

Moral Responsibility, Moral Exemption, and Victimization

Sam Konstan

The essence of providing an accurate analysisⁱ of moral responsibility lies in finding general principles which match our intuitions. In cases where we question the degree to which one can act freely, we tend to rely on our instincts to determine the line of where one should be considered morally responsible. From there, we develop the necessary and sufficient conditions which govern these instances. Through the example of JoJo, the son of an evil dictator, Susan Wolf in “Sanity and the Metaphysics of Responsibility”ⁱⁱ argues that a necessary condition for responsibility is sanity. She defines sanity as the ability to learn morals that align with those of one’s community. In “Culture, Responsibility, and Affected Ignorance,”ⁱⁱⁱ Michelle Moody-Adams objects to the sanity condition, refuting that it can lead to cases of affected ignorance. Affected ignorance occurs when somebody claims to be incapable of learning or intentionally avoids learning a moral code that coincides with their community’s values. People may use affected ignorance to support something morally wrong which benefits them. An example that Moody-Adams brings up is Ancient Greek slaveowners: they benefit from slavery, so they avoid learning its wrongness. These two views interact directly with theoretical

morality and do not consider cases wherein one’s moral desires fail to effectively influence the will. In this paper, I introduce a new thesis about moral responsibility, and I extend the theory from mere theoretical morality into the *dimension of moral action*. In section I, I elaborate on the search for precise principles regarding moral responsibility and more thoroughly introduce the Susan Wolf’s and Michelle Moody-Adams’s views. In section II, I evaluate both ideas. In section III, I propose a framework that I believe solves the problems found in section II. In section IV, I explain how this thesis pertains to moral action. Finally, in section V, I consider possible objections to these views.

I

The question at hand is, at its core, one about free will. A premise taken by both Wolf and Moody-Adams is that free will is a necessary condition for moral responsibility—one cannot be morally responsible for an action over which they had no control. The philosophical community generally accepts that anybody who acts in contradiction with certain instinctual universal principles,^{iv} such as not murdering, is morally responsible. Less clear, however, is the case where the morals one was taught do not coincide with those of their community or with instinctual universal principles. Is someone morally responsible for an action that is wrong but aligns with the views they had been trained in from birth? It is primarily in this realm that the opinions of Wolf and of Moody-Adams clash. The question of responsibility arises in this form

ⁱ An analysis provides necessary and sufficient conditions for one to be morally responsible. Form: *P is morally responsible if and only if A and B*

ⁱⁱ Gideon A. Rosen et al., *The Norton Introduction to Philosophy*, 2nd ed. (New York: W.W. Norton & Company, 2018), 645.

ⁱⁱⁱ *The Norton Introduction to Philosophy* (2018), 764.

^{iv} An unstated premise of Wolf’s thesis is that she assumes a universalist framework. While this is not made explicit, she uses only examples for which we have strong intuitions, such as torturing and killing, and assumes them to be morally wrong. We will put this framework under further investigation later on.

in many situations, which will be touched on later.

One final key ingredient to moral responsibility, which we will discuss more thoroughly in the following sections, is the ability to revise oneself according to proper (whatever that may mean) morality. We will soon turn to Wolf's and Moody-Adams' views along with the thought experiments of JoJo and Ancient Greek Slavery.

Before proceeding, let me introduce you to ordered desires and, from there, the distinction between a desire and volition. The importance of this will be made explicit later in this section and in Section IV. Presented by Frankfurt,^v a desire of the second order entails a preference regarding the existence of or importance placed on a first-order desire. Suppose a person has both a desire to do X and a desire not to do X. Then, possible second-order desires include wanting a desire to do Y, wishing they had no desire to do X, and hoping that the desire to do X prevails over the desire not to do X.

