

**A TYPOLOGY OF BAD CHARACTERS:
UNDERSTANDING MORAL BADNESS AS MENTAL ILLNESS**

by

Aidyn Catherine Cooper

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree of

Master of Arts
in Philosophy

at

The University of Wisconsin-Milwaukee
August 2023

ABSTRACT

A TYPOLOGY OF BAD CHARACTERS: UNDERSTANDING MORAL BADNESS AS MENTAL ILLNESS

by

Aidyn Catherine Cooper

The University of Wisconsin–Milwaukee, 2023
Under the Supervision of Professor William Bristow

This paper presents Susan Wolf's theories on freedom and responsibility. It places special emphasis on her Reason View as presented in *Freedom Within Reason*¹. I analyze three types of bad characters, where a "bad character" is defined as someone with a stable and pervasive pattern of acting badly. I argue that Wolf's Reason View entails that bad characters are psychologically incapable of doing the right thing for the right reasons. Therefore, according to the demands of Wolf's Reason View, we cannot hold them responsible for their actions. This spells trouble: aren't bad characters precisely the type of people that we should feel empowered to hold responsible? Wolf's view accurately predicts that bad characters should not be held responsible for the fact that they lack a psychological ability to do the right thing for the right reasons. However, Wolf's Reason View fails in its insistence that this ability is what moral responsibility requires. I argue that moral responsibility requires an ability to take responsibility for one's own health. I provide a re-conceptualization of moral badness as mental illness. Understanding bad characters as mentally ill allows us to hold them responsible for their own health. This vision of responsibility, however, must be partly constituted by the convictions of one's society.

¹ Wolf, Susan (1990). *Freedom Within Reason*. New York: Oup Usa.

TABLE OF CONTENTS

I.	Introduction	1
II.	Wolf's Reason View	3
III.	A Psychological Ability to Do Otherwise	9
IV.	Bad Characters	16
V.	Moral Badness as Mental Illness	28
	REFERENCES	36

I. Introduction

I present Susan Wolf's theory of freedom and responsibility. She calls this view the Reason View. In her book, *Freedom within Reason*, Wolf argues that freedom and responsibility requires a psychological ability to do the right thing for the right reasons. This ability is cashed out in a number of ways: as an ability to act in accordance with "the True and the Good²," as an ability to act in accordance with "true beliefs and good values³," and as an ability to act in accordance with "Reason⁴." In response to Wolf, I argue that her Reason View has the consequence that we can't hold bad characters responsible for their actions. I define a bad character as someone who has a stable and pervasive pattern of acting badly.

In the second section of this paper, I will present Wolf's Reason View as well as two alternative theories of freedom and responsibility that Wolf rejects. These are the Autonomy View and the Real Self View. A brief discussion of these alternative views is necessary to uncovering why Wolf views an ability to act in accordance with "Reason" as the only worthwhile candidate for establishing our freedom and responsibility.

In the third section of this paper, I will dive into the details of what Wolf's view entails. Wolf argues that freedom and responsibility requires an "ability" to act in accordance with the True and the Good. In this section, we will examine closely what this ability consists in. As it turns out, determining whether someone has this ability (or not) comes with a certain degree of epistemic limitations.

² Wolf, Susan (1990). *Freedom Within Reason*. p. 95

³ *Ibid.*, p. 131

⁴ *Ibid.*, p. 65

In the fourth section of this paper, I utilize Wolf's notion of ability to examine the freedom and responsibility of three bad character types. These three bad character types are: the villain, the malmaton, and the akratic. Consequently, Wolf's Reason View would have us accept that none of these three bad character types are free and responsible actors.

In the final section of this paper, I aim to soften the consequence of Wolf's Reason View by suggesting an alternative approach to understanding the freedom and responsibility of bad characters. I argue that treating bad characters as mentally ill allows us to hold them responsible for their own health. This vision of responsibility, however, must be partly constituted by the convictions of one's society.

II. Wolf's Reason View

Wolf presents her theory of free will and moral responsibility in two major works. Her ideas are first developed in a paper titled, *Sanity and the Metaphysics of Responsibility* and later in a book titled, *Freedom Within Reason*. This paper will primarily focus on her theory as it is fleshed out in her book. However, some references will be drawn to her earlier work as is useful for the presentation of my project. Remember, my goal is to show that Wolf's theory has the consequence that we cannot hold bad characters responsible for their actions. Let's begin by setting up Wolf's theory of free will and moral responsibility as presented in *Freedom within Reason*.

In chapter four of Wolf's book, *The Reason View*, she sets out to provide a fresh perspective on free will. She argues that free will is an ability to act in accordance with the

“True and the Good”.⁵ She comes to this realization only after settling that the two alternatives to her view are failing in some significant respect. She argues that what we are looking for is neither mere autonomy nor an ability to act in accordance with our deepest values. What freedom and responsibility requires is the ability to act in accordance with Reason. It is ‘reason’ with a capital ‘R’ because it is *right* reason; born out of an ability to form true beliefs and good values. Going forward, “Reason” (capital ‘R’) refers to one’s ability to reason in accordance with truth and goodness; while “reason” (lowercase ‘r’) refers to a common cognitive capacity to deliberate among many alternatives. It does not have the special normative component that Reason has.

The Autonomy View, according to Wolf, is the view that states that free will requires the “ability to do otherwise”.⁶ This view poses free will as being directly incompatible with the laws of cause and effect. Those laws, upon reflection, would make it the case that we are committed to doing all and only that which we are caused to do. If I am “able to do otherwise” then, it would seem, I must be metaphysically independent from these causal laws. I must not be determined to act in virtue of them but rather in virtue of some other power.

What ignites us in thinking that we could possibly possess a power of this sort? Perhaps it is that we sometimes, unlike other animals, are able to go directly against our appetitive inclinations. In the face of desire we are able to restrain ourselves; often using a deliberative capacity called reason. This capacity for reason is typically understood as being what separates us from lower-order animals. Being capable of acting on reasons and not on whims, desires, and impulses is what makes us free. This power of reason, then,

⁵ Ibid., p. 71

⁶ Ibid., p. 69

must be the feature which renders us metaphysically extraordinary; capable of thwarting the laws of cause and effect and capable of paving a path of our own. However, although the Autonomy theorist may want to claim this fact about reason, they cannot.

