The Bad, the Evil, and the Monstrous

By

Alyssia Fogarty

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

the Degree of Master of Arts in Philosophy.

November 2015, Halifax, Nova Scotia

Approved:	Dr. Todd Calder	
	Supervisor	
Approved:	Dr. John MacKinnon	
	Reader	
Approved:	Dr. Greg Scherkoske	
	External Examiner	
Date:	November 27, 2015	

The Bad, the Evil, and the Monstrous

by Alyssia Fogarty

Abstract

This thesis focuses on defending the conceptual distinction between evil persons and irredeemable monsters. I argue that applying the comparative methodology to borderline cases can make sense of the intuition that evil persons are the worst sorts of people—because they are worse than those who are merely bad—without being forced into too narrow a definition of evil persons as monstrously terrible. Using this methodology, I show that a hybrid motive-act-based regularity account is the most satisfying theory of evil personhood because it can make sense of the distinction between merely bad persons, evil persons, and monsters. One practical consequence of this view is that we can consider both theoretical and practical conditions of reformation and redemption for evil persons.

November 27, 2015

Table of Contents

Chapter I: Introduction

1.1. The Comparative Methodology

Evil is an extreme moral concept. Evil actions and evil characters deserve our

strongest moral condemnation (Singer 2004, 185). For this reason, many theories of evil

personhood limit the application of the concept of evil to only the very worst kinds of

people. Those who are the most terrible and monstrous are the most likely to be

considered evil people, even if it turns out that such people do not actually exist, or are

incredibly rare. Many take the rarity of evil persons as something that ought to be

captured by a theory of evil personhood (Russell 2014, 135). We may also think that the

extremity of the term "evil" shows that the concept is meant to exclusively capture the

vilest sorts of people. Peter Brian Barry calls his thesis that an evil person is the morally

worst sort of person a "modest proposal" (Barry 2010, 26), since it seems to be a modest

suggestion that if evil is anything, it is the worst possible kind of thing. However, this is

not the only way that we can go about trying to conceptualize the nature of evil.

Todd Calder argues for the use of what he calls "the comparative methodology"

(Calder 2015, 118) to develop a theory of evil action. This approach focuses on directing

our critical reflection to cases of uncontroversial evil actions and mere wrongdoings.

Claudia Card relies on a similar comparative methodology, stating that atrocities are her

paradigms of evil and they are useful as such because atrocities are uncontroversial cases

of evil that can serve to distinguish between evils and lesser wrongs (Card 2010, 6-7).

Additionally, Marcus Singer argues that when we say that something is evil, we are

saying that it is something "over and above ordinary badness" and so we must search for

uncontentious cases that are "clear beyond any reasonable doubt" in order to come to a

stable conclusion about the nature of evil (Singer 2004, 190-1).

Calder and others advance this method for developing theories of evil action, but I

contend that we can adopt a similar method for developing a theory of evil personhood.

However, the comparative methodology can be used in different ways. One way to use

the comparative methodology is to develop a theory of evil personhood by comparing the

vilest sorts of people with people who are clearly only merely bad. Another way to use

the comparative methodology is to develop a theory of evil personhood by comparing

borderline cases of evil persons with borderline cases of the merely bad. When evil

personhood is understood from the position that to be an evil person is to be the worst

possible kind of person, then those who are not the worst possible, but who are worse

than merely bad, are left unaccounted for. This suggests that a beneficial use of the

comparative methodology for theories of evil personhood will be to focus our attention on

borderline cases of those who seem intuitively evil and those who do not. Rather than

drawing from obvious paradigms of the vilest evil villains, we can appeal to cases that sit

on the boundary of evil to bring out the characteristics that we find ourselves unwilling to

describe as merely bad.

We can imagine an obvious and uncontroversial case of an evil person named

Sam. Sam is the kind of person who not only kidnaps, tortures, rapes, and kills dozens of

innocent people, but he also takes great pleasure in inflicting pain on others. Sam knows

that his behaviour is wrong, and he gets an added thrill from that. He is the kind of person

who writes letters to the police to mock them. Sam is a persistent evildoer whose

character is sadistic, malicious, and defiant of morality, and he has no feelings of guilt,

empathy, or compassion for other human beings. It is not a stretch for us to say that Sam

is an evil person. We can also imagine a fairly uncontroversial case of a merely bad

person named Paula. Paula is the kind of person who steals money from her friend's

wallet and spreads injurious lies about people she doesn't like, but she stops short of

causing anyone significant harm. She does genuinely care about some people and

sometimes even feels guilty about her behaviour, though not every time, and not enough

to stop. Paula has a selfish and deceitful character, which makes her a bad person. While

Sam and Paula illustrate two uncontroversial cases of individuals who occupy different

areas of the moral spectrum, it should be clear that there are many other intermediate

kinds of cases. It is unlikely that we know of many people like Sam outside of fiction, but

we are likely to know of many who are worse than Paula. Perhaps even worse enough

that we want to call them evil.

Consider the case of Joe. Joe is a pimp and his job is to prepare victims of human

trafficking for sex work. He does not choose his victims or abduct anyone himself, but he

does force young victims to participate in child pornography as a way of conditioning

them into prostitution. Joe is not overtly callous or violent and he has some sympathy for

his victims. He rationalizes that pornography is the least traumatic introduction into

prostitution for these minors. Joe is also not sexually interested in children himself, but he

3

knows that other people are. Is Joe merely bad? Or even very, very bad? Or is it the case

that we cannot adequately capture the morally significant badness of Joe without

appealing to a stronger moral concept, like evil? Joe is not like Sam, since Joe could be an

even worse kind of person who is callous, violent, and who finds his victim's pain to be

intrinsically pleasurable. Nonetheless, a child pornographer like Joe still seems to strike

us as an evil person. For one reason, Joe wilfully refuses to acknowledge that his actions

are extremely wrong and harmful by rationalizing his behaviour. And he uses these

rationalizations to justify exploiting kidnapped children for his own gain. Joe rationalizes

that forcing children into pornography will make their entrance into prostitution easier.

While this might be true, we should fault Joe for never considering that he can prevent

more pain to the children by helping them to escape prostitution altogether.

By focusing the comparative methodology on borderline cases we can make sense

of the view that evil persons are the worst sorts of people, because they are worse than

those who are merely bad. In addition, we can do so without being forced into too narrow

a definition of evil persons as monstrously terrible. We can make sense of clear cases like

Sam and Paula, but we can also approach less obvious cases like Joe. This means that

attending to borderline cases confers an advantage over other ways of thinking about the

nature of evil, since it can speak to a wider range of cases and can address more precisely

what distinguishes an evil person from a merely bad person. However, some argue that a

restricted conception of evil, as belonging to only the very worst, is precisely what an

account of evil personhood is meant to capture. Even if we allow that Joe is significantly

worse than Paula, he is still not as bad as Sam. This is why some argue that we can only

call someone like Sam evil, since Sam is the worst kind of person and we want to be able

to capture the moral significance of that. What we need, then, is to be able to distinguish

between evil and mere badness, while also accounting for the monstrous.

1.2. The Monstrous Conception of Evil Personhood

Phillip Cole claims that the monstrous conception of evil persons is the view that

evil people are monsters in human shape who "constitute a distinct class, different from

the rest of humanity, with a different nature" (Cole 2006, 13). Daniel Haybron argues that

an evil person is someone "between the human and demonic" whom we call a monster to

"emphasize the profound moral and psychological gulf between them and us" (Haybron

2002b, 277). Michael Stone claims that the term "monster" is reserved for behaviour that

is so unlike our own "as to create a comfortable distance between our self-image and our

perception of those freakish aberrations of nature" (Stone 2009, 247).

These claims focus on the view that evil people are monsters, and monsters are

those who are so horrendously terrible that we cannot recognize them as being anything

like us or conceive of them as ever becoming better kinds of people. A monstrous

conception of evil personhood is the view that evil people are the conceptually worst

kinds of people, and, because of their monstrosity, are irredeemably evil. To be

irredeemably evil in this way is to have an unchangeable evil disposition or propensity, or

deeply entrenched evil desires or motivations that cannot be altered through typical

means. Haybron, for instance, argues that the ascription of evil affords us with

5

simplifications, since we can confidently predict that the evil person will never behave

from decent motives and we can never expect good deeds or genuine friendship from

them. This means that we need not worry or hesitate about hostile dealings with evil

people, since we will never be bothered by sympathies or affections for them (Haybron

2002b, 277). Luke Russell provides a weaker version of this view, arguing that evil

persons have an especially highly fixed disposition to perform evil actions (Russell 2014,

168), and this supports the intuition that an evil person is a "moral write-off" (169). As a

moral write-off, an evil person is not a suitable candidate for attempts at moral reform

since they are "beyond rational persuasion and cannot be made good" (225).

Whether we adopt this conception of evil personhood matters because it sways

how we evaluate evil people. If we think that only monsters can count as evil, then our

accounts will be too narrow to make sense of cases where someone seems worse than

bad, but not monstrously so. Recall the case of Joe. On a monstrous conception of evil

personhood, Joe does not qualify as an evil person because we are unlikely to describe

him as demonic, or as a freakish aberration of nature, despite his appalling actions. Joe is

recognizable as being somewhat like us, in that he is motivated by self-interest and

rationalizes his wrongdoing. Joe purposely deceives himself into thinking that he is not

causing significant pain, and so we are right to judge him harshly, but this kind of self-

deceptive rationalization is a familiarly human trait. We might think that Joe has made

some bad decisions or has been caught up in bad circumstances, so it is possible that he

could change for the better. But if evil people are monsters who cannot be understood in this way, then a child pornographer like Joe is a merely bad person.

A monstrous conception of evil personhood implies that someone like Joe is either a merely bad person or he is a monster. Because this conception presents a false dichotomy, it is more likely to be misused as a concept of evil in particular cases. By thinking that all evil people are monsters we are susceptible to accurately calling someone evil in the comparative sense—as worse than merely bad—but inaccurately inferring that this means they are a monster beyond reform. For instance, someone might resist the implication that Joe is a merely bad person and call him evil, but they infer that he has characteristics that he does not have. Susan Neiman argues that we hold a view of evil as inevitably connected to evil intentions because this "is more soothing than alternatives" (Neiman 2003, 271). Neiman argues that we find it too disturbing to acknowledge an apparent absence of malice aforethought in evil, so we prefer to argue that cruelty and malicious intentions are always there in subliminal form. We might think that what Joe does is monstrous and that anyone who could do what Joe does could not be anything but a monster. We prefer the more comforting and simplistic view that there can be no such thing as a child pornographer who also genuinely cares about children to some degree. From this, we deny that Joe really is sympathetic, or dismiss his sympathy as a monstrous perversion of the term. We deny that someone like Joe could have any relevant morally redeeming character traits (or, at least, could only have very few to a limited degree), and so we deny that someone like Joe could ever become a better person. In virtue of a

7

monstrous conception of evil personhood, we infer that Joe is either merely bad or he is a

monster. Since we are unlikely to think that a child pornographer is a merely bad person,

we conclude that Joe must be a monster.

Consequently, the monstrous conception of evil personhood also determines how

much attention we give to conditions of reformation and redemption for evil persons. If

we think that only monsters can count as evil, and monsters are those who are beyond

reform, then we are committed to the view that evil persons deserve simply to be killed or

locked away (Russell 2014, 213), that they are moral exiles (Haybron 2002b, 277), and

that they are beyond communication and negotiation (Cole 2006, 236). Thus, it would be

entirely pointless to consider conditions of reformation and redemption for evil persons.

Yet, as Laurence Thomas points out, insofar as the possibility of redemption arises at all,

it does so in the context of evil. It would be inappropriate to talk of moral redemption in

cases of trivial minor wrongdoings (Thomas 2009, 130). A deceitful Paula may become

an honest person, and we would praise her for it, but it would be strange to refer to her

transformation as a case of moral redemption when her worst traits led her to commit

only trivial wrongs.

The monstrous conception of evil personhood faces three potential challenges.

The first challenge is that an account of evil persons as monsters cannot provide an

adequate bad/evil distinction. This is because it cannot make sense of non-monstrous

people who we nonetheless still think are worse than merely bad. A second, related,

challenge is that the monstrous conception can lead to falsely ascribing monstrous

characteristics to non-monstrous evildoers. This occurs in virtue of the false dichotomy

that extreme wrongdoers are either merely bad or they are monstrous. Since we are

unwilling to allow that certain kinds of evildoers are merely bad people, and we fail to

recognize that there can be non-monstrous evil people, we judge certain kinds of

evildoers to be worse than they actually are. The third challenge is that a monstrous

conception of evil personhood precludes theories of reformation for evil. This is based on

the presupposition that theories of reformation are irrelevant because evil people are

necessarily irredeemable. I offer a solution to these challenges for those who want to

maintain that people like Sam belong in their own moral category, as well as for those

who want to maintain that people like Joe are relevantly distinct from people like Paula. If

we allow that the concept of an evil person is distinct from the concept of an irredeemable

monster, then we can reconcile the problem of distinguishing between the bad, the evil,

and the monstrous.

Some have pre-emptively rejected this distinction in favour of the monstrous

conception of evil personhood, where the concept of an evil person and the concept of a

monster are identical. Haybron promotes such a view, arguing that the distinction we

ought to recognize is not one between evil persons and monsters, but between the evil

person and the moral criminal. The moral criminal is one who has "performed seriously

immoral acts that reflect major deficiencies of character" (Haybron 2002b, 272).

According to Haybron, those non-monstrous people whom we may want to call evil "are

better described as an extreme variety of criminal" (273). While moral criminals are

9

disposed to perform frequent acts of evil, they can also have a good side and exhibit genuine compassion and other moral virtues. Evil persons, on the other hand, will always be so extremely bad as to never have a good side or ever be capable of entering into any human relationship (277). For Haybron, to regard people as evil is to treat them as "monsters who are not fit even for friendship or familial relations," and we are simply mistaken when we call mere moral criminals evil (278).