Further, Frankfurt classifies two types of second-order desires: mere desires and volitions. A desire is exclusively the want to have a specific first-order desire (want to want to do X), whereas a volition entails that the person truly wants their desire to do X to prevail. To make this distinction clear, Frankfurt uses the example of a physician versus a drug addict. A physician wants to feel the urge to use drugs to understand his patients better but does not want that urge to push them into action. Frankfurt classifies this case as a mere second-order desire. They want to have a desire, but it is not one which they wish to become effective. On the other hand, an addict who wants to quit and has no desire for the drug, trying everything in their power to reduce the drug's hold on them, has

a second-order volition not to use the drug. They have first-order desires to both use and refrain from using the drug, and at the second order, they want the latter to overcome the former. Simply put, they wish to make the first-order desire effective, moving them to action.

Now, returning to the views of Wolf and Moody-Adams. Susan Wolf begins by examining what she labels as the Deep Self View. This view argues that free will, as well as responsibility, requires that (a) the person-in-question has a deep self (defined as having second-order volitions or values) and (b) that their deep self can directly motivate their will. If a person meets both conditions, the Deep Self View advocate argues, they can be morally responsible for any action. Wolf sees two problems with this. First, she notes that it only submerges the question of how determinism plays into responsibility. The Deep Self View attempts to define the person at the level of the second order but fails to take into account the fact that the second-order desires and volitions are the resultants of higher-order desires, which are, eventually, products of external factors. In a certain sense, nothing is in the deep self's control because it cannot change the factors that shaped it. This lack of control is a serious problem in moral responsibility, and Wolf argues that the Deep Self View fails to solve it.

The second problem that she sees in the Deep Self View, Wolf illustrates through the case of JoJo, the son of an evil dictator. He is taught from birth that it is okay to torture, imprison, or jail people as he desires. JoJo is free in the technical sense—nothing is restricting his ability to act otherwise—but he willingly follows these principles. The Deep Self View would clearly state that JoJo is morally responsible, though Wolf believes otherwise because his turning out the way he

^v *The Norton Introduction to Philosophy* (2018), 635.

did was inevitable. However, Wolf questions, if JoJo is not morally responsible because his wrongdoings were inevitable, and as none of us are holistically in control of ourselves, does that mean that no one can ever truly be morally responsible?

To resolve these two problems, Wolf introduces the sanity condition. The sanity condition requires that, to be morally responsible, one must be capable of accurately perceiving the world and adjusting their values accordingly. This condition allows for self-revision—the realigning of one’s values and morals according to the world’s mainstream (yet subjective) interpretation of morality—by stating that those who have the potential to justifiably^{vi} self-revise are morally responsible. Wolf believes that this resolves both of the above problems with the Deep Self View. First, it solves the issue surrounding determinism by explaining that one need not have total control over themselves to be morally responsible. Instead of getting caught up with the part of oneself that is out of their control, the Sane Deep Self View restricts who can be morally responsible to the sane and, within that, defines moral responsibility as living up to proper moral values. Second, the Sane Deep Self View can account for the case of JoJo. As his deep self is, according to Wolf, not sane, he cannot be held morally responsible.

There are two different possible and equally valid interpretations of the Sane Deep Self View. The first coincides with a universalist ethical framework: there are objective moral laws, and sanity should be judged accordingly. A second possibility is that Wolf could be using a relativist framework. Perhaps

what she means is that one may be considered insane and thus not be morally responsible *to a specific society*. While the former seems more likely, they are both plausible interpretations and need to be addressed. Either way, Wolf believes that, alongside the Deep Self View’s requirements, the sanity condition provides necessary and sufficient conditions for responsibility. Put as simply as possible—if somebody is capable of knowing the right (morally) action and opts not to do it, they are morally responsible.