Autonomy cannot be the ability to act in accordance with reason. After all, this ability to reason must come from somewhere. It comes from being born with a particularly adept cognitive capacity and perhaps from learning the connections between reason and action utilized by our parents and educators. Furthermore, what reasons we recognize as being reasons worthy of acting likewise must have originated in our past. For example, as children we are told explicitly what we should and should not do. Therefore, as a moment of thoughtful analysis demonstrates, our capacity to reason does not apparently arise beyond the laws of cause and effect. It arises in tandem with them – along with the rest of nature.

Wolf concludes that what the Autonomy theorist must really be after – if what they are after is genuine autonomy – is the ability to act on no basis whatsoever. This is what Wolf describes as “the ability to make radical choices”.⁷ What the Autonomy theorists should say, if they wish to take autonomy seriously, is that free will gives us an ability to thwart the commands of reason. While I may be capable of recognizing what it is most rational for me to do, this is not what matters for my responsibility according to the Autonomy View. What matters is my ability to also be irrational. If I am to maintain the ability to do otherwise, I must not be determined to act in accordance with reason. Even if what reason demands is so obvious and forceful, I must still be able to do otherwise.

⁷ Ibid., p. 53

Therefore, what bona fide autonomy demands, is not an ability to act on the basis of reason but rather on no basis whatsoever. In chapter 3, *The Autonomy View*, Wolf discusses what such an ability might look like as well as how it might affect our judgements concerning moral responsibility. Ultimately, she concludes that an ability to reason (rather than an ability to make radical choices) more directly corresponds with our ordinary intuitions concerning freedom and responsibility.

Leaving the Autonomy View behind we can examine a second view that Wolf rejects in favor of her Reason View. Picking up where we left off, we found that a capacity for reason might be a suitable candidate for the instantiation of our freedom and responsibility. Would this mean, then, that whenever we exercise our reason we are acting freely? The following example might prevent us from believing that this is so. Consider an alcoholic who is trying with all of his might to maintain his sobriety. In a moment of weakness, perhaps after a fight with his partner, he may reason that one drink won't make or break his resolve. He may convince himself that after that one drink, he'll return to being sober again. Assuming that the alcoholic's condition is severe enough to be a so-called addiction, it is not clear that he is exercising his reason freely. If the alcoholic were free, if he truly had control over his own will, then he would stay sober. Especially if staying sober is what he deeply values, what he wholeheartedly wants for himself, and what he (in a healthy state of mind) is committed to.⁸

The second view, the Real Self View, seeks to capture our intuition that the alcoholic in the above example is not free. This view states that the alcoholic is not free insofar as he

⁸ The contrast between what one values and what one, in fact, does is owed to Gary Watson (Watson 1997) and will be explicated further along in this paper. The contrast between what one does and what one is wholeheartedly committed to is owed to Harry Frankfurt (Frankfurt 1987).

is unable to act in accordance with what he deeply values. This ability to act in accordance with one's deeply held set of values is the origin of our responsibility insofar as we cultivate these values utilizing our capacity to reason. Therefore, we are not free when we are pulled by especially strong desires – as in the case of the alcoholic. We are free when we act in accordance with what we have formerly judged to be the right set of values.

To illustrate the Real Self View, Wolf presents Gary Watson's position on free will.⁹ Watson held that our free will arises out of our ability to make choices that arise from a system of valuation as opposed to a merely motivational system. Our motivational system is made-up of our desires, inclinations, and dispositions. It is what pushes us toward a particular course of action. As human beings, capable of free choice, we are able to place our motivational system under the control of our valuational system. Our valuational system comprises our judgments about which of our desires, inclinations, and dispositions are good and which ones are bad. That we view some aspects of our motivational system as good and others as bad causes us to identify ourselves with some and denounce others. So, it is said that those aspects of myself that I judge as good, and with which I identify myself and fully endorse, comprise my Real Self.

Returning to the case of the alcoholic. It would seem that he judges his alcoholism, and his drive toward it, as bad. It is an aspect of his motivational system that he rejects. The alcoholic does, however, possess a system of values that runs in direct conflict with his desire for alcohol. His system of values tells him that it is better to be sober. He really wants to be sober, in fact. The alcoholic is able to utilize his ability to reason in order to determine what it is best for him to do. Yet, the alcoholic is missing an essential ingredient required

⁹ For a fuller account of Gary Watson's position on freedom and responsibility see Watson (1997).

for free will on the Real Self view. He lacks the ability to act in accordance with his system of values. Therefore, as intuition suggests, free will requires not only the possession of a valuational system but likewise, a motivational system that is capable of submitting to one's prejudged set of values.

Now that we have a grip on what it means to be free according to the Real Self View, we can examine Wolf's rejection of this view. This will have the added benefit of leading us closer to understanding Wolf's own view – which is our ultimate goal. What troubles Wolf about the Real Self View is the fact that we appear to lack ultimate control over who we turn out to be. Whether I possess the sort of character that is capable of developing the right sort of values which make me a good person might not ultimately be up to me. Remember, all that is required for freedom and responsibility on the Real Self View is that I be capable of forming my own set of values utilizing my ability to reason. Recall also that our ability to reason is formulated in conjunction with our previous experiences. The types of reasons that appeal to me might not be good reasons; and they may be based on false beliefs.

As a case and point instance of this possibility Wolf references the infamous “Son of Sam” murders.¹⁰ Serial-killers appear to be agents who are capable of cultivating a unique set of values, and likewise capable of acting in accordance with it, and yet, whose reason leads them to cultivate the *wrong* set of values. Perhaps these psychopathic reasoning processes are caused by trauma in one's past or a defect in one's psychology. Regardless of the cause, it is unclear whether someone can be held responsible for their own poor judgment. The majority of us, it would seem, follow a relatively normal trajectory in

¹⁰ Wolf, Susan (1990). *Freedom Within Reason*. p. 37

becoming who we are. And, as such, are endowed with a capacity to “see and appreciate the True and the Good” and to constitute our value systems in accordance with it. It would seem, however, that individuals who encounter severe trauma in their past, or are victims of severe mental illness, are not capable of developing this same capacity.

It is precisely this line of thinking that leads Wolf to claim that freedom and responsibility requires an ability to see and appreciate the True and the Good. Moreover, freedom and responsibility requires that one is able to act in accordance with this appreciation. If someone is capable of forming true beliefs and good values, then they are capable of developing a valuational system that corresponds with those beliefs and values. One can still lack free will, as in the case of the alcoholic, when they are apparently unable to control their desires in light of their values. Assessment of this ability is going to be crucial to our determination of whether we can hold repeat-offense bad actors responsible for their bad actions.

