Evil and Moral Detachment: Further Reflections on The Mirror Thesis

Abstract: A commonly accepted claim by philosophers investigating the nature of evil is that the evil person is, in some way, the mirror image of the moral saint. In this paper I will defend a new version of this thesis. I will argue that both the moral saint and the morally evil person are characterized by a lack of conflict between moral and non-moral concerns. However, while the saint achieves this unity through a reconciliation of the two, the evil person does so by eliminating moral concerns from her character.

Introduction

"Good as the opposite of evil is, in a sense, equivalent to it."

Simone Weil[1]

Is there a sense in which the evil person is a mirror image of the moral saint? A number of philosophers have been attracted to some version of this view, which I will, following Peter Barry, call The Mirror Thesis. Barry, endorses the following version of the view:

> An agent *a,* is a moral saint just in case *a* possesses *extremely* virtuous states of character […] A person is evil just in case he is a perverse reflection of the moral saint such that he suffers from extremely vicious character traits[2].

Eve Garrard defends a different version of the thesis. She claims that while a virtuous agent is someone for whom the non-moral reasons against a course of action are

---

[1] (1952/2002 p.70).
[2] (2009 pp.171, 173).

1

silenced, for the evil person it is the moral reasons that are silenced.[3] Finally, Hillel Steiner claims that evil acts are the negative counterparts of supererogatory acts as, "evil acts are wrong acts that are pleasurable for their doers, while supererogatory acts are right acts that are painful to perform."[4]

In this paper I want to propose a new version of The Mirror Thesis. I will argue that while moral saints have integrated moral concerns into their identity-conferring commitments, moral monsters have completely removed moral concerns from their character.

We might wonder why anyone has felt the need to spend any time wondering about how to understand The Mirror Thesis. After all, there appears to be a trivial sense in which evil people mirror moral saints: while moral saints are the morally best kind of people, evil people are the morally worst kind of people. This, we might think, is all that needs to be said on the matter. However, as Luke Russell points out, we often cite the evil character of the agent as part of the explanation for the action.[5] Clearly, if we want our account of evil to do this sort of explanatory work we will have to give a more substantive account of evil action than simply stating that evil people are the morally worst sort of people.[6]

Given that this is what is motivating the search for a plausible version of The Mirror Thesis, it also provides one of the criteria on which to judge competing accounts. All

---

[3] (1998).
[4] (2002 p.189).
[5] (2014 Ch.10).
[6] Barry (2009 pp. 163-164) makes a similar point, claiming that we need a thick account of evil, as thin accounts are 'too austere'. Similarly Garrard claim that an account of evil action should give some account of the state of mind of the evil person (2002 pp.321-322). For a discussion of the difference between thick and thin accounts of evil see Russell (2014 Ch.4).

else being equal, one version of The Mirror Thesis will be superior to another if it provides a more informative explanation for evil actions. This, though, is not the only criterion upon which to judge competing versions of The Mirror Thesis. An explanation of evil action is only interesting and useful if it accurately reflects ordinary usage. It is these two criteria upon which I will be evaluating competing forms of The Mirror Thesis.

The discussion will proceed as follows. In §1 I will investigate extant versions of The Mirror Thesis and why these are problematic. I will then, in §2, defend my version of the mirror thesis. I will argue that this version of The Mirror Thesis is more plausible than the accounts considered in §1. I will finish, in §3, by responding to objections that could be raised against this view.

Before I begin it is worth making explicit two assumptions that I will be making in this paper. I will be assuming that the term 'evil' is a useful one and one that it is permissible to apply to actual agents. Not everyone accepts this point,[7] but it does fit with the way in which the term is used in ordinary moral discourse. Moreover, I hope that the discussion in this paper will help to show why the concept is a useful one. Second, I will be assuming that the search for a single account of evil is worthwhile. Again, not everyone accepts this. Russell has recently argued that we should be conceptual pluralists about evil.[8] For the purposes of this paper, however, I wish to investigate what the most plausible single conception of evil is. Whether this single

---

[7] Those that reject this view include Clendinnen (1999), Cole (2006) and Cohen (2011). For a defence of the use of the term see Garrard (2002), Garrard and McNaughton (2012), Morton (2004) Neiman (2003), Russell (2014 Ch1) and Shafer-Landau (2004).