Moody-Adams sees multiple problems with the Sane Deep Self View. First, Wolf relies on knowing what one *can/could* do, not what they *did* do. As we can only know what a person did and not what they were capable of doing, this makes the Sane Deep Self View impractical. Secondly, Moody-Adams points out how beneficiaries of a morally wrong institution can claim insanity and avoid learning about their actions’ moral to perpetuate the institution and avoid responsibility. Moody-Adams cites four primary forms of this phenomenon, which she refers to as affected ignorance (this could also be called the self-imposed insanity condition). The four forms are (1) denying/hiding from the consequences of a morally wrong action, (2) willfully remaining ignorant by asking not to receive information, (3) not asking questions that may reveal moral wrongness, and (4) avoiding challenging deep-seated beliefs. To illustrate affected ignorance, she asks us to turn to the example of Ancient Greek slavery, where slave owners were said by many to have been incapable of understanding the wrongness of slavery due to the customs of the culture that raised them. However, anti-slavery literature existed at the time. Moody-Adams concludes from this that slave owners did not know of the wrongness of their actions because they did not want to know—they were willfully ignorant.

^{vi} Justifiably meaning according to proper moral values. JoJo can self-revise, but he is revising using the foundational values of torture and murder taught by his father. This means that, even upon revision, JoJo would believe that his actions were good.

The third problem that Moody-Adams finds with the Sane Deep Self View is that, by labeling someone as ‘insane’ and exempting them from responsibility, that person is dehumanized and deemed incapable of morality. If an entire culture does not align with either values of the culture assigning responsibility or objective moral laws (as is the case in societies with slavery), then they are actively being othered. Because of its impracticality, self-imposed insanity, and dehumanization, Moody-Adams rejects the sanity condition as a requirement for moral responsibility.

In review: we have discussed the search for necessary and sufficient conditions for responsibility, the distinction between volitions and desires, and the views of Wolf and Moody-Adams (accompanied by the thought experiments of JoJo and Greek Slaveowners). I now move to evaluate the two theses before giving an alternative that I believe does not fall victim to the same problems encountered by these two theses.

II

To evaluate Wolf’s and Moody-Adams’s views, we must first establish the criteria on which to judge them. I see three main areas which must be considered: responsiveness to examples such as JoJo and slaveowners, impact on other people/cultures, and relation to meta-ethics/metaphysics. First, let us look at the two examples already presented. If we are to decide whether JoJo and the slaveowners are morally responsible, we must first define the term.^{vii}

^{vii} This is deserving of its own work and it is far too large a question to fully address in this one tangentially related paper. I will give a few specifications for the range of moral responsibility as it pertains to this paper, though this is not a complete definition and is certainly subject to further consideration.

The general form of responsibility found in society has a few main purposes, including sparking change, supporting the victim, and setting an example. The first focuses on having a positive impact on the actions of the person in question. Even if somebody had no control over a wrongdoing of theirs, by being held responsible, they might be made aware of their mistakes in a manner that encourages them not to repeat what they did. Although JoJo is (in Wolf’s view) not morally responsible for his actions, punishing him may be a way to ‘retrain’ him with a more conventional moral code. The second purpose, supporting the victim, comes into place in cases such as sexual assault, where some form of detainment and exclusion are required of the aggressor for the victim to feel safe in society. Finally, setting an example aims at the benefit of acquaintances of the person who committed the act. Consider slaveowners: if one is punished, even if they were deemed innocent due to their cultural insanity, holding them responsible may show their neighboring slaveowners that slavery is wrong, thus motivating them to abandon their ways. To have this effect, it is not even necessary that the person committed the act in question, as punishing an innocent man may still have a net beneficial impact on the goodness of his society.

However, I believe that none of these purposes constitute moral responsibility. The kind of responsibility of interest to philosophers relates more to the metaphysical self and the objective rightness/wrongness of its actions—not the practical benefits of holding someone accountable. Any instincts used to govern the discussion of responsibility must be heavily scrutinized, as they are likely to play into the nonphilosophical conception of moral responsibility. Responsibility must in some way relate to how one’s actions compare to their, cultural, or universal values.

Past that, though, we are again left with the question of what moral responsibility is and, again, we are unable to produce a satisfactory answer. With that said, it is valuable in our exploration to know that moral responsibility is viewed in the eyes of the universe, God, or some other power displaced from humanity. Moral responsibility is a metaphysical state, not something assigned by a culture or person.