A “repeat-offense” bad actor is anyone who has a stable and pervasive pattern of acting badly. These are who we will henceforth refer to as “bad characters”. In section four of this paper we are going to take up an analysis of three bad character types and show that we can’t possibly hold them responsible on Wolf’s Reason View. It’s doubtful that Wolf, herself, would contest that judgment. Her view is that freedom and responsibility is established by one’s ability to do the right thing for the right reasons. As one might expect, someone described as having a stable and pervasive pattern of bad behavior will likely lack such a capacity. This apparent truism, however, is not nearly as obvious when examining a spectrum of bad character types.

To understand why determining responsibility is such a great difficulty on Wolf's Reason View, we must familiarize ourselves with the epistemic limitations of her view. In order to determine whether someone is capable of doing the right thing for the right reasons we must make a handful of inductive judgments. For instance, we may look to their past behavior as some indication that it is difficult for them to behave well. We may also look at facts about their past experiences: how they were raised, whether they were cared for and supported, or abused and neglected. As it turns out, when we look for answers to explain why someone behaves the way that they do, we are able to churn out quite a bit of evidence. However, despite our talents at drawing connections between someone's past and expectations about their character; these judgments are ultimately uncertain.

Wolf's Reason View claims that freedom and responsibility requires that one be able to do the right thing for the right reasons. This view has a huge epistemic limitation baked into it. Namely, that we — as fellow arbiters of moral responsibility — have no direct access to this fact about others. It's not even clear that we have direct access to this fact about ourselves! Therefore, we must rely on the inductive judgments that we can make about others based on the facts we have about *who they are*. How these inductive judgments are made will be fleshed out in Section III.

III. A Psychological Ability to Do Otherwise

In *Freedom Within Reason*, Wolf argues that what is required to hold someone responsible for their bad actions is the presumption that they could have acted well — even though they did not. Our purely physicalist descriptions typically do not allow for this possibility. A reductionary account of these events would have us admit that if someone did

not do the right thing for the right reasons then they could not have. The ability to do otherwise is removed.

We may point to broader explanations in making the case that someone could not have done the right thing for the right reasons – when they, in fact, didn't. For instance, perhaps the right thing to do is to answer my friend's phone call when I'm aware there is an emergency. But, perhaps I didn't detect the urgency of her text messages and sent her to voicemail instead. Or, perhaps I don't hear the phone ring because the neighbor's are blasting their music too loudly. Either way we flesh out this scenario, what we are implying is something like the following: given the laws of cause and effect and the fact that we, as human beings who are subject to these laws, anyone's failure to do the right thing for the right reasons must have been caused by something.

This description is at the core of the dilemma we have found ourselves in the middle of. That is, the dilemma of determining how it is possible to hold people responsible for their actions if there is always some external causal factor at work. Wolf's position on freedom and responsibility is a compatibilist position. As such, Wolf maintains that our freedom and responsibility is compatible with the deterministic laws of cause and effect. However, Wolf is not prepared to concede that all of our rational choices are reducible to those laws. The laws of cause and effect are physical laws and a rational choice is a *psychological* phenomenon.

Wolf adopts a non-reductionist picture of freedom and responsibility. Insofar as one's freedom is constituted by a psychological ability – that psychological ability need not be reducible to physical facts about one's biological organism. The primary push behind Wolf's non-reductionist position is to show that when we discuss matters of responsibility,

we can do so without making reference to the physical laws. We can make sense of one's freedom and responsibility by appealing to all and only psychological facts about the person's past and present life. For this reason, in discerning what sort of "ability" is required to assess whether someone is capable of doing the right thing for the right reasons, Wolf focuses on the notion of *psychological ability*. The story of Rose, presented in chapter five of *Freedom within Reason*, illustrates Wolf's conception of psychological ability.

Rose is a professor who, one night, for no particularly strong reason, decides to watch *The Philadelphia Story* on TV – for the 7th time. She chooses to do this instead of grading papers. The story aims to show, in what sense, Rose could have graded papers despite the fact that she did not. According to the story, there is at least one obvious sense in which Rose could not have graded papers. Assuming, as the story does, that what Rose does in fact do is entirely determined by physical laws outside of her control, Rose couldn't have graded papers. That is because in *this* world, the world in which this story is taking place, Rose chooses to watch TV instead.

However, what happens according to the physical laws is not the whole story. Since this story posits that there are psychological laws at work, as well. According to the psychological laws, Rose *could* have graded papers – even though she did not. Prior to the point at which Rose decides to watch *The Philadelphia Story* it is perfectly compatible with Rose's psychological history and all the relevant psychophysical laws that she does otherwise. Consider this excerpt from Wolf's chapter, *Ability and Possibility*:

“. . . Rose is not coerced or hypnotized; that no addiction, obsession, or pathological aversion is involved. Rather, Rose just thinks about whether to watch TV or grade papers. It is compatible with her psychological history up to the moment in question, in conjunction with all the psychological and psychophysical laws that apply to her, that she choose TV, and

it is also compatible with these that she choose to grade papers. As it happens, on the evening in question she chooses TV.”¹¹

Given that Rose was clearly not constrained in some psychologically overbearing way (as if she had been addicted or hypnotized), Rose is thought to be “psychologically free” on Wolf’s view.¹²

At this point, we should have a sturdier grasp on what it means when Wolf claims that Rose was “psychologically able” to grade papers – even though she did not. Furthermore, we should have a more firmly cemented understanding of what it means to make a psychologically free and unconstrained choice. Based on the resources Wolf has given us, I’d like to consider in more detail how we might determine whether someone has a “psychological ability” to do the right thing for the right reasons. Most importantly, this will help us assess one’s psychological ability in cases when they do not actually *do* the right thing for the right reasons.

Right off the bat, we may grant Wolf that people who tend to do the right thing for the right reasons are psychologically able to do the right thing for the right reasons. The trickier question is: “Were they able to do the right thing for the right reasons – even though they did not?” Given the criteria that Wolf has provided us with for discerning psychological abilities I’m tempted to answer, yes. Or, at least, admit that it is readily conceivable. Perhaps the easiest way to lend her this point is to use ourselves as test subjects. If you conceive of yourself as someone who is – for the most part – a person with good character; then, you may be able to think of a time where you could have done the right thing, and yet didn’t. You may attribute your moral failing to a number of factors –

¹¹ Ibid., p. 108

¹² Ibid., p. 94

negligence if nothing else – and yet, despite these factors, it was compatible with *who you are*, moreover, that you could have acted better. For this reason, I will assume that what I call “good characters” (or individuals who have a stable and pervasive pattern of acting well) are free and responsible on Wolf’s view.