In another way, Marcus Singer argues that additional terms of extreme moral condemnation, such as "monstrous" or "inhuman", all imply evil. Singer argues that such terms do not go beyond evil in meaning but are simply forms of emphasis and expression (Singer 2004, 195). For Singer, an evil person is one who knowingly performs, wills, orders, or allows evil actions. Singer defines an evil action as one so awful and horrendous that no ordinary decent reasonable human being could conceive of themselves doing such a thing (196), so it seems that he accepts the view that evil persons and monsters are conceptually synonymous. Yet, Singer also argues that "evil" lies at the end point of a scale of badness, "with 'monstrous' ... perhaps going off the scale" (196). So, while Singer initially rejects the conceptual distinction between evil persons and monsters, he does leave open the possibility that "monstrous" may belong to its own category. Singer is right to point out that monstrous implies evil, since all monsters will be evil people, but this need not mean that all evil people are necessarily monsters.

There are theorists who do draw a clear distinction between evil persons and monsters, however. John Kekes argues that there are two kinds of evil characters: moral

monsters who cause evil autonomously and moral idiots who cause evil

nonautonomously. Kekes defines autonomous evil actions as those that the agent chooses

to perform without force and with complete awareness of the significance of the choice

and the action. The agent also makes this choice after evaluating and comparing the

available alternatives. Nonautonomous evil actions, on the other hand, are chosen by the

agent and are not forced, but the agent is not completely aware of their significance and

they fail to evaluate their actions reasonably in light of alternatives (Kekes 1998, 217). In

my earlier examples, Sam, the knowingly defiant sadistic torturer, would be a moral

monster in virtue of his autonomous evildoing, while Joe, the sympathetic child

pornographer, would be a moral idiot. This is because Joe wrongly believes that forcing

his victims into child pornography is justified, since it will make their entrance into

prostitution less harmful. Joe believes that he wants to cause less harm, so he believes this

is the best option, but he fails to consider the available alternative of providing real help

to his victims by helping them escape prostitution.

Importantly, Kekes points out that the autonomous evil actions of moral monsters

cannot account for the majority of evildoing, but rather the prevalence of evil is largely

the result of nonautonomous patterns of actions performed by moral idiots (218). This is

because moral monsters are exceptionally rare. Such individuals would need to not only

have a perfectly clear understanding of their preferred evil actions among all possible

alternatives, and a robust strength of character to consistently follow their evil plans—but

they would also need to keep their true nature concealed from others to avoid detection

and social censure. To be a moral monster would be incredibly difficult for most people, and even more difficult for one to remain undetected and free. Consequently, moral monsters would be unable to cause large-scale evil on their own, since major atrocities would require the participation of numerous assistants, many of whom would be moral idiots rather than like-minded moral monsters. Thus, it would be implausible to attribute the prevalence of evil to a few rare monsters.

Todd Calder adopts Kekes' distinction and amends it for his motive-based account of evil personhood. Calder agrees with Kekes' view that moral monsters are the more basic type of evil character who often desire other people's significant harm for an unworthy goal, such as for their own pleasure or for entertainment (Calder 2003, 373). Calder also contends that moral idiots possess similar kinds of desires, but the difference between them and moral monsters is that moral idiots "believe the goal for which they desire the harm morally justifies the harm" when it does not. Calder argues for an additional class of moral idiots who do not count as evil persons when the belief that their goal justifies their victim's harm is defensible (373). This kind of belief would be defensible, Calder argues, if it would be reasonable "to someone who had sufficient cognitive and deliberative powers" to think that the goal is worthy "given the best scientific and sociological considerations available" (370).

In a different way, Paul Formosa acknowledges the distinction between evil persons and monsters by denying the existence of monsters outright. Formosa argues that evil persons "are not an utterly distinct class of beings, totally dissimilar to the rest of us"

and by acknowledging this we can "avoid the mistake of demonizing them, the evil ones, while monopolizing humanity for us, the good ones" (Formosa 2008a, 235). Arguing against the demonization of evil persons, Formosa points to the language used in an SS pamphlet:

> From a biological point of view he [the Jew] seems completely normal. But in fact he is a completely different creature, a horror. He only looks human, with a human face, but his spirit is lower than that of an animal. A terrible chaos runs rampant in this creature, an awful urge for destruction, primitive desires, unparalleled evil, a monster, subhuman (Formosa 2008b, 400)

Formosa suggests that a view of evil persons as monsters faces not only a conceptual problem, but also a moral one. Similar to how anti-Semites dehumanize their victims, the way in which we often describe evil people as monsters serves to dehumanize them through demonization. While evil people do deserve to be strongly morally condemned, evil people are not subhuman or devoid of any identifiably human characteristics. Formosa argues that we ought to recognize that evil people "are not inhuman monsters, but deeply flawed human beings" and they deserve our strongest moral condemnation because they have the kind of characters that result in repeatedly inflicting extreme harms on others (Formosa 2008a, 235), not because they are demonic.

The conceptual distinction between evil persons and monsters has the advantage of helping us to understand the differing views on the scope, extremity, and permanence of evil personhood. In the subsequent chapter I will investigate theories of evil personhood that either endorse or rely on a monstrous conception of evil, as well as the alternatives. To show if a theory does rely on a monstrous conception of evil personhood,

I will test whether or not specific accounts can deal with borderline cases. The goal is to

determine which accounts can adequately distinguish between merely bad people, evil

people, and monsters, without conflating these distinctions or leaving them unaccounted

for. While some theorists utilize a comparative methodology to provide a bad/evil

distinction, many present a false dichotomy of only people like Paula and Sam to be

considered. This method of comparison often leaves borderline cases, such as that of Joe,

inadequately addressed. If an account leads to the counterintuitive result that certain kinds

of putative evil people are merely bad people, just because they are not monstrously

terrible, then it is more likely to be an account of monsters. While an account of monsters

is theoretically useful, it is not itself a comprehensive account of the concept of evil

personhood.

Chapter II: Evil Personhood

As a preliminary step, it is important to make clear how theories of evil personhood are different from theories of evil action. Accounts of evil personhood tend to be derived from prior accounts of the concept of evil action, or are at least dependent on some presupposed notions about what counts as doing evil. This highlights the important difference between evil people and evildoers. An account of evil action can tell us who counts as an evildoer, since whoever performs an evil action is someone who does evil. An account of evil personhood says something more than this; it tells us who counts as someone with an evil character, rather than someone who has done some amount of evil. Theories of evil personhood focus on an agent's character over and above what an agent happens to do. I will not defend any particular account of evil action here, but for the sake of clarity I will take a harm-based approach to evil action. That is, evil actions are most basically extreme culpable wrongs,[1] and what makes an action an extreme culpable wrong is that it results in someone else's significant harm that could have been, and ought to have been, avoided. What makes a harm significant enough for evil has been defined by various theorists as a "life-wrecking harm ... that violates the minimum conditions of human well-being" (Formosa 2008, 229), which "interferes with the functioning of a person as a full-fledged agent" (Kekes 1998, 217). Or it is harm that "deprives, or seriously risks depriving, others of the basics that are necessary to make a life possible and tolerable" (Card 2002, 16) and it is "so extreme that it stands out in one's life and has

---

[1] Russell calls this a "psychologically thin account of evil action" since evil actions can come from a broad range of motivations and intentions. This is contrasted with "psychologically thick accounts of evil action" which posit that something like malice or sadism is necessary for evil actions (Russell 2014, 74).

long-lasting negative psychological effects" or at least would do so "on a human being

ordinarily situated" (Calder 2015, 120).

2.1. Non-Dispositional Accounts

Various theories of evil personhood can be categorized under two main headings,

as either dispositional or non-dispositional accounts. One sort of non-dispositional

account is what Luke Russell calls "aggregative accounts of evil personhood." The basic

version of this kind of account holds that "an evil person is someone who has done more

than a specified amount of evil" (Russell 2014, 133). One sort of aggregative account is

an "aggregate-of-harms" account, according to which someone is an evil person only if

the aggregate of unnecessary and significant harm they cause by culpable wrongdoing

exceeds a specified threshold (141). Another sort of aggregative account is a "balanced

aggregative" account, according to which the aggregate of harms minus the aggregate of

the agent's culpable benefits exceeds a certain threshold (143). In other words, on an

aggregate-of-harms account someone is an evil person if they are responsible for causing

a certain amount of unnecessary significant harm, and on a balanced aggregative account

someone is an evil person if they are responsible for causing a certain amount of

unnecessary significant harm and they are responsible for not causing enough significant

benefits.

These kinds of non-dispositional accounts are ultimately unsatisfying. The

threshold of what would count as "enough harm" is unclear, and consequently, these

accounts can blur the line between merely bad evildoers and evil people. It cannot be the

case that "enough harm" is determined by causing any degree of significant harm, since a

merely bad evildoer will also cause significant harm in order to count as performing an

evil action. But then it seems arbitrary to state that someone is an evil person just because

they cause some measurement of harm more than their merely bad counterpart. It would

be implausible to say that someone who kills five people in an explosion is an evil person,

but had one of their victims not been in the area of the blast, then they would be a merely

bad evildoer because they did not cause enough harm to count as evil.

A more plausible non-dispositional account is what Russell calls the "regularity

account of evil personhood" according to which someone is an evil person if they

frequently or regularly perform evil actions (138). On a regularity account we can

discover who counts as an evil person by noting whether they perform evil actions that

are not "a one-off moment of weakness or emotional frenzy" but are instead evil actions

that are "part of a larger pattern of evil behaviour" (Formosa 2008, 234). For example,

Stone argues that when we call someone evil we are implying "that the person can be

counted on to commit such acts habitually and often" (Stone 2009, 23). Kekes similarly

argues that a person is evil if they "habitually perform evil actions" (Kekes 2005, 2), and

Singer claims that engaging in a pattern of evildoing is sufficient for evil personhood

(Singer 2004, 197). Haybron, who rejects regularity accounts, nonetheless concedes that

they do have *prima facie* appeal. This is because accounts based on regular or frequent

evildoing can offer a plausible reason to care about evil, since evil people will be those

who commit the worst crimes the most often, and they will be the most dangerous kinds

of people. Additionally, a regularity account can provide reliable practical criteria for

determining whether someone really does count as the worst kind of person, since the

repeated performance of evil actions will often be the easily observable expression of an

evil character (Haybron 2002b, 272).

It seems right to say that the performance of only one, or very few, evil actions

alone is not enough to qualify someone as an evil *person* rather than an *evildoer,* so a

regularity account of evil personhood appears to offer a plausible theory. If we are

interested in distinguishing between evil persons and merely bad persons then our account

will need to be able to make sense of the view that some evildoers will be merely bad

people. An account that includes a frequency of evildoing might help make sense of this.

Someone who performs an evil action only once, or very rarely, may not be bad enough to

count as an evil person. Yet, while it is widely accepted that not all evildoers are evil

people, some go even further to argue that not all evil people are necessarily evildoers at

all. On this view, a regularity account fails because it is not the case that a necessary

condition for evil personhood is the performance of evil actions.

Haybron presents this kind of criticism, arguing that frequent evildoing accounts

arise from the misconception that moral criminals count as evil. Remember that Haybron

claims that the moral criminal is one who has "performed seriously immoral acts that

reflect major deficiencies of character" (Haybron 2002b, 272), but that they also exhibit

some moral virtues unlike the wholly evil person (273). In this way, a frequency of

evildoing could only tell us who counts as a moral criminal, but not who counts as an evil

person. Haybron argues that evil persons will always be so extremely bad as to never

have a good side or ever be capable of any human relationship (277). It follows from his

view that so long as an evildoer has any morally decent qualities they will not be an evil

person, and so long as someone has no morally decent qualities they will be an evil

person, regardless of whether or not they perform evil actions. But this would mean that

someone who regularly kidnaps and murders women, but who is also a devoted and

loving father, would not be an evil person (Calder 2009, 25). And, on the other side,

someone who lives their entire life secretly despising humanity, but performing good

deeds to keep up appearances, would be an evil person. We have reason to reject the claim

that a regularity account fails because it mistakes moral criminals as evil if at least some

moral criminals *do* count as evil persons, like a devoted and loving father who spends his

weekends murdering women. If this is the case, then the performance of evil actions is at

least sufficient, if not necessary, for evil personhood.

A regularity account does face a related problem, however. On a regularity

account we cannot make sense of why we take failed attempts at evil to indicate that the

perpetrator may be an evil person, since such attempts cannot be counted as performing

evil actions. Failed attempts to inflict extreme harm are sometimes considered sufficient

to count the perpetrator as having performed an evil action (Russell 2014, 53-4), but this

view conflates theories of evil action with theories of evil personhood. We might think

that someone who seriously attempts to set off a bomb in a crowded mall does something

wrong enough to count as evil, even if his bomb malfunctions or he is apprehended before

it can go off. However, our judgments ultimately point to the perpetrator rather than the

action. That is, the bomber's failed attempt at evil indicates that he is the kind of person

who can and will do evil—that he is (potentially) an evil person. Calder makes this point

by arguing that "just as we make a distinction between murder and attempted murder, we

should make a distinction between evil and attempted evil" (Calder 2015, 121). We

should accept that "[e]vil hasn't been done until someone else gets hurt" [2] while also

recognizing that failed attempts at evil "may still be indicative of an evil character"

(Calder 2003, 367). In this way, we can better understand cases of what some call

"harmless evil actions" (Russell 2014, 53; Kekes 2005, 193) as failed attempts at evil.

This presents a serious challenge for the regularity account, which holds that it is a

necessary condition of evil personhood that an evil person performs evil actions. If failed

attempts at evil lead us to judge perpetrators as potentially evil persons when they do not

qualify as having performed evil actions, then it does seem that there can be evil people

who have never done evil. One way to solve this problem is to include intentions, so that

a person is evil if they perform a certain amount of evil actions, or if they intend to

perform a certain amount of evil actions. Formosa and Card offer comparable views of

this type. Formosa argues that an evil person is someone "who repeatedly perpetrates, or

at least intends to perpetrate, evil acts" (Formosa 2008, 233), and he adds the intention

clause in order to account for failed attempts at evil. Similarly, Card argues that we can

rightly judge someone as an evil person "on the basis of persistent and effective evil

---

[2] Calder distinguishes between three types of evil action: causing-harm-evil, allowing-harm-evil, and witnessing-harm-evil. For causing-harm-evil, the action is only evil if the harm caused is significant since only significant harms are morally grave enough for evil (2015, 119).

motives or intentions" (Card 2002, 21) and that someone "whose evil intentions are

ineffective can also be an evil person, if the intention persists" (22). By incorporating

intentions into a regularity account, we can say that failed attempts at evil indicate that the

perpetrator may be an evil person because they intend to perform evil, even if they are

inevitably unsuccessful.