[8] (2014 Ch.5-6).

conception is superior to a pluralist conception is a question that I will be putting off for now.

1. Existing Accounts

Before I give my own account of The Mirror Thesis I will first briefly discuss some extant versions of the view and why they are unsuccessful. This will not only motivate the search for a new version of the view but also highlight a number of features that we should look for in a successful version of the thesis.

A. Affective Mirror Thesis

Both Steiner and Colin McGinn defend affective accounts of The Mirror Thesis. Steiner claims that evil acts are the mirror image of supererogatory acts. The mirroring is to be found, claims Steiner, in the affective properties of those who perform the acts. He outlines his version of The Mirror Thesis in the following, "Evil acts are wrong acts that are pleasurable for their doers, while supererogatory acts are right acts that are painful to perform."[9] McGinn, on the other hand, asks us to imagine two kinds of being, one who takes pleasure in other people's pleasure and the other who takes pleasure in other peoples' pain.[10] These beings, McGinn claims, provide a model for understanding evil.

---

[9] (2002 p.185). In addition to the general problems that face affective versions of The Mirror Thesis Steiner's account is also committed to an implausible view of the supererogatory. There seems no good reason to restrict the supererogatory to acts that are painful to perform, as both Ferry (2013 p.579) and Horgan and Timmons (2010 p.54) point out. Note that Barry's criticism (2009 p.165) of Steiner's view misses the target, as Barry criticizes the view that no one who fails to take pleasure in the performance of an act that goes beyond duty truly supererogates. Steiner's view though, is that no one who fails to experience *pain* from the performance of such an act supererogates.

[10] (1997 p.61).

While these views are not identical, both identify evil with a characteristic affective response.[11] This gives us the following version of The Mirror Thesis:

AFFECTIVE MIRROR THESIS: While a moral saint is someone who takes pleasure in the good (either performing it or merely witnessing it) the evil person is someone who takes pleasure in the bad.

This account of evil does well with regards to the first criterion by which accounts of The Mirror Thesis are to be judged. If we accept that evil people act out of a desire for the bad then this helps to explain why evil acts are performed. When confronted with horrendous acts of evil it is tempting to wonder what could have led someone to act in this way. This account of evil can help to explain how this is possible. What enables someone to act in such a way is that they take pleasure in the bad. This account also captures a view of evil that is fairly widespread in fictional accounts of evil people. Horror movies, for example, frequently depict the villain as someone who enjoys inflicting suffering.

However, this version of The Mirror Thesis appears far less plausible when we consider non-fictional accounts of evil people. As Hannah Arendt made clear in her study of Adolf Eichmann, while we may think that evil people are sadists who take pleasure in the bad, the truth is that evil is often much more banal. Arendt sums up this point in the following: "The trouble with Eichmann was precisely that so many were like him, and that many were neither perverted nor sadistic, that they were, and still are, terribly and terrifyingly normal."[12] Eichmann, according to Arendt, was not someone who took pleasure in other people's pain or in performing bad actions. In

---

[11] See Barry (2009 pp.165-166) for a discussion of the differences between these two positions.
[12] (1963 p.276).

fact, as Arendt points out, when Eichmann witnessed the killing of Jews he found the experience far from pleasurable, describing one scene as "a horrible sight"[13] and refusing to look into a truck full of corpses when instructed to.[14] Of course, a supporter of The Affective Mirror Thesis might take this as evidence that Eichmann was not really evil. While this will surely strike many as a fairly large bullet to bite, the view becomes even more implausible when we consider that Eichmann's feelings of revulsion were far from unique amongst those working in or around concentration camps. As Arendt describes:

> What stuck in the minds of these men who had become murderers was simply the notion of being involved in something historic, grandiose, unique […] Hence the problem was how to overcome not so much their conscience as the animal pity by which all normal men are affected in the presence of physical suffering. The trick used by Himmler […] consisted in turning those instincts around, as it were, in directing them toward the self. So that instead of saying: What horrible things I did to people!, the murderers would be able to say: What horrible things I had to watch in the pursuance of my duties, how heavily the task weighed upon my shoulders.[15]