Ultimately, my aim is to show that Wolf’s view fails in allowing us to hold “bad characters” responsible for their actions. Again, bad characters are individuals who have a stable and pervasive pattern of acting badly – if only in isolated circumstances. It is also part of my project to convince the reader that this outcome – namely, the fact that we can’t hold bad characters responsible for their actions – is not as unintuitive or unappealing as it may at first seem. However, before reaching a full assessment of the consequences, let’s ensure that we are clear on the notion of psychological ability, first. In particular, let’s get clear on what we mean by a “psychological ability to do the right thing for the right reasons”.

We might say, for instance, of someone we generally recognize to be of a morally-upright nature that they could have done the right thing even when they fail to. For instance, consider when a close friend cancels on your lunch plans at the last minute. We may hold them responsible at the psychological level of explanation. We may claim, albeit ambiguously, that, given their beliefs about politeness norms, and so on, they should have recognized the moral irresponsibility of canceling plans at the last minute. We might, at some point, learn more about our friend’s situation which excuses this minor transgression. But, as the facts are presented to us at this stage, we can hold them responsible with respect to a *prima facie* recognition of their psychological abilities.

The point I'm trying to convey is that we typically don't appeal to any physical facts when holding our friend's morally responsible for their actions. Perhaps it is true that, at the highly convoluted physical level, our friend was both physically and psychologically unable to commit to our lunch date. Perhaps she was unable to commit to our lunch plans given a complex set of neuronal firing that she had no conscious awareness of. This pattern of firing was, as the story goes, caused by a handful of epistemically inaccessible features about her distant and most recent past.

The availability of such explanations, however, hinges upon our acceptance of a physical reductionism that Wolf herself resists. Indeed, a rich philosophical history suggests that we have good reason to question whether such an explanation like the one above is plausible. Therefore, if we admit that psychological abilities are not always reducible to physical abilities we can reasonably make sense of our friend's being responsible for canceling our lunch plans at the last-minute. We can do this, that is, without appealing to any unnecessary or high-falutin physical descriptions.

Returning to the story of Rose we can appreciate Wolf's tactics in assessing why our intuitions lead us to understanding, in what sense, Rose could have graded papers even though she did not. What is crucial in determining Rose's ability to do otherwise in this case is our epistemic access (fabricated as it may be) to Rose's "psychological history." On the subject of a person's psychological history Wolf has very little to say. She does, however, state that for Rose's choice to be "psychologically free" then it must be the case that it is (a) compatible with Rose's psychological history that she could have graded papers, and (b) compatible with all of the psychological and psychophysical laws that Rose could have graded papers.

If all of these facts about Rose, and the relevant laws hold true, then Rose is responsible for choosing not to grade papers. Wolf utilizes this notion of “psychological history” without much explication of it. Presumably she expects it to be a pre-theoretical notion that bears no lengthy explanation. For our purposes, a succinct definition certainly couldn’t hurt us. Going forward, one’s psychological history can be understood simply as that which can be reasonably said about someone’s distant and recent psychological past. What is significant, psychologically speaking, may be one’s past behaviors, past experiences, and facts about their past and present emotional and cognitive lives.

Consider, again, the case of a close friend canceling on your lunch plans at the last-minute. Perhaps I am aware that this friend has a habit of canceling plans at the last-minute. I can conclude that it is, for whatever reason, difficult for this friend to follow through on commitments. This behavior suggests that it might be incompatible with my friend’s psychological history that she could have done otherwise. That is, when she cancels our lunch plans it’s incompatible with her psychological history to say that she could have followed through — since, she usually doesn’t. However, determining whether one’s psychological history is compatible with their ability to do the right thing for the right reasons isn’t (in reality) this straightforward.

It can be, I think, made more straightforward when we have more than merely behavioral accounts of the agent under analysis. For instance, perhaps it is quite easy to conceive of our friend as lacking the psychological ability to do otherwise when we also know that she struggles with depression. We now know not just that she usually cancels last-minute but also *why* she usually cancels last-minute. Here she seems both less able to follow through on lunch commitments as well as less responsible for canceling them.

It bears repeating here that Wolf's view relies upon our epistemic access to the psychological histories of others. Wolf's claim is that one's psychological history contains the secrets that will express one's psychological ability to do the right thing for the right reasons. The more we know, the more we can assess this ability. That is, the more robust our psychological histories of agents are, the more easily we can determine whether they have a psychological ability (or *inability*) to do the right thing for the right reasons.

IV. Bad Characters

Wolf has given us the resources to assess moral responsibility according to her view. We now know how to spot a psychological ability to do the right thing for the right reasons. This will be useful as we undergo our analysis of three bad character types. Our project, in this section, is to take a closer look at three types of bad characters and ask the following question: according to the Reason View, are they morally responsible for their actions? What responsibility requires, on Wolf's view, is the psychological ability to do the right thing for the right reasons.

Importantly, Wolf anticipates the concerns I hope to lay bare. Presumably she is comfortable with the idea that good characters sometimes do bad things. But, insofar as good characters have the ability to do the right thing for the right reasons, they should be held responsible for their actions. It is perhaps less likely, however, that Wolf would be willing to admit that bad characters are responsible for their bad actions. In a sense, aren't these persons precisely the individuals who Wolf is trying to rescue from responsibility; insofar as they weren't necessarily involved in their own *becoming* bad?¹³

¹³ Bad characters are not involved in their own becoming bad insofar as the events that led to their becoming bad were not up to them. In other words, they were not responsible for the events that set

To remind ourselves of these intuitions that inspired Wolf's general project, let's return to her earlier work, *Sanity and the Metaphysics of Responsibility*. In this paper, Wolf introduces the character of Jojo. Jojo is the son of an evil dictator who, raised in his father's likeness, adopts an utterly skewed set of values. Given the extremely isolated and corrupt circumstance of Jojo's upbringing, it's unclear that Jojo is responsible for the man that he becomes. The case is illustrated as follows:

"Jojo is the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father's special feelings for the boy, Jojo is given a special education and is allowed to accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little Jojo takes his father as a role model and develops values very much like Dad's. As an adult, he does many of the same sorts of things his father did, including sending people to prison on the basis of whim. He is not coerced to do these things, he acts according to his own desires. Moreover, these are the desires he wholly *wants* to have. When he steps back and asks, 'Do I really want to be this sort of person?,' his answer is resoundingly, 'Yes,' for this way of life expresses a crazy sort of power that forms part of his deepest ideal."¹⁴

Wolf posits the case of Jojo as a refutation of the "Deep Self View" concerning free will.¹⁵ The Deep Self View is merely an earlier iteration of what she later compiles under the title of the Real Self View; of which we are already familiar. This view holds that so long as someone is capable of acting from their "Real Self" then they can be held responsible for their actions. Furthermore, acting out of one's "Real Self" constitutes one's acting *freely*.¹⁶

To re-illustrate these points briefly, let's return to Gary Watson. Watson argues that insofar as we human beings are capable of establishing a system of values, and capable of acting in

their reasoning processes askew and made them unable to reason in accordance with truth and goodness. They were, however, involved in their own becoming bad insofar as their own deliberative processes are what led them to being bad; to maintaining and endorsing a set of values that condones bad behavior.