A larger problem facing the regularity account is that it cannot provide a strong

bad/evil distinction. A regularity account seemed initially plausible because it could help

make sense of the distinction between merely bad evildoers and evil people. But extant

regularity accounts do not provide a way to determine the threshold between bad and evil.

It remains unclear how much evil must be performed and so we do not know exactly how

much evildoing is enough for evil personhood. Moreover, it seems unlikely that these

accounts could provide a threshold that is theoretically satisfying. As some have argued,

the threshold between bad and evil could be indicated by a pattern of evildoing (Singer

2004, 197), or the habitual (Kekes 2005, 2), repeated (Formosa 2008, 233), or persistent

(Card 2002, 21) performance of evil actions. In order to solve the problem of failed

attempts at evil, we could add that this also includes a pattern or habitual performance of

evil intentions. So we could say that a regularity account states that a person is evil if they

reliably engage in a pattern of performing, or intending to perform, evil actions. Even

still, this understanding remains too vague for an adequate bad/evil distinction, and can be

entirely arbitrary. In judging the case of a periodic murderer, this view would suggest that

we could go from calling him a merely bad person to calling him an evil person only after

he has completed some selected number of murders. But this would mean we could say

that he is a merely bad evildoer after killing three people over a span of a month, but then

qualifies as having killed regularly enough to count as an evil person after an additional

murder the next month. A regularity account would need to provide a more specific

measurement of "enough evil" to be a plausible theory of evil personhood, and it would

need to provide good reasons for why that measurement is not an arbitrary threshold.

In a similar way, a regularity account is unable to fit well with our intuitions about

who actually counts as an evil person. Russell argues that an evil person could be

someone who performs his evil actions sporadically, infrequently, or all in a rush (Russell

2014, 139). A brooding spree killer, for example, could spend most of his life merely

fantasizing about murder without ever inflicting harm, and then suddenly kill several

people all at once. He is still plausibly an evil person even though he did not regularly

perform his evil actions. Russell suggests that the reason we think a brooding spree killer

is an evil person is that his evil character is what explains his eventual attack, not the

other way around. A regularity account is in tension with the intuition that evil

personhood is not just an ascription we give someone in virtue of their regular evildoing.

Rather, evil personhood is more plausibly an assessment of a person's characteristics, and

those characteristics can result in regular evildoing. In other words, we think that the

brooding spree killer and the periodic murderer are both evil people because they are the

kinds of people who are characteristically *disposed* to evil.

2.2. Dispositional Accounts

Dispositional accounts can be defined broadly as accounts which hold that "an evil person is someone who would do evil if given the chance" (Russell 2014, 133). Proponents of dispositional accounts argue that not all evil people are necessarily evildoers, but an evil person is someone who is disposed to evil in a particular way. Act-based dispositional accounts argue that an evil person is someone who is disposed to perform evil actions. As we will later see, there are also affect-based and motive-based dispositional accounts in which an evil person is someone who is disposed to have certain kinds of feelings or motivations. For now, I will follow Russell in focusing only on act-based dispositional accounts in contrast to regularity accounts.

Act-based dispositional accounts can make sense of failed attempts at evil, since we can rightly judge that someone who attempts to perform evil actions is disposed to perform evil, even if their attempts ultimately fail. Dispositional accounts can also provide a more precise bad/evil distinction by pointing to an evil disposition as a relevant distinguishing characteristic between evil people and merely bad evildoers. While all evildoers can be said to be disposed to perform evil actions in some way, the kind of disposition necessary for evil personhood is one that is characteristic of the agent. That is, the disposition of an evil person is the reflection of a more stable character than the occurrent dispositions of mere evildoers. For example, Russell argues that an evil person is strongly disposed to perform evil actions, meaning that an evil person is markedly likely to perform evil actions compared to merely bad evildoers (Russell 2014, 156). The

disposition necessary for evil personhood is such that the evil person is more likely

overall to perform evil actions, while a merely bad evildoer is only weakly disposed and

thus less likely overall to perform evil actions.

Yet, dispositional accounts face problems that a regularity account does not, since

they imply that there can be blameless evil persons, and they can imply that most people

are evil given that dispositions to perform evil in some situations are widespread. In the

first case, because a dispositional account of evil personhood only requires that an evil

person is disposed to do evil, there can be evil persons who have never actually done

anything wrong, and are thus blameless. But judging someone to be an evil person is one

of our strongest forms of moral condemnation, so it seems at least odd to say that

someone who is blameless nonetheless deserves to be seriously morally condemned.

In this way, a regularity account has the advantage over dispositional accounts,

since it maintains that an evil person is someone who performs, or intends to perform,

evil actions. While failed attempts and ineffective intentions to perform evil do not count

as evil actions, they are blameworthy. Continuing the comparison drawn by Calder, under

Canadian law both murder and attempted murder are indictable offences for which the

perpetrator is liable (Criminal Code, s 463) and this is also the case for conspiring to

commit murder in the future (Criminal Code, s 465). While a failed attempt to murder

someone, and a plan to murder someone in the future, are not themselves instances of

actually committing murder, they are nonetheless legally blameworthy. Similarly, failed

attempts and intentions to perform evil are not, themselves, instances of actually

committing an evil action, but they are nonetheless morally blameworthy. If we maintain

that it is necessary for evil personhood that an evil person performs, or attempts to

perform, evil actions then all evil persons are morally blameworthy.

A second challenge to dispositional accounts is that it turns out that a large

majority of people are disposed to perform evil because of the significant extent to which

situational factors influence our behaviours. Social psychology provides compelling

evidence that this is the case, most notably in Phillip Zimbardo's Stanford prison

experiment (Zimbardo 2007) and Stanley Milgram's obedience experiments (Milgram

1974). Russell is particularly concerned with the latter and with what he calls "Milgram

scenarios" (Russell 2014, 170-1). Subjects of Milgram's experiments, known as teachers,

were asked to administer electric shocks to unseen fellow participants, known as learners

(who were confederates of the study and pretended to receive the shocks), for a study on

memory and learning. The teacher participants were asked by an experimenter in charge

to continue administering increasingly painful shocks, up to 450 volts, to the learner

participants whenever they gave an incorrect answer. Milgram's study showed that 65

percent of teacher subjects were willing to administer the maximum voltage just because

the experimenter requested that they continue, even when the learners could be heard

crying out in pain, begging for them to stop, and going ominously silent (Milgram 1974).

Russell tries to reconcile the intuition that most people are not evil with the fact that in

Milgram scenarios otherwise decent people are strongly disposed to perform evil actions

(Russell 2014, 161).

To solve this problem, Russell adopts a "fixed and autonomy-favouring dispositional account of evil personhood" (173), such that an evil person is one who is strongly and highly fixedly disposed to perform evil actions when in autonomy-favouring conditions. Where the strength of the disposition is measured by the likelihood that an individual will perform evil actions, fixity is a measure of how difficult it is to change that disposition over time (168). Fixity is determined by whether or not the disposition can be changed through typical means without difficulty, such as teaching, negotiation, rational persuasion, punishment and reward, or other various kinds of reformation attempts (178). Russell suggests that adopting a fixed dispositional account could allow that a large majority of us are strongly disposed to perform evil actions in Milgram scenarios, but since only a small proportion of us have a highly fixed disposition, most of us are not evil (170).

However, Russell acknowledges that it is not clear that our disposition to perform evil in Milgram scenarios could be easily changed. To address this, he indexes the disposition to autonomy-favouring conditions, so that someone is an evil person only if they are strongly and highly fixedly disposed to perform evil actions when under these specified sets of conditions. Russell describes autonomy-favouring conditions as conditions in which an agent is not under the kinds of pressures that would typically alienate her from her actions. In autonomy-favouring conditions the agent is able to do what she really wants to do and her actions are a reflection of her true self. Autonomy-limiting conditions, on the other hand, are conditions in which an agent might describe

herself as feeling alienated from her actions in some way, such as in Milgram scenarios

(173). Those who are disposed to perform evil in extreme circumstances are not as

deplorable as those who are disposed to do evil in autonomy-favouring conditions. This

allows us to claim that even though a large majority of people are disposed to perform

evil when under certain autonomy-limiting conditions, many of us are not so bad as to be

disposed to perform evil when in autonomy-favouring conditions.

The claim that an evil person is someone who is disposed to perform evil actions

in autonomy-favouring conditions begins to fall apart when we acknowledge that,

ultimately, there are no such perfect conditions. We can make general claims about

comparative autonomy-favouring, such as claiming that the conditions in Canada in 2015

are more autonomy-favouring than the conditions in Germany in 1943. But we cannot

reasonably claim that the conditions in Canada in 2015 are such that no autonomy-

limiting conditions remain. In fact, it seems that any social conditions whatsoever will

have some autonomy-limiting factors. For example, studies show that something as

simple as manipulating the smell of a room can have a surprising effect on our judgments.

In one case, inducing disgust by spraying a foul odour throughout a room caused

participants to evaluate gay men more negatively than they had prior to entering the room

(Inbar, et al. 2012). In another study, participants who reported their political attitudes in

the presence of a hand-sanitizer dispenser reported a less liberal political orientation, and

participants who were offered hand wipes before using a computer keyboard reported

harsher judgments of sexual acts (Helzer & Pizarro, 2011). If these kinds of influencing

factors qualify as sufficiently autonomy-limiting, then we would be unable to say with

any certainty that those who seem evil in our ordinary conditions are actually evil people.

Russell likely does not mean that the conditions need to be definitively and

perfectly autonomy-favouring, but he also does not provide an explanation for what

would and would not count as appropriately autonomy-favouring for evil, in light of

ordinary situational constraints. Another way that Russell describes autonomy-favouring

conditions is that they are conditions under which an agent is not "deceived, threatened,

coerced, or pressed" (Russell 2014, 173). But if an agent is deceived, threatened, coerced,

or pressed to a strong degree, then we think they are not evil because they are not

culpable.[3] That is, we would not say that you performed an evil action if you killed a

stranger because I demanded it from you while holding a gun to your mother's head. On

the other side, if an agent is only weakly deceived, threatened, coerced, or pressed such

that they are responsible for their actions, then this constitutes what most of our social

conditions are often like. Our judgments are made harsher by strong odours, our helping

behaviour increased by something as innocuous as finding a dime in a pay phone (Isen &

Levin, 1972), and our fondness for others can be swayed by whether or not they mimic

our facial expressions and mannerisms back to us (Chartrand & Bargh, 1999). If these

sorts of factors are enough to say that our ordinary conditions are autonomy-limiting, then

we would be unable to say that people like Ted Bundy or Dennis Rader are evil people.

---

[3] Calder makes this point in a forthcoming review of Russell, "Evil Persons."

We should reject an autonomy-favouring condition that suffers from this kind of

ambiguity if better alternatives are available.

So what are our alternatives? That is, how can we address the problem that a large

majority of people are disposed to perform evil when in certain conditions, but we do not

think a large majority of people count as evil persons? We could reconcile this problem

with a regularity account, by recognizing that either attempting or performing evil actions

is required for someone to count as an evil person. This way we can avoid the implication

that the majority of us are evil because we are disposed to do evil, since the fact that we

might be so disposed if we were in certain kinds of hypothetical conditions says nothing

about us as we currently are. If we neither perform evil, nor attempt to perform evil, then

we are not evil. A regularity account thus avoids both problems that face dispositional

accounts. A regularity account that includes intentions can ensure that there are no

blameless evil people, since performing, attempting, and intending evil actions are all

morally blameworthy. And a regularity account can ensure that situational dispositions to

perform evil does not imply that the majority of people are evil, since the majority of

people do not actually perform, or attempt to perform, evil actions.

But a regularity account remains too vague to provide a sufficient bad/evil

distinction and it fails to capture our intuition that evil personhood is meant to be at least

partly an explanation of why an evil person performs evil actions. Dispositional accounts

have the advantage over non-dispositional accounts in this way, since a characteristic

disposition to engage in evil behaviours can distinguish those with evil characters from

those who are merely bad evildoers. In the end, a dispositional account does seem more

practically and theoretically advantageous as a theory of evil personhood, despite its

problems. This is because a regularity account can only aggregate and categorize amounts

of evildoing, providing a weak distinction between bad and evil and providing no

satisfactory help for borderline cases. A dispositional account, on the other hand, is able

to provide a stronger distinction by focusing on the differences between bad character and

evil character more specifically. However, what counts as an evil disposition can vary. To

say that a view is a "dispositional account" is to say that it belongs under a broad heading,

and there are different theories available to determine which characteristics are

specifically required to say that someone has a disposition that is constitutive of an evil

character.

2.3. Consistency and Extremity

On a consistency account, an evil person is one who has characteristics of evil

consistently, or all of the time. Daniel Haybron argues that "evil persons lack morally

redeeming qualities of even the most modest sort" (Haybron 2002a, 63), such that they

have a "kind of monstrous consistency of character" with no good side and are "vile

through and through" (70). Extremity accounts argue that to be an evil person is to have

characteristics of evil to an extreme degree. Unlike consistency accounts, an evil person

can have a relative good side on an extremity account, but this good side is dwarfed by

the significant extent to which evil persons possess extremely vicious character traits.

Peter Brian Barry argues that an evil person is someone who has the worst kinds of vices,

such as cruelty and malice, and they have these vices to a significant degree (Barry 2009, 172).

Consistency and extremity versions of dispositional accounts are the most explicit proponents of the monstrous conception of evil personhood. The monstrous conception of evil personhood views all evil people as monsters, and monsters are those who are so horrendously terrible that we cannot recognize them as being anything like us or conceive of them as ever becoming better kinds of people. Both consistency and extremity accounts rely too heavily on a monstrous conception of evil to make sense of the distinction between borderline cases of evil persons and merely bad persons. They cannot adequately provide a bad/evil distinction because these accounts define evil personhood solely in terms of the worst possible monsters. As I will show, both Haybron's consistency account and Barry's extremity account lead to the counterintuitive results that those we would normally want to call evil people are merely bad, just because they are not monstrous.