Thus, we return to asking whether JoJo and Ancient Greek slaveowners are morally responsible with little more than when we started. First, I believe that JoJo shall not be morally responsible. As Albert Einstein famously said, “everybody is a genius. But if you judge a fish by its ability to climb a tree, it will spend its whole life think that it is stupid.” It is not fair to judge a fish by its ability to climb a tree, and analogously it is not fair for somebody to be morally responsible according to a moral code to which they did not agree.^{viii} On this point, I side with Wolf’s interpretation.

The slaveowner example is much simpler. I think that, if what Moody-Adams says about slaveowners being capable of knowing the institution’s wrongdoing, the two agree that the slaveowner is morally responsible. Wolf bases sanity on whether it is possible to know one’s wrongdoing (not whether one does), suggesting a much more metaphysical interpretation of moral responsibility. Her conception is based in the eyes of the universe, not in those of a specific person or group of people. As such, while Moody-

^{viii} Somewhat in conflict with religious and universalist beliefs around moral responsibility, I believe that the described type of responsibility is unreasonable. It is vital to take into account that my argument on its own does not disagree with the religious or the universalist—rather, we disagree in our interpretations of the term moral responsibility. It very well may be that the opposing view of the universalist is correct for his definition of responsibility and mine is correct for mine.

Adams frames affected ignorance as a critique of the sanity condition, Wolf would agree that there is moral responsibility in cases of affected ignorance because the person in question could have known—they just actively chose not to. Hence, there is much less disagreement between the two views in these cases.

The second criterion to judge both views by is their impact on other people/cultures. Moody-Adams rightly argues that labeling other individuals and cultures as insane has a dehumanizing effect, placing one’s own culture as superior to that being assessed. Wolf might respond to this by expressing that sanity compares to universal moral rules that are largely agreed upon and sought by many societies. However, this does not adequately address Moody-Adams’s point about the toll of labeling somebody as incapable of change. To that end, I believe that Moody-Adams’s view better addresses this point.

The third and final criterion relates to a difficult question in moral responsibility: if one’s deep self is governed by deeper selves, which are in turn governed by yet deeper selves, eventually leading to external factors, then none of us have control over who we are. How can we then be morally responsible? Wolf brings up this issue with the Deep Self View. Her solution is simply to accept it. By acknowledging that external factors are at play but allowing responsibility, Wolf defines the self within a deterministic world and provides an explanation for how we can be morally responsible: her notion of moral responsibility comes into play within the deep self, regardless of how that self was formed. It is, in her view, about living up to the potential of your deep self. While it may not be the only solution to the issue of determinism, addressing the problem is a significant benefit of Wolf’s view.

We have discussed the merits and drawbacks of both views and, while both contribute to a proper view of moral responsibility, it is clear that neither is complete. Now, I propose a new way of defining moral responsibility which I believe solves those problems which we have discussed.

III

It seems as though no meta-ethical formulation of responsibility can cover all examples without encountering problems such as dehumanization. As such, I believe that we can attain the same features of moral responsibility without the drawbacks by viewing responsibility as a condition of the self. Before fully explaining my view, though, I must introduce the idea of Platonic Forms^{ix}—this can be done most clearly through examples. First, imagine a triangle. While triangles can be seen anywhere in the world around us, what actually is a triangle? I can't reach out and touch the concept of a triangle—does it even exist? Plato answers that there is a pure idea of what a triangle is—a Platonic Form—and all triangles seen in the world are mere representations of that ideal. Other Platonic Forms include humanness and goodness. From this, let me construct the idea of *Platonic Values*. Suppose that somebody confronted with the trolley dilemma^x decides to let the five die by not getting involved. This decision is a reflection of the person's Platonic Value that killing is wrong. Now, imagine somebody who donates to the poor. This action is a result of the person's Platonic Value of equality. Essentially, Platonic

Values are a person's core values in their most pure sense.