¹⁴ Wolf, Susan (1987). *Sanity and the Metaphysics of Responsibility*. pp. 53-5

¹⁵ *Ibid.*, 50

¹⁶ For the purpose of my project we can assume that "deep self" and "real self" may be used interchangeably; though Wolf herself may have her own reasons for wanting to distinguish between the two.

accordance with those values, we are capable of acting freely.¹⁷ However, as Wolf points out, Jojo's values are utterly askew given his deeply depraved upbringing. Simply, he values that which he ought not to value; activities like "sending people to prison on the basis of whim".¹⁸

It is our own compassionate tendency that allows us to recognize that, given Jojo's unusual past, he is not responsible for his own Real Self – and the actions which that self endorses. Jojo is not responsible, according to Wolf, because he is *insane*. Sanity on Wolf's view contains both a cognitive and normative component. It is the first condition of her Reason View: "the ability to see and appreciate the True and the Good."

We are already familiar with the age-old insanity defense. This is the defense utilized in court systems to suggest that the defendant is not guilty for the charges that are being issued against them because they are insane. And insanity, as it were, excuses one from moral responsibility. This is the exact intuition that Wolf's earlier view – and one might argue her later view, as well – hangs on. Wolf's argument in *Sanity and the Metaphysics of Responsibility* is that Jojo, and others like him, are not responsible for their actions (and not free agents) because they are insane.

Insanity, however, does not arise out of nowhere. At the root of insanity is typically a pathological presence which disrupts the individual's moral upbringing. In the case of Jojo, that presence is his morally depraved childhood. As in the case of many psychopaths, it may be a severely traumatic experience or a psychological defect. It is these intuitions that drive

¹⁷ According to Wolf, an agent is free if they have the capacity to act in accordance with their Real Self. One can still be free and not act in accordance with their Real Self; so long as they were able to act in accordance with it. For more on this see Wolf (1990).

¹⁸ Wolf, Susan (1987). *Sanity and the Metaphysics of Responsibility*. p. 50

Wolf away from the Real Self View. The development of one's Real Self may, to no fault of one's own, have been pathologically disrupted.

If we understand insanity as having a normative as well as a cognitive component, then we are pushed toward thinking of all bad characters as insane. However, one should stay fixated upon the fact that it is not just the fact that these bad characters are insane, it is that they are insane due to the presence of a pathology. Their development was pathological in a way that the rest of ours was not. Each of the bad characters we will discuss are therefore mentally ill.¹⁹ In being mentally ill they are psychologically unable to do the right thing for the right reasons. However, I encourage the reader to make that judgment on their own.

Bad Character #1: The Villain

Villains, our first bad character type, can be spotted by a recognizable pattern of bad behavior. They are likewise the biggest proponents of their own bad actions. The most obvious case and point of "villainy" is Wolf's character, Jojo.

Given the robust knowledge we have about Jojo's psychological history, we can very quickly and easily apply our Wolfian analysis to determine if Jojo is morally responsible for his actions. Although Jojo shows up day-in and day-out as a horrible leader to an undeserving country, could he do otherwise? May he one day wake up and decide to follow a path of righteousness as opposed to constantly partaking and pleasuring in cruelty? On the face of it, not without a miracle.

¹⁹ More will be said about the connections between moral badness and mental illness in the final section of this paper.

As was already discussed above, Jojo is insane. He is insane not only because he commits awful and abhorrent acts, but because he commits these acts as a result of a depraved childhood. I return to this consideration because it will be crucial to squaring our Wolfian determination of moral responsibility with our ordinary intuitions. It is the fact that Jojo has the kind of upbringing that he does which helps us to understand the ways in which he is not responsible. It is our ability to point to external factors which were outside of Jojo's control (e.g. that he is the son of a sadistic ruler) that solidifies our judgment. Indeed, these are the "external factors" which tend to grip us most when grappling with the issue of Determinism.

Villains, unfortunately, do not exist in story books alone. Certain serial killers who don't show remorse for their actions are villains. Leaders of organized crime who reap large monetary rewards at the expense of others are villains. Other commonplace villains include scam artists, sex-traffickers, politicians, and more. Being a villain is more than simply behaving badly. It requires more than an ability to turn a blind-eye to goodness and common decency. It is a relishing in and full endorsement of one's bad actions.

Let's return to what I have said about the existence of a pathological presence lurking in the past of every bad character. On this point, I must admit that I have to rely on my own intuitions and the intuitions of philosophers like Susan Wolf. I am likewise relying on the positions of Ancient Greek philosophers and Neo-Aristotelian philosophers, as well. In the final section, I will return to discuss these influences in greater depth. For now, I can only hope to convince the reader with a few brief remarks.

Consider the case of Edward. Edward is a bad scientist, both evil and incompetent, who traps small furry animals in order to test his highly acidic chemical concoctions.

Imagine that we have full access to Edward's entire psychological history. Everything about his life seems perfectly normal. His parents supported and cared for him. He performed well, perhaps not extraordinarily, in school. And, he fits no obvious diagnostic-criteria for a psychiatric disorder.

I am prepared to toss this case out with the incomprehensible. When we encounter bad characters like Edward we do not stop short of asking "why?". Bad behavior begs to be explained because it falls beyond the scope of normal expectation. It can also, at times, be downright irrational. It is apparently irrational to be bad, for instance, when one's bad behavior makes it impossible for them to make or maintain stable relationships, or winds them up in jail. Furthermore, our intuitions, the majority of our media representations, and developments in contemporary psychiatry would suggest that every bad character has a backstory.

I ask the reader to keep these judgments regarding bad characters in mind as we move on to discuss two more bad character types: the malmaton and the akratic.