Haybron's consistency account leads him to claim that if the typical Nazi war criminal had been a true friend and loving companion to some, then he was not evil (Haybron 2002b, 270). But this means that a Nazi war criminal who is a true friend and companion to some, and a callous murderer of others, would be a merely bad person. Haybron also argues that someone who kills innocent people to protect an endangered species of fish does not count as evil if she believes that murder is a regrettable, but permissible, means of achieving a morally worthy goal (Haybron 2002a, 72). But then a

violent racist is also a merely bad person if they believe that the eradication of one group

of people is a permissible means of protecting another. It may be right to say that an

unloving Nazi and a murderer who knowingly does wrong are worse sorts of people than

a loving Nazi and a murderer who tries to rationalize that they are doing good, but it does

not follow that companionship and false rationalizations of harmful actions means that

someone is not evil. Rather, it means that they are not monsters.

Barry advances the "modest proposal" for his extremity account, which states that

an evil person is the morally worst sort of person. This means that "for all persons, if X is

a morally worse sort of person than Y, then Y is not evil" (Barry 2010, 26). This claim

captures a primary supposition of the monstrous conception of evil personhood—a person

can count as evil only if they are the most horrendous sort of person that we can conceive

of, and if we can think of anyone worse, then they are probably not evil. One common

objection to Barry's thesis is that it implies that no one would count as an evil person,

since we can always imagine a potentially worse counterpart. Barry responds by pointing

out that "morally worse" means that an evil person (E) is comparatively worse than a bad

person (B), in the sense that (E) is more motivated to act wrongly than (B), and (E) is

more vicious than (B) (Barry 2012, 71-2). In other words, just because two serial rapists

differ in that one happens to be lazier than the other, so that the lazy rapist can only

assault one woman a week compared to the two a week that his more active counterpart

can manage, this does not mean that the lazy rapist is a morally better kind of person.

But we can imagine cases where someone is not the morally worse sort of person as Barry intends it, because they do have a comparatively worse counterpart, but we nonetheless still tend to think of them as evil. My earlier examples of Sam and Joe capture this problem. Sam, a malicious torturer who gets pleasure from other people's pain, is comparatively worse than Joe, a child pornographer with some sympathy for his victims. Sam is more motivated to act wrongly than Joe since Sam performs evil for his own pleasure while Joe performs evil for instrumental reasons, such as for money and employment. Sam is also more vicious than Joe, since Sam is malicious while Joe is not. Barry would contend that only someone like Sam could be called an evil person, and someone like Joe would be a merely bad evildoer.

According to Barry, a plausible conception of evil personhood will make sense of the psychological properties and states that constitute an evil character (66) which will explain why an evil person's evildoing is more than a mere accident (Card 2002, 22). In other words, a plausible theory of evil personhood will have to account for evil persons' motivations to perform evil. Barry challenges Haybron on this matter, since a consistency account only requires that an evil person has no good side or motivation to act rightly, but not that they necessarily have a bad side or are motivated to act wrongly. Evildoing on a consistency account, then, could be something of an accident since it is only a contingent matter that evil people happen to perform evil actions. On an extremity account, evil personhood "is a function of suffering from extreme vices, and being disposed to evildoing is partly constitutive of those vices" (Barry 2012, 66). This means, though, that

motivations to perform evil that are not the result of extreme vices are not enough to

count an evildoer as an evil person.

Another reason we might think that Joe is merely a bad evildoer while Sam is an

evil person is that a "plausible account of evil personhood must say something about the

affective states of evil people" (84) and so "must include an affective component to

accurately describe what evil people are like" (85). Barry argues, in agreement with

Russell, that an evil person will be someone who identifies with their evildoing. This

identification will be a kind of self-satisfaction of the evil person with their will to do

evil, which means that they will lack feelings of resistance or uncertainty (84). In other

words, in order for an evil person to identify with their evildoing, they will lack "morally

appropriate feelings prompted by recognition of their wrongdoings" (85), that is, they will

lack feelings of guilt or remorse. Barry concludes from this that an affective component

of an account of evil personhood would be that evil people are unaffected by a morally

significant range of choices that a decent person would find difficult and painful (86). For

Joe, since he does have some feelings of sympathy and he is resistant to causing more

pain than he thinks he needs to, it seems that he does not have the kind of affective state

necessary for evil. Except Joe does identify with his evildoing, in the sense that he does

not feel alienated from, or uncertain about, his actions. After all, Joe has convinced

himself that forcing young victims of human trafficking into child pornography is the

right thing to do, given that it can potentially alleviate the later trauma of prostitution. The

problem with Barry's account is that an evil person's lack of morally appropriate feelings

34

requires a recognition of their own wrongdoings. This means that an evildoer cannot

count as an evil person if they do not recognize that their actions are extreme wrongs.

This makes sense for Barry's modest proposal and the monstrous conception of

evil personhood. Someone who unremorsefully performs evil actions that they recognize

are evil is certainly a morally worse sort of person than someone who unremorsefully

performs evil actions because they do not recognize that their actions are evil. But, as was

the case for Haybron's examples and as Kekes has argued regarding the prevalence of evil

due to moral idiocy, it is implausible to suggest that simply believing that one is not doing

wrong means that one is not evil. It can mean, though, that they are not a monster.

In a somewhat ironic turn, some putatively evil persons tend to hold a view of evil

that is similar to a consistency or extremity account. In order to avoid recognizing

themselves as viable contenders for evil personhood, evildoers themselves point out that

they are not wholly bad people and that they do feel something like remorse. For

example, in an ethnographic study of perceptions of evil by sex offenders in prison

populations, one offender claims:

> Bad, yes, the things I've done. Evil? I don't fully agree with evil. But bad, yes. ... I'm
> not exactly sure what the word "evil" means but in my sense it's like ... there's bad then
> there's really, really bad which gets to the point of almost evil, you know, where you
> don't care. Step all over people ... you don't care about feelings any more. Feelings ...
> don't bother you. There's guilt. There's guilt after an assault. It's always there. Every
> time I've committed an assault, right after that ... initial rush is over, where you're
> coming down from [it] ... that guilt sets in, right? But you block that out too, like
> anything else. You know, "I didn't hurt them." [You] go through all kinds of
> rationalizations, minimizations, everything. Saying, "Okay, I didn't hurt them, all I did
> was have sex with them." You know, "it's not like the first time they had sex or anything
> like that." So you go through all these little things in your head, and it's okay. You
> know? "They're not dead. I didn't hurt them, all I did was have sex with them (Waldram
> 2009, 228-9).

When a sex offender defends himself against the label of evil by pointing out that he did

not rape virgins or kill his victims, and he felt guilty after every assault (though, clearly,

not enough to stop), we are likely to dismiss his claims as insincere at best. At worst, we

are slightly horrified at the idea that someone could believe that these things matter when

all is said and done. Yet, this kind of defence is precisely what both Haybron and Barry

propose. In another example, an offender points out that their actions were not motivated

by extreme viciousness, so it would be a mistake to call them evil:

> … as a child, sexual abuse was love. And that's what I learned. And I only felt
> comfortable with children. I didn't feel comfortable with adults because I didn't trust
> them, and it was my perverted way of showing love. Mentally I could distinguish
> between the right and the wrong of it, but the wrong always seemed to … win over.
> So   I   don't   think   I'm   evil   (Waldram   2009,   227).

By defining evil personhood only in terms of a monstrous conception of evil,

consistency and extremity theories cannot account for those who are worse than bad, but

not monstrously so. Nonetheless, though Haybron and Barry each present accounts of

what it means to be a monster rather than an evil person, they do indicate that a plausible

theory of evil personhood needs to account for other factors beyond habitual evildoing.

They point out that conditions of affect and motive can also play a significant role in

determining who counts as someone with an evil character rather than someone who is a

merely bad evildoer.

2.4. Affect-based and Motive-based Accounts

Affect-based accounts claim that evil persons have specific feelings or emotions

that distinguish them as evil. On a dispositional affect-based account, evil persons are

disposed to have certain sorts of feelings. Colin McGinn argues that an evil person is someone who "derives pleasure from pain and pain from pleasure" (McGinn 2003, 62). Because these are individuals who find other people's pain intrinsically pleasurable, they have "no other purpose than that of harm and destruction" (64-5). For McGinn, evil persons are those whose characters produce "malevolent motiveless action" the paradigm of which are sadistic torturers. This is because McGinn deals only with what he calls "pure evil" as opposed to "instrumental evil" and is unconcerned with "your average rogue, cheat, or traitor … with something to gain from his misdeeds" (63-4). In this way, McGinn seems to agree with Haybron's distinction between the evil person and the moral criminal.

It is clear that taking pleasure in another person's pain, particularly if we are the cause of that pain, is morally reprehensible. However, it is less clear that this kind of affective response is either necessary or sufficient to call someone an evil person. As Calder points out, we are just as likely to consider someone evil if they are the type of person who indifferently runs down pedestrians with their car, even though they do not take any pleasure in the harm (Calder 2003, 367). Moreover, taking pleasure in witnessing someone else's pain is not enough on its own to count someone as an evil person. McGinn argues that an evil person can be a mere "spectator of the suffering he relishes" and that he need not bring about that suffering to count as evil (McGinn 2003, 66). Yet there are cases of those who do feel pleasure in another person's pain, but whom we do not think are evil.

One compelling example is the case of the pseudonymously named Adam, a 20-year-old self-identified pedophile who harbours strong sexual desires for boys between the ages of three and seven, and girls aged five to eight. Adam has feelings that we would unequivocally call evil. The desire to sexually abuse children amounts to nothing less than desiring to do something that would bring the abuser pleasure from the child's pain. But Adam not only purposefully refrains from ever acting on his feelings, he also provides support to others like him to keep them from behaving wrongly. As a teenager, Adam established a support group for young pedophiles, emphasizing two important rules: members cannot have offended or have any intention to do so, and members must be committed to stopping the use of child pornography. As an adult, Adam works as an official advisor to the Moore Center for the Prevention of Child Sexual Abuse at Johns Hopkins University (Malone 2014). While we might find it difficult to allow that someone is a morally good person when they confess to once being addicted to child pornography, Adam's case is a clear example of someone whose evil feelings ultimately led to actions that minimized, rather than perpetrated, other people's significant harm. Whatever it is that we think of him, Adam is not an evil person.

Nonetheless, we might think that feeling pleasure from the harm that an evildoer causes themselves is what distinguishes them as an evil person. We don't consider Adam evil because he does not cause any harm to take pleasure in. But if he did cause harm, then we would think he was evil not just because of the harm that he causes, but because of the harm he causes *and* the pleasure he derives from it. One way to understand this is

38

to consider the role that evil feelings play in theories of evil action. Russell points to a category of theories of evil action that he calls "psychologically thick accounts" (Russell 2014, 69) which claim there is a psychological hallmark of evil action. On this view, hallmarks such as malice, sadistic pleasure, or knowing defiance of morality distinguish evil actions from other extreme culpable wrongs (79). Russell suggests that these accounts may arise from our preoccupation with paradigmatic evildoers in the form of serial killers like Ted Bundy. He notes that these serial killers "seem to be monstrous … in the sense that they are significantly psychologically different from ordinary people." By focusing on their example as archetypes of evil action, we might judge that other actions are evil in so far as they resemble them (70).

Malicious and sadistic torture in knowing defiance of morality is certainly an evil action. However, reluctant or indifferent torture for purely instrumental reasons also seems to be comparatively evil, as morally worse than a mere wrong. Malice, sadistic pleasure, and defiance are all exacerbators of culpably wrong actions, and that may be why the presence of these hallmarks are considered necessary for evil (Russell 2014, 81, 83, 88). Proponents of psychologically thick accounts define the concept of evil action by the presence of these exacerbators. They recognize that there is something additionally morally despicable about maliciously inflicting extreme harm, or deriving sadistic pleasure from that harm, and they propose that this difference marks the boundary between extreme culpable wrongs and evil actions. Affect-based accounts of evil personhood like McGinn's are similar, in that they attempt to understand evil personhood

by focusing on affective exacerbators as a determinant of evil character. McGinn is right

to note that there may be something additionally morally despicable about an evildoer

who also thrives on his evildoing, but he is wrong to conclude that deriving pleasure from

pain and pain from pleasure are necessary or sufficient conditions for evil personhood.

Instead of relying on affect, an account of evil personhood might focus on an

evildoer's motivations to distinguish them as evil rather than bad. Motive-based

dispositional accounts argue that an evil person must be disposed to be motivated in a

certain way. For example, Calder argues that an evil person is someone who has a regular

propensity for a certain kind of motivational state that he calls "e-desire sets" (Calder

2003, 366). This motivational state is, first, a determinant of what counts as an evil action.

An e-desire set, according to Calder, is a combination of desires for other people's

significant harm, or objects and states of affairs inconsistent with their being spared harm,

together with the absence of a desire that they not be harmed. These e-desire sets

constitute the motivation that is necessary for evil actions, and to have an evil character

"is mostly just a matter of having a consistent propensity for e-desire sets" (373).

Motive-based accounts have the benefit of being able to deal with cases of

culpable ignorance, which previous accounts took to exculpate someone from evil. As we

have seen, Haybron suggests that someone who kills innocent people to protect an

endangered species of fish is not evil, because she mistakenly believes that her actions are

for a morally worthy goal (Haybron 2002a, 72), and Barry's account implies that an

evildoer cannot count as an evil person if they do not recognize that their actions are evil

(Barry 2012, 85). Both of these claims, though, are implausible. In certain cases, an

individual may be motivated to perform evil actions by false beliefs that they sustain by

self-deception. Self-deception is described by David Jones as the evasion of self-

acknowledgement of some truth that is consciously engaged in and within the control of

the self-deceiver (Jones 1999, 81). Jones claims that a sustained project of this kind of

evasion results in a state of self-deception that can be characterized as "willful

ignorance." The state of ignorance is "wilful" because it is purposefully and intentionally

sought, and once achieved, is maintained and protected (82). Willful ignorance is thus

culpable ignorance, since the self-deceiver can be held responsible for initiating and

sustaining their false beliefs. Jones outlines how culpable ignorance can be achieved

through several tactics of self-deception, including the avoidance of explicit thoughts

about the beliefs in question, by distracting ourselves with rationalizations in favour of

our false beliefs, and by systematically failing to make inquiries and ignoring available

evidence in order to avoid acknowledging that our beliefs are not reasonable.