From this, I construct my view on moral responsibility, which requires only one condition: a person is morally responsible if and only if their actions do not align with their Platonic Values. By restricting responsibility to the level of the self, this view escapes any issues of dehumanization or harmful labelings of cultural insanity. From now on, I will refer to this framework as the *Platonic Value View*. It is essentially Wolf's view but with the modification of living up to one's own Platonic values instead of restricting moral responsibility to the sane. Returning to attributes of responsibility and examples, we will see how this view fares.

As stated in section I, one of Wolf's most critical points is the necessity of self-revision. According to her, JoJo cannot justifiably self-revise because he is unable to see his faults. However, under this new framework, even JoJo can self-revise according to his core Platonic Values (which may be loyalty to his kingdom, continuing his father's legacy, et cetera). As previously stated, this view is quite similar to Wolf's in that it limits responsibility to when one knows what is right and fails to do it. However, it is less restrictive to account for Moody-Adams's criticism.

I now turn to the examples of JoJo and Ancient Greek slaveowners. Starting with JoJo—while we may disagree with his actions, they match his core values. Under the Platonic Value View, we see that JoJo is not morally responsible, matching our instincts, but we find this without dehumanizing him in the way which Wolf's view does. We say that he is following his core values rather than Wolf's description of him as culturally insane and incapable of change. Returning to Ancient Greek slaveowners, I argue that their

^{ix} From the Ancient Greek philosopher Plato (~425 BC - 347 BC)

^x A famous thought experiment designed by Philippa Foot. Essentially, the problem questions the difference in morality between killing and letting die. Philippa Foot, "The Problem of Abortion and the Doctrine of the Double Effect," *Oxford Review*, no. 5 (1967): pp. 5-15.

Platonic Values actually contradict the institution of slavery. Slaveowners likely believe in some form of equality or altruism among their peers; the problem is that they fail to extend this belief to all people. Platonic Values cannot be conditional because of their intrinsic purity—the wealthy should be respected is not a Platonic Value in itself; it is a corrupted representation of the core value that all should be respected. As such, slaveowners are acting in contradiction with their Platonic Values and are thus morally responsible. Again, this fits our instincts.

Finally, I will discuss the most problematic aspect of moral responsibility: determinism and the self. Wolf’s view tends to this problem in a very similar way, by saying that one’s lack of control has no bearing on their ability to self-revise according to their determined values. Similarly, our new framework resolves this by acknowledging its existence by admitting that we do have core values which are out of our control, but that living by those values, the basis for moral responsibility, is entirely up to us.

To conclude this section, I want to direct attention to one distinction: the Platonic Value View addresses metaphysical responsibility—not what one should be held responsible for in society. As seen in the previous section, practical responsibility can have other benefits.

In summation, the Platonic Value View, which argues that a person is morally responsible when their actions do not align with their core values, seizes the benefits of Wolf’s view without encountering the problems pointed out by Moody-Adams.

IV

We have now examined the arguments of both Wolf and Moody-Adams, and we have determined that the Platonic Value View

prevails. Up to this point, the question of determining necessary and sufficient conditions for moral responsibility has largely revolved purely around theoretical morality. It is clear that there are many people with good morals who fail to live up to them; this is because the will does not always directly follow from one’s morality. In closing, I would like to extend my thesis into the dimension of moral action. Specifically, the question I am addressing in this section is, *when can a person be exempted from moral responsibility for having good intentions?*

To answer this, allow me to introduce what I will refer to as the *Self-Victimization* case.

(1) Consider JoJo the Second, son of the evil dictator JoJo. Despite his upbringing, wherein his father and his grandfather view the torturing, killing, etc. of enemies to be the morally righteous thing to do, JoJo the Second knows that these actions are wrong; he is, by Wolf’s account, sane. Unlike his predecessors, his Platonic Values center around equality. However, he still cannot bring himself to stop these actions, which he knows are wrong. He is often heard exclaiming that he “simply can’t change” or that he wants to do what is right but cannot succeed. In this case, let it be that JoJo the Second truly wants to change but is for some reason incapable.