Bad Character #2: The Malmaton

This second category of bad characters comprises agents who apparently lack the requisite cognitive faculties necessary for doing the right thing for the right reasons. Unlike the villain, they do not fail to do the right thing due to a skewed set of values. Rather, they appear to lack a moral sense altogether. It is in virtue of lacking this ability to form moral judgements that makes them psychologically unable to do the right thing for the right reasons. These individuals may be incapable of reasoning in accordance with what morality demands due to the severe progression of a psychiatric illness. Or, they may be caught in

the midst of a psychotic episode and incapable of making rational decisions on the basis of reality. It is possible that neurodegeneration and psychosis can lead one to commit violent or malicious acts against innocent persons.

A malmaton, similar to the villain, has a stable and pervasive pattern of acting badly. However, contrary to the villain they do not appear to be acting in correspondence with a Real Self. The pathological presence that gives rise to their psychological inability to do the right thing for the right reasons affects their entire valuing infrastructure. In the case of the villain, the pathological presence at work affects only the values they affirm as good. In the case of the malmaton, their ability to form value-based judgments is altogether disrupted. A Wolfian analysis of moral responsibility determines that the malmaton is incapable of doing the right thing for the right reasons.

This group of bad characters encompasses paradigmatic mental illness. They are not merely normatively insane, like our villains, but diagnosably-so. When it comes to letting bad characters off the hook for their bad actions, this should be our easiest pill to swallow. The following example provides an illustration of the character-type, malmaton:

Alma's Alzheimers

Alma is the mother of Marie and grandmother of Marie's two young boys. Marie recently had to find a nursing home for her mother because Alma was punishing her two grandsons for miniscule, and by any standards, forgivable transgressions. They were spanked for crumbs on the counters, scolded for hair in the sink, pinched for talking too loudly, and sent to their bedrooms for no reason at all. Alma's Alzheimer's Disease had made her callous, impatient, and downright mean.

Alma's cruelty is not representative of who she was prior to the progression of her mental illness. Nor, does she appear to be acting out of a brand new set of values that she cultivated later on in her life. So, it would seem that she is acting badly as a result of her Alzheimer's. And as such, she is psychologically unable to do the right thing for the right reasons.

Bad Character #3: The Akratic

Let's explore a third type of bad character, the akratic. In examining the akratic's failure to do the right thing for the right reasons we are forced to unpack a third instance of psychological inability; a psychological inability that arises as a result of one's being a slave to their own desires and emotions. In what follows, I will introduce two examples of characteristic akratic behavior and leave the reader to assess whether these agents possess a psychological ability to do the right thing for the right reasons – or not. My own determination, using Wolf's analysis, is that they do not possess this ability.

I am defining the akratic, for the purposes of my project, as an unwilling slave to one's emotions and desires. This psychological enslavement may bind one to acting badly when faced with particular temptations, in the midst of especially strong emotions, or when placed in specific situations.

The akratic is, in essence, quite different from the other bad characters we have analyzed. Unlike the villain, for instance, an akratic may be thought of as a generally good person beyond the scope of their psychological enslavement. In the examples to follow, I introduce two akratics. Their introductions provide a broader biographical narrative which

aims to illuminate this aspect of their condition. One way to briefly illustrate this point is by way of the addict. If we understand addiction as a severe case of akrasia then although sometimes addicts act badly, we can still conceive of them as being “good people” in a world where they are not addicts — or, in a world where they have overcome their addiction.

This brings me to a second, and perhaps more crucial difference between akratics and other bad characters. Unlike other bad characters, akratics experience a painful psychological tension. The guilt and shame endured by those who feel incapable of living up to their values is not an experience shared by all bad characters. In fact, in calling these characters “akratic” (a.k.a “weak-willed”) we presume a psychological discord evident in how these persons want to act and in how they actually act. There is a split, it would seem, between their valuing-self and their acting-self. For example, while they may value faithfulness, loyalty, and various virtues associated with a monogamous lifestyle; they may nevertheless engage in adulterous behavior.

It is worth noting that this split does not occur in our other bad characters. Jojo, as Wolf showed in *Sanity and the Metaphysics of Responsibility*, was acting in accordance with his own predetermined set of values. While Jojo’s upbringing may have determined that he would cultivate a morally-skewed set of values, he nevertheless identifies with them. Villains like Jojo, unlike akratics, view their actions as representative of their values.

With regard to this split in the case of the malmaton, we may wonder whether these persons are even capable of acting from a valuing-self. When malmatons act it’s not apparent that they are capable of acting in accordance with a predetermined set of values. In the case of Alma, her Alzheimer’s may render her incapable of accessing that value

system which she operated upon prior to the progression of her illness. In the case of malmatons, then, the split between one's valuing-self and acting-self dissolves again.

I bring these differences to the forefront of our analysis in order to highlight important differences regarding the *causes* of persistently bad behavior. The villain, the malmaton, and the akratic are all psychologically incapable of doing the right thing for the right reasons. However, this castration of one's psychological moral capacities is different in each bad character type. For the villain it is a skewed set of values that causes them to act badly. For the malmaton it is a lack of moral sense that causes them to act badly. And, as the following examples aim to show, for the akratic it is one's unwilling enslavement to their own desires and emotions that cause them to act badly.

Case #1: The Over-Indulger

Nick is a younger man in his early 30s, who works thankless hours for an unimpressive sales-based corporation. Every weeknight, Nick picks up a tall can of Modelo at the corner store near his office and, drinking and driving, finishes the can just before he approaches his home. After giving his loving wife, and children of three, an affectionate greeting, he steps outside to indulge in a cigarette and marijuana. Make no mistake, Nick is a devoted husband and father. He plays with his children, interacts lovingly with his wife, and is as present as his other indulgences will allow him to be. Nick is well aware that his drinking and driving places his life and his family's stability at risk, yet he is admittedly weak in this area. He is also aware that his pattern of smoking and drinking makes him less emotionally available to his family, but he can't shake the habit.

Case #2: The Hothead

Melanie has been going to therapy to manage her anger for over two years. The incident that led her to therapy was the time she threw a glass vase at her husband's head – thankfully, he ducked out of the way. Melanie loves her husband, she is an excellent daughter, and a loyal friend. She works as an elementary school teacher and is greatly admired by her students. Yet, Melanie is a hothead. She loses her temper often and these emotional outbursts often have collateral damage in the form of strained relationships. Melanie is in therapy working toward anger management but she continues to yell, scream, and occasionally throw things.