What this passage from Arendt makes clear is that many of those working in the camps found their task extremely unpleasant and had to devise ways of managing this unpleasantness. If The Affective View is right then this would show that these people are not evil. This, though, is an unacceptable conclusion. If anyone is to count as evil

---

[13] (1963 p.89).
[14] (1963 p.88).
[15] (1963 pp.105-6)

then Eichmann, Himmler and others who orchestrated The Final Solution should do so.[16]

In addition, The Affective View also faces an objection to its account of moral sainthood. While it may well be true that many moral saints generally do find performing morally good acts pleasurable this is not an essential property of sainthood. To see why consider Susan Wolf's distinction between the loving saint and the rational saint. While the loving saint performs morally good acts because it makes her happy to do so and so acts in line with her self-interest, the rational saint sacrifices her own interests for the interests of others.[17] The rational saint then is not someone who takes pleasure in the good. If we accept that the concept of the rational saint is a coherent one then it looks like the taking pleasure in the good is not a necessary condition for moral sainthood and is at best a contingent feature of actual moral saints.

To sum up, The Affective Mirror Thesis provides an account of evil that appears to offer explanatory potential and fits with many fictional accounts of evil personhood. Unfortunately, it fails to fit with high profile non-fictional accounts of evil personhood and, given the implausibility of withholding the term 'evil' from these cases, it must be rejected for this reason.[18] In addition, the view fails to account for the possibility of rational saints.

---

[16] We might reasonably worry that Eichmann's psychology was not as Arendt describes (See Cesarani 2007). Nevertheless, as Russell points out, we can still draw philosophical conclusions about the nature of evil using the picture Arendt paints of Eichmann, even if this picture is inaccurate, particularly given that many other cases of evil people do seem to possess something like this psychological profile (2014 p. 73).

[17] (1982 p.420).

[18] For more on the difference between fictional and non-fictional portrayals of evil see Baumeister (1997 Ch.3).

B. *Silencing Mirror Thesis*

An alternative account of The Mirror Thesis is proposed by Garrard, who argues that we should identify evil actions not by affective properties but by the psychological silencing of certain reasons for action. According to Garrard, when a virtuous person faces a choice between an act favoured by moral reasons and a different act favoured by non-moral reasons the non-moral reasons are psychologically silenced.[19] While a continent person would feel the force of the non-moral reasons in such a case and overcome the temptation to do otherwise, the truly virtuous person would have no temptation to overcome, as the non-moral reasons would be silenced. The evil person, on the other hand, is someone for whom the moral reasons that count against performing a moral act are psychologically silenced.[20] This gives us the following view:

THE SILENCING MIRROR THESIS: While a moral saint is someone who psychologically silences non-moral reasons when they conflict with moral reasons, an evil person is someone who psychologically silences moral reasons when they conflict with non-moral reasons.

This view is capable of explaining evil actions. While we may find it hard to comprehend how someone like Eichmann could act as he did, The Silencing Mirror Thesis provides an explanation.[21] What enabled Eichmann to act in this way is that he had psychologically silenced the moral reasons counting against his actions. He does

---

[19] (1997 p.50).
[20] (1997 pp.53-55), (2002 p.330)
[21] As Garrard points out (2002 pp.332-334).

not appreciate the force of the reasons counting against his actions and this is what allows him to act in ways that we would find it hard to imagine ever acting ourselves.

The Silencing Mirror Thesis also does a better job than The Affective Mirror Thesis in handling cases of banal evil. According to The Silencing View, those who perform acts of evil are blind to the moral reasons that count against their actions. It is no surprise, then, that their actions may appear banal. They are simply unaware of the strong reasons counting against their actions and this is why they are able to perform horrendous acts in the same way that they would perform any other humdrum task.

However, like The Affective View, The Silencing Mirror Thesis fails to pick out an essential property of moral sainthood. While it may well be true that many moral saints are generally psychologically blind to non-moral reasons when they clash with moral reasons, again this is not an essential property of sainthood. A rational saint may well be someone for whom self-interested reasons are not psychologically silenced. Again it looks like this view picks out traits that are at best contingent features of actual moral saints.