(2) Imagine a junior high bully, Blake McNeilson,^{xi} who feels bad about himself. He enjoys bullying and gets surface-level enjoyment out of other people’s pain. In contradiction, he has Platonic Values revolving around respect. His actions do not align with his core values, suggesting moral responsibility. However, Blake feels so bad about himself that he wishes he wanted to treat people better. He enjoys bullying and

^{xi} If you’re reading this, Blake, I still haven’t forgiven you.

does not honestly want to stop, but if he could at least have some desire to stop bullying, he would be able to sleep at night.

In these two cases, both JoJo the Second and Blake publicly consider themselves the victim of the situation. They both want (yet are unable) to change per what they, sanely, know is morally right. However, my intuition is that Blake shall be considered morally responsible for his bullying (specifically, for contradicting his Platonic Values), whereas JoJo the Second is not morally responsible for his malpractices. To articulate why, let us revisit the distinction drawn in section I between mere desires and volitions.^{xii} As stated, JoJo the Second's wish to change his ways is one he wants to be effective—it is a volition. Since this volition matches his core values, I believe that he is not morally responsible. Blake, however, does not truly want to quit bullying: he merely wants to *want to stop* bullying because he feels bad about himself. From this case, I draw the distinction: *a person may be exempted from moral responsibility due to good intent if and only if they have a second-order volition to act in accordance with their Platonic Values.*

One aspect of this view is called into question by the expressivist view of meta-ethics. During his discussion of meta-ethics, R. Jay Wallace in "Moral Subjectivism"^{xiii} explains the second-ordered expressivist view: the idea that saying "X is right" represents a desire that the desire to do X prevails (over other desires such as not doing X and doing Y). A valid critique of the expressivist view, as brought up by Wallace, is that there may be a case of endless higher-order desires, each undermining the last. To resolve this, we can go back to Frankfurt. He states that a decisive

desire nullifies the question of any higher-order desires. If there is only one second-order desire and it is undisputed, then there are no higher-order desires wishing otherwise. Now, it seems, there are two possible directions we may go. Either we can decide that a decisive volition of any order may exempt one from responsibility, or we must choose where to draw the line. While this question warrants a paper on its own, it is my belief that the line shall be drawn at decisive second-order volitions. While I am leaving this up in the air, this decision is because, as the order of the volition gets higher, the desire becomes more and more distinct from reality. Past the second order, volitions turn from an attempt to make something one's will into meaningless, endless mind games. Thus, we arrive at the final form of our analysis: *a person may be exempted from moral responsibility due to good intent if and only if they have a decisive second-order volition to act in accordance with their Platonic Values.*

To see the implications of this stance, I will examine some real-world cases where this would apply. I believe it is extremely difficult to reject the notion that many around the world, despite wanting equality, have biases that reflect racist, sexist, etc. instincts. Likely, much of this group would commonly be heard saying things such as, "I wish I could change, but I was raised this way." As in the case of JoJo the Second and Blake, this statement shows self-victimization—putting oneself in a position of having no control over their actions. Within this group of people, I would assume that there are two subsets: those who genuinely want equality but are fighting some barrier^{xiv} that they cannot overcome and those who want equality merely because that very

^{xii} Desires are merely the want to have a specific first-order desire; volitions entail the desire for a specific first-order desire to be the one that motivates the will.
^{xiii} *The Norton Introduction to Philosophy* (2018), 865.

^{xiv} For example, mental or environmental barriers; there are many reasons why it may be challenging for one to actively fight their instincts.

want puts them in good moral standing.^{xv} To the former, I say that if there is an actual volition to pursue equality (presumably a Platonic Value of theirs), but an insurmountable barrier is encountered, a person shall not be morally responsible due to their good intent. However, in the case of the latter, the person is morally responsible because no such volition exists.

V

As my theses have been fully developed, I will now turn to address concerns that the reader may have. I see two main objections to the views presented in this paper.