Calder's account allows that someone can be an evil person when they are

culpably ignorant about the worthiness of the goal for which their evil actions aim, or

they are culpably ignorant about the amount of harm that their actions cause (Calder

2015, 122). Someone can still count as an evil person even if they hold the false belief

that when they are causing significant harm it is justified for some end, or the false belief

that their actions do not cause significant harm. According to Calder, "false beliefs

preclude us from doing evil only if we are not responsible for them" (121). To be

responsible for our false beliefs is to engage in a deliberate process of self-deception by ignoring available evidence or avoiding an unbiased assessment of that evidence (122).

False beliefs can be an indication that someone is not evil if the individual is non-culpably ignorant. Those who are non-culpably ignorant are those who have false beliefs for which they are not responsible. Calder gives the example of Sarah from the fictional town of Usville. Sarah grew up in Usville, an isolated and racially homogenous community, and through her entire life was warned about the dangerous and evil inhabitants of Otherville. Sarah was taught that Othervillers are demons who want nothing more than to torture and kill her. However, Sarah has not been told the truth, and people from Otherville do not want to cause her or her community harm. If one day Sarah comes across an Otherviller, Calder argues that it would not be evil for her to cause him significant harm (Calder 2015, 124-5). Sarah falsely believes that causing harm to an Otherviller is for a worthy goal—to protect herself and her community. She is wrong, but she is not responsible for being wrong since she has no good reason to think otherwise.

In cases of putatively evil people it is unlikely that we will find many cases of the non-culpably ignorant. Any false belief that leads to a propensity to be motivated to perform evil actions will likely be unreasonable beliefs for the agent to hold. Cases like Sarah's long-term systematic indoctrination aside, we are more likely to be dealing with self-deceptive evildoers who are culpably ignorant. We can consider the case of the two serial rapists again, but this time instead of a difference in laziness, one of the rapists is motivated by false beliefs. The first rapist is aware that his actions cause his victims to

42

suffer significant harm, but he just doesn't care about it, or he takes pleasure in that fact.

The second rapist is motivated to sexually assault women because he believes that his

victims implicitly want him to. He believes that all women are too shy and submissive to

admit when they want sex, so they all secretly want him to be forceful. It is clear that we

are unwilling to say that a serial rapist is not evil just because he falsely believes that his

victims secretly consent to the abuse. On a motive-based account, we can make sense of

why the second rapist is also an evil person. Both rapists are motivated to cause someone

else's significant harm for an unworthy goal. The second rapist falsely believes that he

does not cause significant harm because his victims consent. Or he might acknowledge

that some harm is caused but falsely believes it is for the worthy goal of liberating

sexually repressed women. In either case, these are not reasonable beliefs to hold in light

of available evidence, including his victim's cries, and the prevalent discussions in the

media of the harms caused by sexual assault.

Culpable ignorance, as one kind of motive for evil, can reach beyond the limited

monstrous conception of evil by including those who are not recognizable as

irredeemable monsters, but are still reasonably considered to be evil persons. On a

motive-based dispositional account, an evil person is someone who is disposed to have a

certain kind of motivational structure which can lead to evil actions. But a motive-based

dispositional account would mean that an evil person is not necessarily ever motivated in

this way. Since they need only be disposed to have the right motives, an evil person could

be the kind of person who is likely to have the motivations or e-desire sets in some

situations, but these motivations may never occur for them. It is not hard to conceive of a

number of scenarios in which many of us would be appropriately motivated to perform

evil, not the least of which might include someone offering us just the right amount of

money or power at the exact moment we find ourselves facing hard times. If all that is

required for evil personhood is that we would be motivated if the right situation came up,

then many of us will count as evil people. Similarly, the possibility of blameless evil

people is increased to an implausible extent if we allow that someone who has never done

as much as desire someone else's harm can count as an evil person.

We might conclude from this that Russell was right to index dispositions to certain

kinds of circumstances, such as autonomy-favouring conditions. In this way, an evil

person would be someone who is strongly disposed to be motivated in the right way when

they are in the right conditions.[4] But the problem remains that what counts as the right

conditions is too ambiguous to offer any real help. As we have seen, an autonomy-

favouring condition falls short in light of the evidence from social psychology, which tells

us that our autonomy is often limited by a wide variety of factors in ordinary social

circumstances. Indexing dispositions to normal circumstances will face the same kind of

ambiguity. We are even less clear about what makes certain circumstances "normal

conditions" than we are about what makes conditions autonomy-favouring.

This brings us to an important consideration: a motive-based account does not

need to be a dispositional account. Instead, we could have a motive-based regularity

---

[4] By "motivated in the right way" I mean that evil persons have a particular kind of motivational structure
that qualifies them as evil. Obviously, I do not mean that they are motivated in a way that is morally correct.

account. This would mean that an evil person is someone who has the right motivations and desires on a regular basis. Calder's later account fits this view, where an evil person is someone who is regularly motivated to perform evil, and performing evil means to cause, allow, or seriously risk causing or allowing, someone else's significant harm for an unworthy goal (Calder 2015, 126). In a similar way, Singer argues that an evil action must come from "a motive or intention to do something that one knows or believes, or has reason to believe, is horrendously wrong or bad" and an evil person is someone "who knowingly performs, wills, or orders such actions" (Singer 2004, 196-7). A motive-based regularity account would avoid the problems associated with dispositional accounts by requiring that an evil person is someone who is regularly, rather than potentially, motivated to perform evil actions.

A motive-based regularity theory can account for both monstrous and non-monstrous evil persons, as well as the merely bad. Being regularly motivated to cause someone else's significant harm for an unworthy goal is a plausible distinguishing characteristic of an evil person. In comparison, a merely bad person would be regularly motivated to cause only minor harms or uncharacteristically motivated to cause significant harm for an unworthy goal. A motive-based regularity account can also provide a solid explanatory ground for act-based dispositional accounts. If an evil person is someone with a certain kind of motivational structure, then that motivational structure will tell us why evil people are disposed to perform evil actions.

2.5. The Take Away

An act-based regularity account states that someone is an evil person only if they regularly perform, or intend to perform, evil actions. An act-based dispositional account states that someone is an evil person if they are disposed to perform evil actions, but they need not ever actually perform an evil action. A regularity account has some advantages over a dispositional account, since on a dispositional account there can be blameless evil people who have never done anything wrong. A regularity account does not allow that there can be blameless evil people since performing, attempting, and intending evil actions are all morally blameworthy. But act-based regularity accounts attribute an evil character to someone in virtue of a vague or arbitrary frequency of evildoing that is unhelpful when applied to borderline cases. Dispositional accounts, on the other hand, can tell us something about how evil people are relevantly distinct from merely bad people, by pointing to the relative likelihood that an evil person will act, feel, or be motivated in certain ways compared to a merely bad evildoer. However, certain dispositional accounts can go too far when attempting to define what a sufficiently evil character would be, such as with consistency and extremity versions. These accounts can tell us which evil people may be monsters, but they cannot tell us who are plausibly evil persons compared to those who are merely bad.

Affect-based accounts claim that an evil person has certain evil feelings, such as deriving pleasure from other people's pain. This type of account fails to capture the full range of individuals that we tend to think are evil people. Someone can still plausibly

46

count as an evil person even if they are unaffected by the significant harm that they cause,

such as someone who indifferently runs down pedestrians with their car (Calder 2003,

367). Additionally, taking pleasure in witnessing someone else's pain is not enough to

count someone as an evil person. A pedophile like Adam might derive immediate pleasure

from viewing child pornography, but he also feels deeply ashamed about that fact and

purposefully restrains himself in the future. Calder gives a similar example of a bystander

to a violent assault. To the bystander's surprise and horror, he takes pleasure in witnessing

the pain of the victim being assaulted. Calder argues that if the bystander is disgusted

with himself for taking pleasure in harm and if he has no desire to witness harm in the

future, then it is too harsh to call him evil, even if the bystander happens to be so

constituted that he will always take pleasure in witnessing harm inadvertently (Calder

2003, 368).

   Motive-based accounts provide the most plausible distinction between merely bad

people and evil people, and they can help explain why evil people perform evil actions. If

an evil person is someone with a certain kind of motivational structure, such as desiring

other people's significant harm for unworthy goals, then that motivational structure will

tell us why evil people are disposed to perform evil actions. But a dispositional motive-

based account means that an evil person does not need to ever have these motivations to

count as evil. On a motive-based regularity account, an evil person is someone who has

the relevant motivational structure on a regular basis. This is more plausible than the

dispositional alternative because it requires that evil people are motivated in the right way

often enough, which means that it is not sufficient for evil personhood to be only disposed

to have the right sorts of motivations. Furthermore, having the relevant motivational

structure on a regular basis is a plausible distinguishing characteristic of an evil person.

On Calder's motive-based regularity account an evil person would regularly desire other

people's significant harm for an unworthy goal, compared to a merely bad evildoer who

causes significant harm for an unworthy goal uncharacteristically.

But a motive-based regularity account implies that some who have this kind of

motivational structure might nonetheless never act from their motivations and remain

blameless evil people. Even though the motivation must occur regularly, an evil person

need not ever act from their motivations, and so, an evil person need not ever do evil to be

evil. To avoid this implication, we might consider a hybrid regularity account that is both

motive and act based. A hybrid motive-act-based regularity account would require that an

evil person is regularly motivated to perform evil actions, and they either do perform, or

attempt to perform, those actions. In this way, we could say that an evil person is

someone who is regularly and effectively motivated to perform evil, where an effective

motivation is one that moves the agent to action. A hybrid regularity account would be

able to meet the challenges faced by both regularity and dispositional accounts. We can

distinguish an evil person from a bad person in terms of a characteristic motivational

structure, while at the same time ensuring that there are no blameless evil people by

requiring that an evil person is effectively motivated to perform, or attempt to perform,

evil actions.

Yet, even though a hybrid regularity account does solve the problems that we have been considering thus far, stating that an evil person must perform evil actions is a contentious claim. In the next chapter, I will respond to those who argue that the performance of evil actions is not a necessary condition for evil personhood. I will address counterexamples to this view and argue that the performance of evil actions is both an epistemic and moral requirement for the concept of evil personhood. I will then show the practical advantages to adopting a hybrid motive-act-based regularity account, particularly as it relates to the concepts of reformation and redemption for evil persons.

Chapter III: The Right Conception

3.1. A Hybrid Regularity Account

A hybrid motive-act-based regularity account states that an evil person is someone who is regularly and effectively motivated to perform evil actions. In other words, an evil person is someone who is regularly motivated to perform evil actions, and they either do perform, or they attempt to perform, those actions. Drawing on Calder's account of the motivational structure of evildoers, I contend that an agent is regularly motivated to perform evil actions if she regularly desires someone else's significant harm, or states of affairs that are inconsistent with others being spared significant harm, for an unworthy goal (Calder 2003, 366). To be effectively motivated to perform evil actions means that the agent's desires are strong enough that they motivate her to act on them. I am using Frankfurt's definition of "effective desire" as the desire which "moves (or will or would move) a person all the way to action" (Frankfurt 1971, 8). It is enough that the agent attempts to perform an evil action for the motivation to be considered effective, so even if her attempt turns out to be unsuccessful she can still count as an evil person. However, intending to perform an evil action without at least attempting to act on that intention will be insufficient. As Frankfurt argues, even if someone has an intention to act in some particular way, they may act in a different way when their desire to do the original action is weaker or less effective than another conflicting desire.

This means that someone is an evil person if (1) they have a consistent motivational structure such that they regularly desire other people's significant harm (or

states of affairs inconsistent with others being spared significant harm) for an unworthy

goal, and (2) they are effectively motivated such that they act from a desire for other

people's significant harm (or states of affairs inconsistent with others being spared

significant harm) for an unworthy goal by causing or allowing (or attempting to cause or

allow) someone else's significant harm. However, (2) is a contentious requirement for the

concept of evil personhood. Russell argues that such hybrid accounts fail because they

cannot allow for the possibility that certain evil persons are merely deterred from doing

evil due to fear of punishment or inability (Russell 2014, 162). Haybron gives the

example of a malevolent quadriplegic who can neither act nor communicate, but who

"wishes nothing more than the greatest suffering" for others. Haybron argues that this

kind of person is an evil person despite her inability to perform evil actions (Haybron

2002b, 264). Calder similarly states that someone can count as an evil person even if he is

too cowardly or incompetent to perform evil, since neither cowardice nor incompetence

makes someone a better kind of person (Calder 2003, 367).

The account I offer states that an evil person is someone who is regularly

motivated to cause or allow someone else's significant harm for an unworthy goal. But I

also argue that either attempting or performing evil actions is required for someone to

count as an evil person. So it is not enough to say that someone is evil because they are

disposed to perform evil actions, if they have the motivation to perform evil but that

motivation is always ineffective. Calder argues that the motivational structure to do evil

must include effective desires. That is, the desire for someone else's significant harm,

either as a direct goal or as a by-product of some other goal, must be stronger than the desire not to cause someone else's significant harm. If our desire for our victim's harm is outweighed by our desire for her well-being, then we cannot do evil. Evil acts only occur when the desire for our victim's significant harm wins out among our competing desires and we are effectively motivated to perform evil (Calder 2003, 365).

This can be expanded into evil personhood easily enough, since we are unlikely to think that someone is an evil person when they only weakly desire someone else's significant harm, but strongly desire that others are not harmed. I might be the kind of person who is prone to quick bursts of revenge fantasies, so that whenever someone insults me I immediately imagine shoving them out into traffic. But since I am also the kind of person who is more strongly motivated to avoid causing harm to others, I never attempt to act on these desires, and so do not count as evil. This is also why we are less likely to think that Adam, the purposefully restrained pedophile, is an evil person. While Adam does have regularly occurring desires for an unworthy goal that is inconsistent with other's being spared significant harm, he has stronger regularly occurring desires to prevent significant harm to others.

The more pressing challenge comes from cases of those who are dissuaded from performing evil, not because of a competing desire that others are spared harm, but because of a competing desire to avoid punishment or because of external constraints on carrying out their intentions. Not being effectively motivated to perform evil because of a competing desire to spare other people harm speaks favourably of a person's moral

character, so it makes sense that we do not think they are bad enough to count as evil. If

someone is not effectively motivated to perform evil because they want to avoid

punishment, or because they are just not physically able to follow through, we do not

have any reason to think that they are a morally better kind of person. But if an evil

person needs to be regularly and effectively motivated to perform evil actions, then

stronger motivations to refrain from performing evil due to fear of punishment or the

recognition of a disability will mean they are not effectively motivated and thus not evil.