The above examples of akrasia should shed light on the fact that Wolfian moral responsibility is sometimes in tension with our own beliefs about moral responsibility. Perhaps it is even at odds with beliefs we hold about whether or not we view ourselves as responsible agents. As was indicated in the above examples: Nick and Melanie both feel responsible for their actions. Part of this feeling of responsibility stems from the fact that they are good people – if only flawed in certain areas. The people close to Nick and Melanie likely hold them responsible, too. If Nick's drinking was to endanger his children, his wife may never forgive him. Melanie's husband is thankful that she is seeing a therapist and praises her for taking responsibility for her health.

Despite these facts regarding considerations of our own moral responsibility, a Wolfian analysis captures details that we might miss in our day-to-day lives. Her view picks up on the fact that while Nick and Melanie are good *sometimes*, they are controlled by factors outside of their control at other times. Both Nick and Melanie are governed by

factors internal to them, but those factors are not governed by Nick and Melanie, moreover. To illustrate this fact we can return to the Real Self View.

In order to be free and responsible on the Real Self View, one must be able to act in accordance with a value system that was determined using one's ability to reason. For Wolf, it is determined by one's ability to Reason. Specifically, it is to Reason in a way that corresponds with truth and goodness. Nick and Melanie possess this value-system but this is not the system from which their actions are derived. For Nick, it is an overwhelming desire to indulge that drives him. For Melanie, it is an energy storage of emotion that sends her into action. For this reason, akratics are sometimes understood as not acting but *reacting*.

It is in this way that Nick and Melanie are psychologically incapable of doing the right thing for the right reasons. Given the information we have about Nick and Melanie's psychological histories, it is difficult to expect that Nick won't drop by the corner store for a can of Modelo; and that Melanie won't fly off the handle on occasion. Therefore, it is incompatible with their respective psychological histories that they are psychologically able to do the right thing for the right reasons. Nick probably won't, anyhow, if the relevant temptations are present. And Melanie will also have an especially difficult time if she is fraught with red hot emotion.

How are we expected to make sense of this psychological inability? In analyzing three different types of bad characters – villains, malmatons, and akratics – we have found that people act badly for a variety of reasons. As was stated earlier, it is my position that anyone with a stable and pervasive pattern of acting badly does so as a result of a pathological presence in their early development. This is the line-up of pathological

candidates: psychological defects (e.g. psychopathy), a deprived (or depraved) childhood (e.g. child abuse), or severe psychological conditioning (e.g. cult-like practices).

Earlier, while discussing instances of villainy, I introduced Wolf's concept of insanity. Sanity, Wolf stipulated, has both a cognitive and a normative component. We are able to determine that Jojo is insane based on the fact that he does not know right from wrong. I would like to push this notion of sanity one step further. I would like to argue that it is not only insane to fail to recognize right from wrong, but it is also insane to fail to act in accordance with it. We were able to call Jojo insane because there was an obvious pathological presence in his past: his screwed-up moral development. I would like to argue that Jojo is not the only one of our bad characters suffering from mental illness in the form of moral badness. Insofar as our malmatons and akratics are victims of a pathological presence, they too, are mentally-ill or insane.

V. Moral Badness as Mental Illness

In this final section, I'd like to present a way forward for appreciating moral badness as mental illness. Understanding bad characters as suffering from mental illness is how we cross-over from a Wolfian conception of moral responsibility into an ordinary one. We begin to see bad characters as being bad due to no fault of their own. We begin to encounter them with compassion and recognize that rehabilitative efforts aimed at bolstering one's psychological health are preferable to plain punishment.

While I think that Wolf's notion of insanity traverses a similar train of thought as my own, it is different in two important respects. The first has already been glossed. Insanity, on Wolf's view, implies that one fails to see and appreciate the True and the Good; one can't

know the difference between right and wrong. What I am calling a strain of mental illness, on the other hand, can mean either (a) that one fails to know what the right thing is, or (b) that one fails to do the right thing – despite knowing what the right thing to do is. One can only fit the criteria for mental illness, however, if they have a stable and pervasive pattern of acting badly.

The second difference between Wolf's view and my own highlights the difference in our choice of terminology. I prefer the term mental illness as opposed to insanity – though one could, in theory, use them interchangeably – because it applies to a variety of conditions that exist along a spectrum of badness. One is the mentally ill sort of bad character if they are psychologically incapable of doing the right thing for the right reasons. Among the conditions that cause this incapacity include, as formerly stated: a skewed set of values, a lack of moral sense, and an unwilling enslavement to one's desires and emotions.

Another aspect of my position pressuring the coinage of moral badness as mental illness is the pathological presence I have spoken of. Again, this presence may not be known, but it must be taken for granted. We are operating under the valid assumption that every bad character has a backstory. This line of thinking has a home in Aristotelian Virtue Ethics and Neo-Aristotelianism. A full-fledged presentation of this position would burden us too much for our purposes but, given the influence of this position on my own thinking, it deserves a brief detour.

In her book, *Natural Goodness*, Philippa Foot argues that in order to be a good human being, *qua* human being, one must be morally good. In the same way that a good Oak Tree traps water in its roots, and a good flower engages in a process of

photo-synthesis, so a good human being is virtuous. It is, in other words, natural and normal for a human being to be a good person.

Elizabeth Anscombe in *Promising and its Justice* argues that insofar as the good hangs on our ability to keep promises, these moral behaviors are necessary for the flourishing of human societies. From this general schema we can derive many of our moral imperatives. If we want to flourish as human beings we need to be cooperative, productive, and so on. Bad characters fly in the face of these distinctively human ends. Insofar as a thing's essence is derived from its distinctive purpose, bad characters are defective human beings. In other words, their badness is a sign of sickness.

According to Aristotle virtue is the paragon of psychological health. It is what allows for the possibility of true flourishing. Aristotle discusses at length the significance of moral development. In order to become virtuous one must be physically healthy, have good moral educators, and opportunities to exercise their virtue. Aristotle, himself, considers three types of bad characters: the *kakos*, or bad man, the *akrates*, or weak-willed, and the brutish. These served as the inspiration for the villain, akratic, and the malmaton (respectively). Aristotle recognized the bad man, the akratic, and the brutish as suffering from a sickness of the soul. I understand them as mentally ill.

Concluding our detour, let's return to the consequences of Wolf's Reason View. Wolf's view would have us admit that we cannot hold bad characters responsible for their actions. This consequence becomes more appealing when we understand bad characters as mentally ill. After all, mental illness is typically thought to excuse moral responsibility. This is the motivating factor behind legal practices like the insanity defense. However, mental

illness is not always guaranteed to be a moral excuse.²⁰ Perhaps it depends on the extent of one's mental illness and the type of mental illness one has.