Moreover, while The Silencing Mirror Thesis does a better job of handling real life cases of evil than The Affective View, it ultimately fails to provide a plausible account of evil personhood. The first problem with this view is that it rules out the possibility of evil actions that are dependent upon the moral reasons that count against acting in a particular way. For example, John Milton's Satan in *Paradise Lost* desires to do what is bad, saying, "Evil be thou my good".[22] Clearly, though, in order to be successful in this Satan would need to be able to see the moral reasons present in any given situation in order then to perform the acts they count against. Saint Augustine

---

[22] (1674 Book IV Line 110).

describes a similar motivation of doing wrong for the sake of doing wrong in his *Confessions*:

> Let my heart tell you what prompted me to do wrong for no purpose, and why it was only my own love of mischief that made me do it. The evil in me was foul, but I loved it, not for the things that I committed wrong, but the wrong itself.[23]

Another example can be found in the actions of Dr. Jemand von Niemand in William Styron's novel *Sophie's Choice.* Von Niemand forces Sophie to choose which of her two children to save from the gas chamber. If she fails to make a choice then he will take both. Speculating on von Niemand's motivations the narrator, Stringo, says the following:

> It had to do with the absence of sin, and his own realization that the absence of sin and the absence of God were inseperably intertwined. No sin! He had suffered boredom and anxiety, and even revulsion, but no sense of sin from the bestial crimes he had been party to […] Was it not supremely simple, then, to restore his belief in God, and at the same time to affirm his human capacity for evil, by committing the most intolerable sin that he was able to imagine?[24]

Stringo's interpretation of the motivation behind this action then is that von Niemand wanted to perform the worst act he could imagine in an attempt to feel a sense of sin, which would give him reason to believe in God once more. In both of these examples the actions are dependent upon the agent being able to see the moral reasons that count against these actions. Nevertheless, both acts appear to be paradigmatic

---

[23] (1961 p.47)
[24] (1979 p.598).

examples of evil action. The problem for The Silencing View is that on this view the fact that the agents can see these moral reasons prevents these acts from being classed as evil.[25]

C. Virtue/ Vice Mirror Thesis

The final account of to be considered is proposed by Barry:

VIRTUE/ VICE MIRROR THESIS: A moral saint is someone who possesses extremely virtuous character traits, an evil person is someone who possesses extremely vicious character traits.[26]

At least on the face of it, this account possesses a significant advantage over the previous two in that it appears to capture a conceptual claim about saints and heroes rather than a contingent one. As we saw, the previous accounts both suffered from giving an account of sainthood that picked out properties which are at best contingent features of moral saints. As a result, they wrongly ruled out the conceptual possibility of rational saints. Barry's claim can avoid this objection, possessing extremely virtuous character traits is plausibly a platitude about moral sainthood. Moreover, Barry's account can accommodate the possibility of the rational saint, as it seems plausible to think that a rational saint possesses extremely virtuous character traits.

---

[25] Barry (2009 p. 167) suggests that The Silencing View can avoid objections of this kind if we distinguish between two ways in which reasons can be silenced. Reasons may be silenced by a failure to appreciate their existence (a failure of receptivity) or they may be silenced by a failure to be moved by their existence (a failure of reactivity). There are two problems with this response. First, as Barry notes, this new version of the view is problematic, as some evil people seem to react to some moral reasons and some saints react to nonmoral reasons Second, it fails to handle the problems raised by the examples. Both Satan and von Niemand *are receptive* to moral reasons. They are both moved by the existence of the moral reasons to act in certain ways. This response then cannot save the silencing view.
[26] Barry (2009 p.173).

However, this advantage comes at the cost of having far less explanatory power than the previous two accounts. While the previous two accounts offered an informative explanation as to why both saints and evil people perform acts most ordinary people would not, The Virtue/ Vice View does not. In the introduction we saw that the reason to look beyond a trivial version of The Mirror Thesis that holds that moral saints are the morally best kind of people and evil people the morally worst kind, is of no explanatory use. The problem with Barry's account is that it is scarcely any better at explaining evil conduct than this trivial account. If someone were to ask how an evil person was able to perform an act that most people would find too morally repelling to perform they are unlikely to be satisfied with the explanation that the evil person possesses extremely vicious character traits.