Russell argues that a basic intuition of evil personhood is that not every evil

person performs, attempts, or even intends to perform evil actions (Russell 2014, 150).

According to Russell, we could imagine a perpetually frustrated misanthrope who lives in

a society that can efficiently detect and punish extreme wrongdoing so much so that the

misanthrope never even forms the intention to perform evil, let alone actually commits

any kind of evil action. The perpetually frustrated misanthrope is still an evil person,

Russell claims, despite his lack of evil intentions, since he would have those intentions

and he would perform evil actions if he were in a different environment. Russell argues

that the perpetually frustrated misanthrope has a disposition to perform evil actions in

autonomy-favouring conditions, so even though he never does as much as form the

intention to act, the fact that he is disposed to do so when under other conditions is

enough to say that he is an evil person.

Formosa, responding to Haybron's example of the malevolent quadriplegic, argues

that "there is an important moral difference between merely wishing and fantasizing

about evil … and actually intending and inflicting evil" (Formosa 2008, 235). Those

deterred by fear of punishment and inability can wish and fantasize, but since they do not,

or cannot, attempt and perform evil actions, they do not count as evil. One reason to think

so is that we can have just as many false beliefs about why we do not act as we have for

why we do. For example, a principled and self-interested amoralist might claim that she

would kill her competition if it were the most practical way for her to achieve her rational

self-interests. But since murder is rarely, if ever, a practical means to rational self-interest,

because the risks of being caught and punished outweigh the benefits of killing another

person, she never seriously intends to murder her competition. The principled amoralist

might believe that this is the only reason why she never intends murder. We might

conclude that the principled amoralist is the kind of person who would kill others for her

own gain if she could get away with it, and so she is an evil person because she is

motivated to perform evil. Yet, if the principled amoralist should find herself in a position

where murder really is the most practical and rationally self-interested thing for her to do,

it is unlikely that her principled position alone will cause her to kill another person.[5] In

this way, the principled amoralist falsely believes that she does not kill because of her

circumstances, and she falsely believes that she would kill if her circumstances were

different.

---

[5] For one reason, our responses to moral infractions are not often the product of careful and deliberative reasoning, but are more like automatic and intuitive aversions to extreme wrongdoing. See Joshua Greene's *Moral Tribes* (2013) and Jonathan Haidt's *The Righteous Mind* (2012) for detailed explanations of this view.

Or consider the case of the wronged vengeance-seeker. A mother whose son was killed in a drunk-driving accident obsessively plans to exact revenge on the perpetrator by killing his child. But since the perpetrator committed his crime as a teenager, she has to wait ten years before she can carry out her plan. For ten years the mother is consistently motivated to perform evil on the false belief that she is justified in killing the child of someone who killed her own child. After ten years, the vengeance-seeking mother goes to the home of the perpetrator with the intention of killing his young daughter. However, once she is there, and once she is in the presence of an innocent toddler, the vengeance-seeking mother discovers that she cannot actually carry out her plans. The mother had spent a decade imagining and fantasizing about killing the child, but it wasn't until she began to implement her plan that she realized she was unable to go through with achieving her goal.

In the cases of both the principled amoralist and the wronged vengeance-seeker, there is a kind of moment of truth of evil personhood that can only be revealed in the process of actual evildoing. If the principled amoralist kills her competition, or if the vengeance-seeking mother goes through with killing the child, then we have reason to think that they are each evil people. If they do not, then it seems too strong to call them evil, even though prior to the moment of truth they both held persistently strong beliefs and desires that motivated them to perform evil actions. As we have seen, possessing false beliefs alone does not make someone a morally better kind of person.[6] I argue for

---

[6] Specifically, pages 35-38

the additional claim that false beliefs alone do not make someone a significantly morally

worse kind of person, either. What matters in both cases is what the agent ultimately *does*.

An agent who performs evil while falsely believing that she does not is not a significantly

better person than someone who performs evil knowingly. Similarly, an agent who

refrains from performing evil while falsely believing that she would perform evil is not a

significantly worse person than someone who knowingly refrains from evil.

But what about cases, like the principled amoralist, who would kill if they could?

That is, if the principled amoralist does not have false beliefs about her reasons for

refraining from evil, then does she count as an evil person even though she will likely

never perform evil actions? I maintain that she does not count as an evil person until she

fully engages in the process of actual evildoing. This is because the moment of truth of

evil personhood is an epistemic requirement to rightly call someone evil. The only

reliable evidence we have that a person will do evil is that they either attempt to perform

an evil action and fail for reasons external to the agent, or they attempt and succeed. I can

ardently insist that I will go bungee-jumping once I can afford to do it. I can really believe

that I will, and really desire to do so in the abstract, planning and fantasizing about it for

months. But until I am hooked up, standing on a bridge, and looking over the edge, no

one can know if I really am the kind of person who can and will bungee-jump or if I am

someone who will back out in the last minute, going only as far as my fantasizing will

take me.

This indicates an important practical problem with dispositional accounts and motive-based regularity accounts. In one sense, we might want to say that the performance of evil is not a necessary requirement for the concept of evil personhood, but it is a requirement for the application of the concept. We can capture what is conceptually meaningful for evil personhood by stating that an evil person is someone who would perform evil actions under the right conditions, or an evil person is someone who is regularly motivated to perform evil even if the motivation is always ineffective. Someone is an evil person because they have the kind of character that would lead to evildoing, regardless of whether that evildoing ever does occur. The performance of evil actions is only a contingent requirement because we happen to be at an epistemic disadvantage about which actual people the concept of evil personhood really does apply to. As such, someone might object to my view as Haybron does that "the utility of a notion is not simply a function of how often we may be warranted in applying it" (Haybron 2002b, 278). If we had the ability to predict the future behaviours of suspicious individuals, then we could say that certain people are evil prior to the performance of evil actions. So, then, the performance of evil actions is not a necessary condition for the concept of evil personhood, but simply a current condition for any particular application of that concept.

There are two responses to this problem. In the first case, we could grant that the requirement for the performance of evil actions is simply a practical requirement for the extension of the concept of evil personhood. Evil personhood may be theoretically independent from the performance of evil actions, but it is practically dependent on the

performance of evil actions to be of any particular use. It is enough for the concept of evil

personhood that someone is appropriately motivated to perform evil actions, but we rely

on the actual performance of evil actions in order to retrospectively know who was

motivated in this way. Even if this were the case, we should consider including conditions

of use in our theories of evil if they prove to be beneficial. An important part of why we

care about the nature of evil is that we do apply these concepts to actual people and we

want good reasons to think that we are justified in doing so. If a practical requirement can

provide us with good reasons to think that we are right when we apply the concept of evil

personhood in any particular case, then we have reason to accept it as an important

component of that concept as long as our epistemic disadvantage remains. Dealing with

actual cases like the malevolent quadriplegic or the perpetually frustrated misanthrope

would illustrate this problem. We would be unable to say with any certainty whether they

are self-deceptive about their motivation to perform evil actions if they could, and so we

would be unable to know if they are relevantly different from the principled amoralist or

the wronged vengeance-seeker.

A second available response is to reject the assumption that the performance of

evil actions is merely a practical requirement for the application of the concept of evil

personhood. We can argue instead that it is a necessary component of the concept of evil

personhood that evil persons perform, or attempt to perform, evil actions. We do not say

that someone is a liar when they have never lied, or that someone is generous despite

never having the opportunity to share. An intrinsic feature of the moral concept "liar" is

that the concept only applies to individuals who have been dishonest, and an intrinsic feature of the moral concept "generous" is that the concept only applies to individuals who have shown themselves to be charitable. It may be the case that any particular individual displays traits that would lead us to think that they would lie when the opportunity strikes, but there is an important difference between saying, "I bet you would be a liar" and saying, "You are a liar". Dispositional and motive-based regularity accounts can bring us as far as the former, providing only enough reason to say "I bet you would be evil" while a hybrid regularity account allows us to say "You are evil" with sufficient confidence. This suggests that we cannot capture all of what is conceptually meaningful about evil personhood independent of evildoing.

A hybrid regularity account of evil personhood states that someone is an evil person if they are regularly and effectively motivated to perform evil actions. To restate in full, this means that someone is an evil person if (1) they have a consistent motivational structure such that they regularly desire other people's significant harm (or states of affairs inconsistent with others being spared significant harm) for an unworthy goal, and (2) they are effectively motivated such that they act from a desire for other people's significant harm for an unworthy goal by causing or allowing (or attempting to cause or allow) someone else's significant harm. (1) is insufficient for evil personhood. We are regularly motivated to perform evil actions without being an evil person when our desire for evildoing is trumped by our desire to refrain from evildoing. But (2) is also insufficient, since someone who is effectively motivated to perform evil actions is not

necessarily an evil person, but might instead be a merely bad evildoer. To capture this, we

can rephrase (2) to say that someone is an evil person if (1) and (2b) they are effectively

motivated such that they act from (1) by causing or allowing (or attempting to cause or

allow) someone else's significant harm for an unworthy goal. This means that the

evildoing required for evil personhood will be the result of a regular motivational

structure to desire to perform evil actions, rather than the result of an occurrent

motivation uncharacteristic of the agent.

A concern remains for this account, however. It cannot be the case that a

requirement for evil personhood is that the evil person is effectively motivated to perform

evil actions on a regular basis. If we require that the evil person regularly acts from their

evil motivations then evil personhood becomes an arbitrary designation, as we have seen

for the basic regularity account. There is no significant difference between the evil

person's first-time evildoing and their fifth, tenth, hundredth, and so on, that would

warrant a meaningful threshold between an evil person and a merely bad evildoer. It

would be strange to say that someone who kills three people is a merely bad person, while

someone who kills five people is an evil person, all things being equal. Similarly, the

difference between someone who kills every three years and someone who kills every

three months does not strike us as a relevant difference for calling one murderer bad and

the other evil, even though one is more regularly effectively motivated.

But there does seem to be a significant difference between first-time evildoing and

no evildoing. Going back to Calder's example, the would-be evildoer who refrains from

performing evil because he is a coward is meant to strike us as still potentially counting as

evil because cowardice does not make someone a better kind of person (Calder 2003,

367). But we do tend to think that courageous evildoers are worse than cowardly would-

be evildoers. A courageous evildoer who resolutely carries out his evil actions in the face

of punishment is a plausibly worse kind of person than a cowardly would-be evildoer who

has the desires to perform evil but is sufficiently deterred from ever acting on these

desires by stronger desires to avoid being caught and punished. For one reason, this

suggests that the courageous evildoer desires performing evil more strongly than his

cowardly counterpart. The stronger motivational desire of the courageous evildoer passes

a certain minimal threshold that can be understood as a relevant difference between

himself and the cowardly would-be evildoer. In the same way that Russell offered a

strength condition for his dispositional account to distinguish between evil people and

evildoers, we can consider a strength condition for a motive-based account. To determine

the relative strength of a motivation to perform evil actions we can point to whether or not

a desire to perform evil is effective, since an effective desire is stronger than other

competing desires. The performance of an evil action does make someone a worse kind of

person than someone who does not perform evil, all things being equal.

   To avoid an arbitrary threshold and to account for the difference between

evildoing and no evildoing, we can amend (2) further to say that someone is evil if (1)

and (2c) they are effectively motivated such that they act from (1) by causing or allowing

(or attempting to cause or allow) someone else's significant harm for an unworthy goal *at*

least once. In short, an evil person is someone who regularly wants to perform evil actions, and they make at least one attempt to achieve that goal.

A critic might respond that this has the disheartening implication that we cannot deal with evil people until evil has already been done. Part of what we might want an account of evil personhood to do is help us identify those most likely to be evildoers in order to prevent the occurrence of evil. But if we cannot know who counts as an evil person until they have already done evil, then this will be unhelpful. This account does have less of a pragmatic advantage in this respect, but we are nonetheless more likely to accurately pick out those who will do evil in the future. It is reasonable to think that those who have already at least tried to perform an evil action are more likely to continue to try, and to succeed, than those who have merely fantasized about it. This is because those who have already at least tried to perform evil have shown themselves to be more strongly motivated to perform evil than would-be evildoers whose motivations are weaker and ineffective. We are also less likely to misjudge people as evil on this account, while on dispositional and motive-based regularity accounts we often risk misjudgement since it is difficult to know in practice whether someone is appropriately motivated or disposed to do evil.

Admittedly, even on a hybrid regularity account we still face an epistemic disadvantage regarding the motivational structure of putatively evil people. The performance of an evil action alone will not tell us who is regularly motivated, compared to someone who is uncharacteristically motivated, to perform evil. So, the risk of

misjudgement is not altogether eliminated. Nevertheless, requiring effective motivations as part of the motivational structure of evil persons does provide us with more epistemic reliability in our judgments, as well as a more comprehensive conceptual account of the motivational structure necessary for evil personhood.

The account I offer states that an evil person is someone who is regularly motivated to perform evil actions, and they need to be effectively motivated to perform evil actions at least once. What we most often mean when we say that someone is an evil person is that they are the kind of person who can and will do evil. The account I offer can accommodate both moral and epistemic requirements for determining whether someone really is the kind of person who has an evil character. The only way to reliably know if someone is the kind of person who can and will do evil is if they have done, or have tried to do, evil. Additionally, this account can provide a strong bad/evil distinction, since a merely bad person will be regularly and effectively motivated to cause only minor harms or they will be uncharacteristically effectively motivated to cause significant harm for an unworthy goal. All in all, a hybrid motive-act-based regularity account is the most plausible account of evil personhood because it can support both common intuitions about which particular people are evil as well as philosophical intuitions about the conceptual requirements for a theory of evil personhood.

