Clarke, R. (n.d.) Still Guilty. Forthcoming in *Philosophical Studies*.

Foot, P. (2002). *Virtues and Vices*. Oxford University Press.

Fritz, K. (2018). Moral Responsibility, Voluntary Control, and Intentional Action. *Philosophia*, 46(4), 831-855.

Hieronymi, P. (2008). Responsibility for Believing. *Synthese*, 161, 357-73.

Hieronymi, P. (2014). Reflection and Responsibility. *Philosophy and Public Affairs*, 42(1), 3-41.

Kool, W. and Matthew, B. (2014). A Labor/Leisure Tradeoff in Cognitive Control. *Journal of Experimental Psychology: General*, 143(1), 131-141.

Kurzban, R. (2016). The Sense of Effort. *Current Opinion in Psychology*, 7, 67-70.

Lange, K. (2017, February 28). Pfc. Desmond Doss: The unlikely hero behind 'Hacksaw Ridge'. *U.S. Army*. Retrieved August 29, 2020, from https://www.army.mil/article/183328/pfc_desmond_doss_the_unlikely_hero_behind_hacksaw_ridge

Leepson, M. (2015). Desmond Thomas Doss (1919–2006). *Dictionary of Virginia Biography, Library of Virginia* (1998–). Retrieved August 29, 2020, from https://www.lva.virginia.gov/public/dvb/bio.php?b=Doss_Desmond_Thomas

Levy, N. (2011). *Hard Luck*. Oxford University Press.

Maslen, H., Savulescu, J., and Hunt, C. (2020). Praiseworthiness and Motivational Enhancement: 'No Pain, No Praise'? *Australasian Journal of Philosophy*, 98(2), 304-318.

Miller, D. (2014). Answerability, Blameworthiness, and History. *Philosophia*, 42(2), 469-486.

Nelkin, D. (2016). Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness. *Nous*, 50(2), 356-378.

Scanlon, T. M. (1998) *What We Owe to Each Other*. Harvard University Press.

Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Harvard University Press.

Scanlon, T. M. (2015). Forms and Conditions of Responsibility. In R. Clarke, M. McKenna, & A. M. Smith (Eds.), *The Nature of Moral Responsibility*: *New Essays* (pp. 89-111). Oxford University Press.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., and Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, 40, 99-124.

Shoemaker, D. (2015). Ecumenical Attributability. In R. Clarke, M. McKenna, & A. M. Smith (Eds.), *The Nature of Moral Responsibility*: *New Essays* (pp. 115-140). Oxford University Press.

Smith, A. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*, 115, 236-271.

Smith, A. (2008). Control, Responsibility, and Moral Assessment. *Philosophical Studies*, 138, 367-392.

Sorenson, K. (2010). Effort and Moral Worth. *Ethical Theory and Moral Practice*, 13, 89-109.

Talbert, M. (2012). Moral Competence, Moral Blame, and Protest. *Journal of Ethics*, 16, 89-109.

Talbert, M. (2017). Omission and Attribution Error. In D. Nelkin and S. Rickless (Eds.), *The Ethics and Law of Omissions* (pp. 17-35). Oxford University Press.

Talbert, M. (2019). The Attributionist Approach to Moral Luck. *Midwest Studies in Philosophy*, 43(1), 24-41.

Talbert, M. (n.d.). Attributionist Theories of Moral Responsibility, forthcoming in D. K. Nelkin and D. Pereboom (Eds.), *The Oxford Handbook of Moral Responsibility*. Oxford University Press.

Zimmerman, M. (2002). Taking Luck Seriously. *The Journal of Philosophy*, 99(11), 553-576.