Consider again our malmaton, Alma. Alma has a severe case of mental illness: Alzheimer's Disease. She is altogether incapable of acting in accordance with judgments of right and wrong. Furthermore, among the lessons we learned from the Real Self View, Alma is not able to act from her Real Self. The Real Self theorists were right in conceiving of this ability as essential to our freedom and responsibility. So, insofar as Alma is incapable of being the author of her own life, she should not be held responsible for her actions — full stop.

Now we return to the villains. While villains meet the conditions for responsibility on the Real Self View they do not meet those conditions on Wolf's view. Despite their ability to be authors of their own lives; they are not responsible for being the types of authors they are. Villains are terrible authors. They conduct their lives in a way that is alienating, destructive, and antagonistic. Anyone authoring their lives in this way is clearly not well. Wolf picks up on the fact that villains, for this reason, must not be free. A free person would choose happiness and connection over separation and angst.

The Villain, understood as a candidate for mental illness, is obviously not responsible for her mental illness. However, we should ask: is the villain responsible for her mental health? Is she responsible for recognizing her moral badness as mental illness and seeking help? The answer to this is tricky. If the villain finds herself in a society that understands moral badness as mental illness she may be able to find herself relating to her experience as an experience of sickness. After all, what reason do we have to believe that

²⁰ See Pickard (2013), Elliot (2004), King & May (2018)

phobic patients, depressed patients, ADHD or OCD patients, would ever seek treatment if they had no cultural context by which to understand and relate to their experiences (i.e. symptoms) as an experience of mental illness?

What all of this amounts to is a tension between Wolf's Reason View and my own intuitions about moral responsibility. Wolf's test for moral responsibility allowed us to recognize that bad characters are psychologically incapable of doing the right thing for the right reasons. It may not be impossible for them to do the right thing for the right reasons but, close examination of their psychological histories suggests, we can't *expect* them to do the right thing for the right reasons. Therefore, according to Wolf's Reason View, we should not hold them morally responsible for their actions.

As I've come to understand it, the fact that one's psychological history implies that it is especially difficult for them to do the right thing for the right reasons, does not entail that they can't be held responsible for their actions. What it tells us, rather, is that they are mentally ill. What matters, then, for the moral responsibility of bad characters is whether they are capable of taking responsibility for their health.

However, In the case of villainy, this may not be possible in a society that doesn't understand moral badness as mental illness. This is one point in favor of adopting a clinical perspective on moral badness. In a culture that literally paints them as "the villain" they are unable to understand and relate to their experience as an experience of sickness. Therefore, they do not seek the proper rehabilitative treatment; nor is it typically offered to them.

With respect to being capable of taking responsibility for their health, the akratics are already one step ahead of the villains. The psychological distress that one experiences

as an akratic — as someone who lacks control over their own actions in the face of strong desires and emotions — usually leads one to seeking external help. It is no wonder that the Alcoholics Anonymous community has grown to service a startling number of akratics from overeaters, gamblers, and hoarders to the salacious and the codependent.

Akratics recognize that they struggle to be the authors of their own lives. Sometimes they even recognize themselves as sick. In taking responsibility for their own health — like Melanie seeking to manage her anger — they are capable of being free and responsible beings. That is, they are capable of being free and responsible beings despite not always being capable of doing the right thing for the right reasons.

An analysis of bad characters utilizing Wolf's test for moral responsibility allowed us to acknowledge that some agents are psychologically incapable of doing the right thing for the right reasons. An appreciation for this incapacity led us to a recognition of their plight as an experience of mental illness. This may have been Wolf's worry all along. In her earlier work she was concerned about Jojo's insanity; an insanity that she felt morally excused Jojo from the realm of moral responsibility. Ultimately, Wolf is right that Jojo is not morally responsible for his own insanity. But, whether Jojo is responsible for his own becoming *better* is a consideration that should not be taken off the table.

It is not uncommon that we view agents as more responsible in some areas of their lives, and less responsible in others. In the case of physical sickness we can draw this out quite neatly. Suppose my mom suddenly becomes sick with the flu. She has a fever, she is dehydrated, fatigued, and so on. In no way do I perceive my mom as being responsible for the sickness she incurs. However, if her symptoms fail to clear up in a matter of weeks, I expect her to call her doctor; assuming that she is not so ill that she is physically incapable

of doing so. If she failed to call her doctor, I'd hold her responsible for neglecting her own health.

Therefore, insofar as bad characters are psychologically incapable of doing the right thing for the right reasons they are mentally ill. We cannot and should not hold them responsible for incurring this illness. However, we can sometimes hold them responsible for not taking charge of their own health. Again, what matters most in the case of villainy is whether they are able to understand and relate to their experience as an experience of illness.

Furthermore, while absolving bad characters of moral responsibility might at first seem like a compassionate act – it is not. If our society is flooded with bad characters, as the cynic might insist, something must still be done about them. We can't have villains running amok and disrupting our cooperative efforts whenever the mood strikes them. Recognizing villains as lacking agency over their own lives – and essentially lacking any ability to become morally-upright citizens – we funnel them out of mainstream society through our systems of punishment. This assumption that the “bad can't get better” is a dangerous fabrication. It is akin to declaring the common cold incurable.

Let's suppose our society were to adopt the theoretical convictions I have presented in this final section. Ideally, this would pave a way for bad characters to recognize themselves as – not only bad – but sick. They would no longer examine their condition as one of mere difference – a difference upon which they are ruthlessly outcasted – but as one that must be treated. Surely, a villain who sincerely endorses her own bad behavior is not going to want to change. It might be that some of these treatments must be compulsory in nature. Regardless, we should not shy away from cultivating rehabilitative

systems of punishment aimed at improving the bad character's ability to do the right thing for the right reasons.

References

Elliott, Carl (1996). *The Rules of Insanity: Moral Responsibility and the Mentally Ill*. SUNY Press.

Frankfurt, Harry (1987). Identification and Wholeheartedness. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.

King, Matt & May, Joshua (2018). Moral Responsibility and Mental Illness: a Call for Nuance. *Neuroethics* 11 (1):11-22.

Pickard, Hanna (2013). Psychopathology and the Ability to Do Otherwise. *Philosophy and Phenomenological Research* 90 (1):135-163.

Watson, Gary (1997). 3. On the Primacy of Character. In Daniel Statman (ed.), *Virtue Ethics: A Critical Reader*. Georgetown University Press. pp. 56-81.

Wolf, Susan (1990). *Freedom Within Reason*. New York: Oup Usa.

Wolf, Susan (1987). Sanity and the Metaphysics of Responsibility. In Gary Watson (ed.), *Free Will*. Oxford University Press.