Chapter IV: Once Evil, Not Always Evil

4.1. Do We Need Monsters for Evil?

On a plausible hybrid motive-act-based regularity account of evil personhood we are able to see that there is a strong conceptual distinction between evil persons and irredeemable monsters. This view accommodates the contention that monsters are the conceptually worst possible kinds of people and they are incredibly rare. What this view rejects is the assumption that all evil people are monsters. Instead, an evil person is someone who is regularly and effectively motivated to perform evil actions. This means that an evil person can have a relatively strong good side, particularly if they are motivated by self-deception. An evil person can be someone like Joe, who sympathetically wants to cause the least amount of pain possible to young victims of human trafficking, but unreasonably believes that forcing victims into child pornography achieves this goal. Even without false beliefs, an evil person can be someone who is regularly and effectively motivated to perform evil actions, such as a mob boss who violently dominates over wealthy adults, while also being regularly and effectively motivated to perform good actions, such as working to alleviate poverty in his neighbourhood. A monster, however, is an evil person with no good side, who neither believes that their evildoing is morally justified nor has any morally praiseworthy characteristics. Consequently, one individual evil person may be comparatively less vicious than another. An evil person need not be extremely cruel or sadistic, while a

64

monster will be knowingly indifferent at the very least, if not overtly malicious.[7] A hybrid

motive-act-based regularity account of evil personhood is thus able to account for the bad,

the evil, and the monstrous.

This account can preserve our intuition that evil people exist and they can be

distinguished as morally worse sorts of people than the merely bad, while also offering

that evil people need not be so horrendously terrible that we cannot recognize them as

being anything like us or conceive of them as ever becoming better kinds of people.

Because this account distinguishes between an evil person and an irredeemable monster,

it implies that many evil people are potentially redeemable. If we no longer hold to a view

of evil persons as the monstrously worse people imaginable, then we can dissociate

ourselves from the idea that "once evil, always evil." As a result, this account provides us

with reason to consider additional ways to deal with evil persons beyond eradication

through death and imprisonment.

4.2. Reformation and Redemption

In particular, we have reason to consider how individuals may be aided in moral

and psychological development toward reformation and redemption. While many use the

terms "reformation" and "redemption" there is little work on these concepts themselves as

they are used in theories of evil. What is first needed is a conceptual analysis of

redemption, and an investigation into the distinction between what it is to be redeemed

---

[7] I do not propose a strict psychologically thick requirement for monsters, such that they are necessarily sadistic or malicious. But it seems plausible that a monster is necessarily cruel, in the sense that a monster will be at least coldly indifferent to the suffering that he causes others.

and what it is to be reformed. A minimal understanding of what redemption is might be simply that which occurs when someone who once counted as an evil person becomes a good person. But what is required for an evil person to become a good person just is whatever is required for redemption, so this understanding is unhelpful. What we need is an evaluation of the conditions for redemption itself.

We might think that redemption requires forgiveness, such that we are redeemed only if we are forgiven. An evil person may be forgiven by his victims, those affected indirectly by his wrongdoing, or by society in general. So we might think that an evil person is redeemed only when those relevantly related to his wrongdoing are sufficiently convinced that he deserves to be forgiven. We should be hesitant to allow this as either a necessary or sufficient condition for redemption, however. Russell notes that forgiveness alone would tell us more about the redeemers than the redeemed, and it is possible that a person can be forgiven by others while still counting as an evil person (Russell 2014, 209). Those who believe that forgiveness is a moral requirement on their part may offer forgiveness to an evil person even though he has had no change of character at all. The forgiven can still be someone who is regularly and effectively motivated to continue to perform evil actions and so forgiveness is not enough to call someone redeemed.

What's more, if a monstrous conception of evil lurks relentlessly in the background of our judgments, then we will be unwilling to grant even the possibility that an evil person is redeemable. It is highly unlikely in that case that we would be willing to see an evil person as forgivable, even when forgiveness may be warranted by a relevant

change in character. Card, referring to Roy Baumeister's studies on victim and perpetrator

perspectives, notes the "magnitude gap" in these perspectives where "victims are apt to

exaggerate the reprehensibility of a perpetrator's motives [and] to perceive monsters

instead of ordinary people" (Card 2002, 45). In Baumeister's original study, participants

were asked to read a story about a transgression and then to retell the story in their own

words, as if it had happened to them. Prior to reading, participants were randomly

assigned to take the role of either the victim or the perpetrator, along with a control group

who retold the story in the third person without identifying with either character. The

result was that both victims and perpetrators distorted their stories almost equally. The

number of discrepancies between the victim stories and the original story was almost

identical to the discrepancies in the perpetrator stories, and both were significantly higher

than the control group (Baumeister 2004, 90). Baumeister suggests that "the

incomprehensibility of the perpetrator's motive appears to be an important part of how

people define themselves as victims" (Baumeister 1990, 1002) and he shows that victims

tend to depict harm as greater and longer lasting than perpetrators do. While this is partly

because perpetrators often downplay the consequences of their actions, victims are

conversely prone to exaggerate the consequences (Baumeister 2004, 88). Likewise, then,

refusing forgiveness will have more to do with the unforgiving than the unforgiveable.

Victims of an evil person's actions, who do have a legitimate claim to suffering and

resentment, are nonetheless apt to inflate the viciousness of the perpetrator and are likely

to believe that the perpetrator could never sincerely regret their actions and change for the

better.

A better way to conceive of redemption is to consider the conditions an evil

person would need to meet in order to reasonably warrant forgiveness, regardless of

whether forgiveness is given. We can say that those who reasonably deserve forgiveness

as former evil persons will be those who are redeemed. To be redeemed will require that a

set of conditions are met that are independent of whether forgiveness ever follows in

actuality.

One condition for redemption is repudiation by the evil person of his actions.

Russell claims that repudiation requires that the evil person has feelings of remorse and

condemns his previous actions as wrong (Russell 2014, 194). As we have seen, Barry

argues that an indication that someone is not evil is if they have "morally appropriate

feelings prompted by recognition of their wrongdoings," such as guilt and remorse (Barry

2012, 85). Thomas similarly suggests that an evil person can deserve forgiveness when

"[t]he wrongdoer of the evil fully and willingly acknowledges both in private and in

public the wrongs that she or he has committed" and they are "genuinely contrite and

remorseful" with "deep shame and regret" (Thomas 2009, 117). Repudiation, then, is the

requirement that the evil person acknowledges and condemns his wrongful actions, and

displays genuine contrition for perpetuating those wrongs.

Along with repudiation, redemption also requires some form of reparation or

compensation to those who were wronged by the evil person's actions. Card argues that

"mere contrition should not be sufficient for reacceptance while reparable damage

remains" (Card 2002, 172), and Thomas adds that an evil person can only be forgiven if

they are "motivated by a sense of justice … to make right the wrongs committed, and so

to making amends" (Thomas 2009, 118). We might think, though, that compensation is

just part of what it means to be genuinely remorseful and to engage in sincere repudiation.

If I destroy a book that I had borrowed from you, either intentionally in a fit of anger or

through my own negligence, simply saying "Sorry" does not seem like enough. In order

for you to take my remorse seriously and believe my repudiation to be more than lip

service, you would expect that I propose to replace the destroyed book or offer some

other kind of compensation. In this way, compensation would not be an additional

condition for redemption, but rather a component of repudiation.

In terms of the evil person, however, there are two ways in which repudiation can

occur without compensation. It may be that the evil person's actions were sufficiently

complex so that those who were wronged either cannot be compensated, or it is difficult

to discover exactly who it is that ought to be given compensation. Russell argues that one

relevant sense of redemption is when an extreme wrongdoer becomes a good person who

is strongly disposed to do right rather than wrong (Russell 2014, 210). We could say that

a redeemed evil person who is unable to directly compensate his victims would need to be

even more motivated to do right generally. For example, a former member of a gang who

had engaged in the violent and fatal abuse of men and women living on the street cannot

directly compensate his deceased victims. Furthermore, his victims may either have had

no friends or family whom he can compensate in their place, or he is unable to discover

who they are. Instead, the repudiating evil person might work with police to dismantle his

former gang and contribute to bringing about awareness and assistance to other homeless

individuals. By doing right in this way, we might then count the repudiating evil person as

also indirectly compensating. In other cases, indirect compensation may occur without

any connection to the prior evildoing. That is, instead of working with police and raising

awareness of the homeless, the repudiating evil person may engage in other acts of

extreme good. For example, he might donate a significant amount of time and money to

another kind of charitable cause.

A more apparent way in which an evil person might repudiate without

compensating is when his repudiation is so strong that he is incapacitated by his own

guilt. That is, the guilt-ridden evil person recognizes the magnitude of the harms he has

committed and is overcome with the morally appropriate feelings of shame and remorse.

But he is overcome to such an extent that he believes there is nothing he could ever do to

adequately compensate for his terrible actions. The guilt-ridden evil person agrees with

Kekes' claim that even a repentant wrongdoer does not deserve forgiveness, since our

unforgiving blame is "directed against wrongdoers for having done a specific wrong" and

whatever "has been done, cannot be undone" (Kekes 2009, 502). Further, the guilt-ridden

evil person may go so far as to think that attempting compensation would be an insult to

his victims. He might purposefully avoid moral reparation because he is too ashamed to

consider himself worthy of redemption or forgiveness.

On the one hand, strong feelings of guilt and remorse on the part of a formerly

regularly and effectively motivated evildoer seems like persuasive enough evidence to

70

consider his character sufficiently transformed. Yet, his behaviour could also be described as a form of self-indulgent defeatism. Card argues that "the negative self-judgment in guilt and shame poses the danger of focusing perpetrators too much on themselves" (Card 2002, 209) such that it can become "a wallowing in self-deprecation that eventually dulls one's appreciation of the wrong itself" (208). She points out that excessive self-absorption "can hinder moral regeneration" which requires us to act (207), and for this reason "some willingness to forgive oneself, even for evil deeds, may be needed to sustain motivation to fulfill our obligations" (176). A guilt-ridden former evildoer who does not compensate does strike us as no longer qualifying as an evil person, but because his guilt leads to a self-absorbed inaction, it would be too strong to call him a good person. This suggests that we may want to consider additional ways in which an evil person can become non-evil without being redeemed.

Repudiation and compensation are two conditions for an evil person to become a good person. However, "evil" and "good" are not the only kinds of characters along the moral spectrum. As we have seen, an adequate theory of evil also needs to account for the significant range of individuals who are merely bad. Similarly for redemption, we need to provide a way to think about the more minimal shift from one who counts as an evil person to one who no longer counts as evil, but does not necessarily count as good. One way we can go about this is to consider the difference between redemption and reformation. While the idea of redemption carries with it connotations of an extreme moral conversion, the idea of reformation is more modest in comparison. To say that

someone is reformed is still to say that they have undergone a relevant change in their

character and behaviour, but it is not as strong a characterization as saying that someone

has been redeemed from the moral stain of their past behaviours. This suggests the

possibility that a person can go from evil to non-evil in two ways. Through redemption,

the evil person can become a good person, and through reformation, the evil person can

no longer count as evil without necessarily becoming good.

Russell claims that redemption might be described as what happens when an evil

person loses his disposition to perform evil, but while this may be necessary for

redemption, it is not sufficient (Russell 2014, 209). In a similar way, Thomas points out

that a wrongdoer who has become worthy of forgiveness is also one who has become

trustworthy in terms of not committing the wrongful behaviour (Thomas 2009, 124).

While an evil person needs to do more than simply cease to commit evil actions in order

to be considered a good person, losing their propensity to do evil is a clear prerequisite.

We can understand this loss of the disposition, motivation, or likelihood of committing

evil actions as reformation. Reformation is thus a necessary condition of redemption,

since an evil person cannot be considered good if they maintain their motivation to

continue to do evil, but it is not sufficient since redemption also requires repudiation and

compensation. Redemption means that the evil person has become a good person.

Reformation allows that the evil person becomes a *better* person, since they are no longer

motivated to perform evil. This is not enough to say that they are *good*, but it is enough to

say that they are no longer evil. Since reformation offers an intermediate stage between an

evil character and a good character, we can make sense of why we do not think that someone is still an evil person when they are guilt-ridden but uncompensating. They are reformed, in the sense that they will not continue to perform evil, but they are not redeemed from their prior evils.

Reformation requires, as a bare minimum, that someone who once counted as an evil person is no longer the kind of person who is regularly and effectively motivated to perform evil actions. It seems, though, that if this is true then reformation would also require repudiation. It would be strange to say that an evil person can be reformed without acknowledging and condemning his previous evildoing. One indication that someone is no longer motivated to perform evil actions is if they are incapacitated by strong feelings of guilt and shame over their previous evil actions. They are no more likely to perform evil actions than a minimally decent person because they no longer have strong desires to perform evil actions. In this case, we can say that they engage in strong repudiation. But another indication that someone is no longer motivated to perform evil actions is if they are more strongly motivated to refrain from performing evil actions. Deterrence by fear of punishment can be one such motivation, along with the motivation to engage in pro-social behaviours for benefits that evil actions cannot grant. This means that deterrence by fear of punishment, or instrumental pro-social behaviour for reward, are sufficient for reformation if these motivations to refrain from performing evil are stronger than the formerly evil person's motivations to perform evil actions.

In order for us to say that a formerly evil person's motivational structure is changed in this way, however, they will need to at least acknowledge and denounce their previous wrongdoing. In this case, we can say that they engage in weak repudiation. Weak repudiation does not require extreme condemnation of prior evil actions with guilt and shame as does strong repudiation. Instead, a formerly evil person weakly repudiates when he recognizes the undesirable consequences of his past transgressions. Strong repudiation is primarily backward-looking, in the sense that someone who strongly repudiates is focused on contrition for, and condemnation of, his past evil actions. In contrast, weak repudiation is primarily forward-looking. Someone who weakly repudiates recognizes that his prior evil actions are no longer desirable and he is motivated to behave differently in the future. For example, someone who thinks that the risks involved in extreme wrongdoing are no longer worth it could be described as weakly repudiating, since he is motivated to avoid the future performance of those actions.

Either strong or weak repudiation is required for reformation because it is unlikely that someone who does not denounce their prior evildoing will be reliably motivated to refrain from performing evil in the future. Cases of self-deceptive evil persons whose evil actions are motivated by false beliefs are clear examples of this. If those false beliefs are not recognized as the product of self-deception and denounced by the evil person, then they are likely to continue to self-deceive and perform evil actions. Furthermore, while reformation is not the strong moral conversion of redemption, it is still meant to capture a relevant change in the formerly evil person's internal motivational structure, which

74

requires at least weak repudiation. It would be implausible to say that every evil person

who is imprisoned is reformed just because they are unable to continue performing evil

actions, since some will continue to strongly desire to do so. Simply being imprisoned

cannot count as reformation if the evil person's internal motivational structure remains the

same but is impeded only by external constraints.