Objection 1: A Pure Evil^{xvi} Being. This concern is broken into two cases: one wherein the evil being successfully does the evil action in question, and one wherein they do not. To the former, a critic would question how it could be that even a pure evil being is not morally responsible for their wrongdoings^{xvii}. As for the latter, they might contend that this view implicates that an evil being should be morally responsible for not being evil.

Starting with a successful evil being; while the critic would insist that they must be morally responsible, I argue that it depends on the given situation. If the being has concealed good platonic values and does evil, then I agree that they are morally responsible. However, in the case described, regarding a *pure evil* being, they should not be morally responsible because they are doing good

^{xv} The meaning behind good moral standing is left intentionally vague. It could mean wanting to have moral desires to live with oneself (as in the case of Blake), it could mean wanting to be perceived as righteous by others, or it could mean something else entirely.

^{xvi} I use this term cautiously, as evil could be conceived as relative to the self. In this context, I am referring to actions which strongly contradict the common values of society.

^{xvii} Once again, I am using the term wrongdoings to depict an action which contradicts societal norms.

according to their most fundamental, unconditional beliefs. This will be considered further in the second objection.

Now, moving onto the case of an unsuccessful evil being. The critic pointed out that a consequence of this view suggests that an evil being should be responsible for not doing evil. I see two cases within this. In the first, the being cannot bring themselves to do the evil act because of hesitation in them; this provides us with hint that they truly have non-evil Platonic Values, and by avoiding the act in question they are living up to those values.

The second case is messier. Imagine that it is truly a pure evil being and they are simply failing to perform the act in question for whatever reason. Further, presume that the being has a decisive second-order volition to be evil and fails to live up to it. In this case, they are not responsible due to our rules surrounding moral exemption. An educated critic may draw attention to the other case, where somebody is purely evil but has no volition to follow their values. Aren't they, according to this view, morally responsible? To this, I respond that there are few, if any, of such cases. There are a multitude of times when a good person may act in malice because it is to their advantage, such as stealing or lying. Fewer cases exist, however, wherein a pure evil being neglects their values and does good for their own advantage. Further, any one of these cases is in itself contradictory; if an evil being does good to their own advantage, they are not doing it out of goodness but rather out of greed. Thus, no such case, a purely evil being whose volitions are good, exists.

Objection 2: Objective Moral Facts. It may be argued that my view is incompatible with objective moral facts. To this I say that, even if universal moral laws exist, it would be unjust for a person to be judged according to

them without being made aware of their existence (see footnote viii, p. 5). Since we have not discovered or been instructed on an agreed-upon set of moral laws, only two possibilities remain. Either somebody's Platonic Values align—by chance^{xviii}—with objective moral laws, in which case my stance does not differ from the universalist, or somebody's Platonic Values do not align. In the latter case, it is my intuition that the person shall still be judged according to the Platonic Value View; they should not be at a moral disadvantage if their Platonic Values, over which they have no control, do not align with objective moral laws, which they do not know.

I now believe that I have exhaustively addressed the most prominent objections to the views presented in this paper.

VI

Although questions such as the definition of the self, the meaning of moral responsibility, and when to truncate ordered desires remain, I believe that we have thoroughly covered the conditions for moral responsibility. We started by introducing the Wolf's Sane Deep Self View and Moody-Adams's criticisms. Moving onto evaluating both arguments, we found that they both had merits but fell short in certain areas—Wolf's view leads to dehumanization, whereas Moody-Adams (while providing valuable critiques) does not provide conditions for responsibility. To achieve the benefits of both views without the drawbacks, we introduced the Platonic Value View, stating that a person is morally responsible when their actions fail to align with the purest form of their core values. Finally, we considered exemption due to good

intent and decided that somebody may be exempted from moral responsibility if they have a decisive second-order volition to act in accordance with their Platonic Values.

^{xviii} I stress that this is by chance to convey that a person whose Platonic Values *happen to* align with objective moral laws does not receive any sort of moral reward for said alignment.