　　　We might think that weak repudiation is insufficient for reformation because it is

not a plausible indication that the evil person's motivational structure has reliably

changed. To see how weak repudiation can indicate that the motivation to perform evil is

reliably altered, consider the example of the "Today/Tomorrow" program instituted by the

Mendota Juvenile Treatment Center (MJTC) in Wisconsin. MJTC is a juvenile

correctional institution designed to deal with those who have been deemed uncontrollable

and resistant to treatment at other institutions. Youth sent to MJTC are those who are

severely violent, with over half convicted of a violent felony offense (Kiehl 2014, 219-

20). Treatment at MJTC uniquely aims at developing basic pro-social bonding with the

youth, compared to punishment-focused treatment found in most other institutions (221).

The "Today/Tomorrow" program utilizes a positive reward structure of contingency

management, where youth learn that if they are good today, then they will earn positive

reinforcement tomorrow. Rewards are graduated, such that they range from small rewards

of candy bars to larger rewards like the right to play video games in their cells, and youth

earn rewards by consistently engaging in positive social behaviours with staff and other

inmates (222).

In a series of follow-up studies on the effectiveness of the program over a five-year period, the MJTC program resulted in a 35% reduction in recidivism. That is, while 98% of similar youths who were not treated at MJTC were arrested for a new crime within four years of being released from a juvenile prison, only 64% of the MJTC youth were. More importantly, MJTC youth were more than 50% less likely to be convicted of a violent crime after two years, and 45% less likely after four years to have been convicted of a new violent felony (223). In the four years after their release, the non-MJTC youth had killed sixteen people, while the MJTC youth had killed none (224).

A relevant consideration here, as Michael Caldwell notes, is that these changes might be based solely on a self-interested calculation that certain positive interpersonal interactions are simply more rewarding than deception and aggression (Caldwell 2013, 224). Caldwell points out that these studies do not provide any indication of whether or not the internal characteristics of MJTC youth were radically changed such that they would have strongly repudiated their former behaviours. What the results do show is that a rewards-based treatment program for violent high-risk re-offenders decreased their propensity for violence and aggression (224). That is, by promoting and reinforcing a stronger motivational system to refrain from wrongdoing in those who had been strongly motivated to do wrong, the program effectively changed the motivational structure of at least some would-be evildoers by encouraging weak repudiation.

This example does suffer from drawbacks, particularly the fact that it centres on the changes of youth offenders who are arguably more malleable. It might be the case that

76

these young offenders were merely potential evil people, and the Today/Tomorrow

program succeeded in an early intervention to keep some from becoming evil people later

in life, but it did not succeed in reforming those who already were evil. However, my

point is not that this is an example of reformation, but that it indicates a way in which a

motivational structure underlying extreme wrongdoing can be reliably altered without

evidence of strong repudiation. What is particularly compelling about the MJTC example

is that a significant number of participating youth had scored high on the Youth

Psychopathy Checklist. The average score was 28 out of 40, which means that the

majority of youth at MJTC were in the severe range on the test and at a high-risk for

developing a full diagnosis of psychopathy as an adult (Kiehl 2014, 220). Evidence

suggests that psychopathy is not only a problem with behaviour, such as being impulsive

and manipulative, but it is also a neurocognitive disorder such that the capacity for

feelings of empathy and shame are significantly reduced (Anderson & Kiehl 2013, 131-

45). If some of those who were effectively deterred were also those with psychopathic

tendencies, then it is even more likely that their motivational structure was changed with

only weak repudiation.

This means that an evil person can go from evil to non-evil in two ways. Through

redemption, the evil person can become a morally good person. They can do this by

strongly repudiating and compensating for their previous evil actions. To strongly

repudiate, they will acknowledge and condemn their wrongs with an accompanying sense

of guilt and shame. And they will either directly compensate those they have wronged, or

77

indirectly compensate for their wrongdoing by regularly performing right actions more

generally. Through reformation, the evil person can become merely non-evil by either

strongly or weakly repudiating their previous evil actions and becoming reliably

motivated to refrain from performing evil. To weakly repudiate, the formerly evil person

will acknowledge and denounce their wrongdoing, but they need not have the same

feelings of guilt and shame involved in strong repudiation. This is because reformation is

not meant to capture the strong moral conversion required for redemption, but is a weaker

form of moral progression away from evil. We can conclude from this that evil

personhood is not necessarily fixed and there are viable alternatives for dealing with those

who count as evil persons beyond hostility and isolation. By conceptualizing the nature of

evil personhood in terms of the motivational structure to perform evil actions, we are able

to consider the myriad of intervention techniques that have been shown to successfully

alter these motivational structures and the behaviours of putatively evil people.

Chapter V: Conclusion

When we think about the nature of evil personhood, we should not think about cases of "good versus monstrous," but rather borderline cases of "bad versus evil." Evil is one of our strongest forms of moral condemnation, and so we need a theory that can explain why evil remains a unique and useful concept among our other forms of moral opprobrium. By employing the comparative methodology in borderline cases we are able to say with more precision what it is that distinguishes an evil person from a merely bad person. Because borderline cases sit on the threshold between very bad and evil, directing our critical reflection to cases of these kinds can help us to think more clearly about the characteristics and behaviours that we want to call evil and those that we are comfortable calling merely bad. In this way, we arrive at a useful and satisfying theory of evil personhood.

This method also yields the significant result that evil persons and irredeemable monsters are conceptually distinct. An evil person is minimally someone who is worse than a merely bad person. But not all who are worse than bad are monstrously terrible. Those who claim that evil persons are monsters in human form, who are different from the rest of humanity (Cole 2006, 13), between the human and demonic (Haybron 2002b, 277), and are freakish aberrations of nature (Stone 2009, 247), can at best attend to an incredibly small minority of those we call evil. As we have seen, a monstrous conception of evil personhood cannot make sense of a broad range of cases and forces us to conclude

that certain individuals, like a self-deceptive rapist or a Nazi war criminal with close

friends, are merely bad people.

Admittedly, there is a strange comfort in the view that all evil people are

irredeemable monsters. It allows us to feel vindicated when we are hostile and

unsympathetic to those we call evil, since we shouldn't be bothered by compassion or

affections for those deemed barely human (Haybron 2002b, 277). It allows us to maintain

the more soothing view that evil is inevitably connected to malicious forethought, instead

of arising from the kind of self-deception that is common to a shared human psychology

(Neiman 2003, 271). And it allows us to demonize and dehumanize those evil persons

while monopolizing humane characteristics for ourselves, the good ones (Formosa 2008a,

235). It is simply easier for us to think that all evil people are monsters in disguise. But

this is rarely, if ever, the reality of the prevalence of evil in the world.

Unsurprisingly, it turns out that every evil person is still a person. As persons, they

are equipped with the same basic psychological and motivational structures that are

familiar to us all. By applying the comparative methodology to borderline cases in order

to come to a plausible theory of evil personhood, and consequently recognizing that evil

persons are conceptually distinct from irredeemable monsters, we are able to approach

conditions of reformation and redemption for evil persons. The monstrous conception of

evil personhood inevitably means that the best we can do to protect ourselves from evil

persons is to kill them off or lock them away for good. This might be the only way for us

to respond to monsters, but it is an unnecessarily harsh response to the majority of

perpetrators of evil in the world. I have argued that what it means to be an evil person is

to be regularly and effectively motivated to perform evil actions. If this is correct, then we

can direct our attention to rehabilitative techniques that are designed for understanding

and changing the motivations of extreme wrongdoers in order to reduce evil in the world.

This offers the more optimistic outlook that we are equipped to consider practical ways in

which evil can be successfully dealt with and prevented. Contrary to the monstrous

conception of evil personhood, we can recognize evil persons as being very much like

ourselves. And, precisely for that reason, we can conceive of them becoming better kinds

of people.

Bibliography

Anderson, Nathaniel & Kent Kiehl (2013) "Functional Neuroimaging and Psychopathy"
        in *Handbook on Psychopathy and Law,* K. Kiehl and W.P. Sinnott-Armstrong (ed.).
        Oxford University Press: 131-149

Barry, Peter Brian (2009) "Moral Saints, Moral Monsters, and the Mirror Thesis"
American Philosophical Quarterly, Vol.46(2): 163-176
-------- (2010) "Extremity of Vice and the Character of Evil" *Journal of Philosophical
        Research* Vol.35: 25-42
-------- (2012) *Evil and Moral Psychology*. Routledge.

Baumeister, Roy & Arlene Stillwell & Sara Wotman (1990) "Victim and Perpetrator
        Accounts of Interpersonal Conflict: Autobiographical Narratives About Anger"
        Journal of Personality and Social Psychology, Vol.59(5): 994-1005

Baumeister, Roy & Kathleen Vohs (2004) "Four Roots of Evil" in *The Social Psychology
        of Good and Evil*, Arthur Miller (ed.). The Guilford Press: New York: 85-101

Calder, Todd (2003) "The Apparent Banality of Evil: The Relationship Between Evil Acts
and Evil Character" *Journal of Social Philosophy*, Vol.34(3): 364-376
-------- (2009) "The Prevalence of Evil" in *Evil, Political Violence, and Forgiveness:
        Essays in Honor of Claudia Card*, Andrea Veltman and Kathryn J. Norlock (ed.)
        New York: Lexington Books: 13–34
-------- (2015) "Evil and Its Opposite" *Journal of Value Inquiry*, Vol.49(1-2): 113-130
-------- (Forthcoming) "Evil Persons"

Caldwell, Michael (2013) "Treatment of Adolescents with Psychopathic Features" in
        Handbook on Psychopathy and Law, K. Kiehl and W.P. Sinnott-Armstrong (ed.).
        Oxford University Press: 201-228

Card, Claudia (2002) *The Atrocity Paradigm: A Theory of Evil*. New York: Oxford
        University Press
-------- (2010) *Confronting Evils: Terrorism, Torture, Genocide*. Cambridge University
        Press.

Chartrand, Tanya & John Bargh (1999) "The chameleon effect: The perception–behavior
        link and social interaction" *Journal of Personality and Social Psychology*,
        Vol.76(6): 893-910.

Cole, Phillip (2006) *The Myth of Evil*. Westport: Praeger.

Criminal Code, RSC 1985, c C-46 s 463, 465. Web.

Formosa, Paul (2008a) "A Conception of Evil" *Journal of Value Inquiry*, Vol. 42(2): 217-239

-------- (2008b) "The Problems with Evil" *Contemporary Political Theory*, Vol. 7: 395-415

-------- (2013) "Evils, Wrongs and Dignity: How to Test a Theory of Evil" *Journal of Value Inquiry*, Vol.47(3): 235-253

Frankfurt, Harry (1971) "Freedom of the Will and the Concept of a Person" *The Journal of Philosophy*, Vol.68(1): 5-20

Garrard, Eve (2002) "Evil as an Explanatory Concept" *The Monist*, Vol.85(2): 320–336

Greene, Joshua (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* New York: The Penguin Press

Haidt, Jonathan (2012) *The Righteous Mind: Why Good People Are Divided by Politics and Religion* New York: Pantheon Books

Haybron, Daniel (2002a) "Consistency of Character and the Character of Evil" in *Earth's Abominations: Philosophical Studies of Evil*, D. Haybron (ed.). New York: Rodopi: 63-78

-------- (2002b) "Moral Monsters and Saints" *The Monist*, Vol.85(2): 260-85

Helzer, Erik & Dave Pizarro (2011) "Dirty Liberals!: Reminders of cleanliness promote conservative political and moral attitudes" *Psychological Science*, Vol.22(4): 517-22.

Inbar, Yoel, Dave Pizarro & Paul Bloom (2012) "Disgusting smells cause decreased liking of gay men" *Emotion* Vol.12: 23-27

Isen, Alice & Paula Levin (1972) "Effect of Feeling Good on Helping: Cookies and Kindness" *Journal of Personality and Social Psychology*, Vol.21(3): 384-388.

Kekes, John (1998) "The Reflexivity of Evil" *Social Philosophy and Policy*. Vol.15(1): 216–32

------- (2005) *The Roots of Evil*. Ithaca and London: Cornell University Press

------- (2009) "Blame versus Forgiveness" *The Monist*, Vol.92(4): 488-506.

Kiehl, Kent (2014) *The Psychopath Whisperer: The Science of Those Without Conscience.* New York: Crown Publishers

Malone, Luke (2014) "You're 16. You're a Pedophile. You Don't Want to Hurt Anyone. What Do You Do Now?" *Matter*. August 10. https://medium.com/matter/youre-16-youre-a-pedophile-you-dont-want-to-hurt-anyone-what-do-you-do-now-e11ce4b88bdb

McGinn, Colin (2003) "The Evil Character" in *Ethics, Evil, and Fiction*. Oxford
      University Press: 61-91

Milgram, Stanley (1974) *Obedience to Authority: An Experimental View*. New York:
       Harper and Row

Neiman, Susan (2004) *Evil in Modern Thought: An Alternative History of Philosophy*.
       Princeton University Press

Russell, Luke (2014) *Evil: A Philosophical Investigation*. Oxford University Press.

Singer, Marcus (2004) "The Concept of Evil" *Philosophy*, Vol.79(308): 185-214

Stone, Michael (2009) *The Anatomy of Evil*. New York: Prometheus Books

Thomas, Laurence (1993) *Vessels of Evil: American Slavery and the Holocaust*.
       Philadelphia: Temple University Press
-------- (2009) "Evil and Forgiveness: The Possibility of Redemption," in *Evil, Political
       Violence, and Forgiveness: Essays in Honor of Claudia Card*, Andrea Veltman
       and Kathryn J. Norlock (ed.) New York: Lexington Books: 118-133

Waldram, James (2009) "It's Just You and Satan, Hanging Out at a Pre-School: Notions of
       Evil and the Rehabilitation of Sexual Offenders" *Anthropology and Humanism*,
       Vol.34(2): 219–234

Zimbardo, Phillip G (2007) *The Lucifer Effect: Understanding How Good People Turn
       Evil*. New York: Random House