To make matters worse for The Virtue/ Vice View it appears to be incompatible with Arendt's claims about cases of banal evil. As we have already seen, Arendt's view of Eichmann was that he was not a monster but rather 'terrifyingly normal'. What enabled Eichmann to act as he did was not, according to Arendt, some extreme vice but rather lack of imagination. In Arendt's words: "It was sheer thoughtlessness – something by no means identical with stupidity – that predisposed him to become one of the greatest criminals of that period."[27] This, thoughtlessness, though does not seem like an *extreme* vice, for two reasons. The reason for this is that this trait only produced the terrible results that it did as a result of the specific time and place Eichmann found himself in. In almost any other time or place Eichmann's lack of imagination would not have had these results. Indeed, the unquestioning obeying of order might even have led Eichmann to perform a great many morally worthwhile

---

[27] (1963 pp.287-288).

acts given the right external conditions. Yet it is surely a necessary condition for an extreme vice that it *disposes* the agent to perform bad acts, not only that it *happens* to lead the agent to perform bad acts in very specific circumstances. To be considered a vice then this trait would have to be one that would lead the agent to perform bad acts across a sufficiently large range of possible worlds. Eichmann's thoughtlessness does not appear to be such a trait.

Eichmann's thoughtlessness then, is a trait that neither disposes the agent to perform morally bad acts across a sufficiently large range of possible worlds nor does it appear to be a particularly uncommon trait. As a result, there seems little reason to think that this trait is an extreme vice. Nevertheless, this lack of an extreme vice does not seem to prevent Eichmann from being classed as evil.

2. Moral Integration/Detachment Mirror Thesis

In this section I will defend a new version of The Mirror Thesis, which, I will argue, avoids the objections facing the views we looked at in the previous section. Before I begin it is worth briefly reminding ourselves of the problems those views faced. The problem for The Affective Mirror Thesis was that it was incompatible with cases of banal evil. The Silencing Mirror Thesis, on the other hand, appeared to be incompatible with cases of evil where the agent performs bad acts *because* they are bad. Finally, the Virtue/ Vice Mirror Thesis was incompatible with the possibility that an evil person might not be disposed to perform evil acts across a wide range of possible worlds. Ideally, then, we would find a version of the view that is both plausible in itself and that is not vulnerable to any of these objections. In the rest of this section I will present a view that achieves both goals.

Central to our secular understanding of a moral saint is someone who makes a commitment to morally worthwhile projects a fundamental part of his or her identity. As Susan Wolf puts the point, "A necessary condition for moral sainthood would be that one's life be dominated by a commitment to improving the welfare of others or of society as a whole."[28] As it stands, Wolf's claim seems to present an overly self-conscious picture of a moral saint.[29] As Vanessa Carbonell has argued, someone who is committed to morally good pursuits can be described as a moral saint even if he does not present those commitments as moral commitments to himself.[30] Carbonell supports this claim by focusing on the example of Paul Farmer, an American anthropologist and physician dedicated to providing health care to rural and under privileged areas of the developed world. As Carbonell points out, "Although he is almost maniacally driven by morally good pursuits, he does not describe the pursuits to himself as such. He simply wants to help the poor and the sick."[31] The important point then is that moral saints are committed to morally worthwhile projects, not that they necessarily represent these projects to themselves in this way.

In fact this also fits with the most plausible understanding of the most important sense of 'moral concern'. Nomy Araply argues that we should understand moral concern as, "concern for what is in fact morally relevant and not as concern for what the agent takes to be morally relevant."[32] In other words, what is important for how we evaluate an agent's moral behaviour is not, primarily at least, the agent's moral judgements about what acts are right or wrong but whether the agent is moved to perform the

---

[28] (1982 p.421).
[29] Thanks to an anonymous referee for helpful discussion here.
[30] (2009 p.377).
[31] (2009 p.377).
[32] (2003 p.84).

right actions by the appropriate kinds of reasons. In order to be morally praiseworthy an agent must be acting for the relevant moral reasons, where this is read *de re* rather that *de dicto*.[33] What this shows then, is that The Moral Integration Thesis should be understood in the *de re* sense rather than the *de dicto* sense. What is important for both these views is not whether the agent has integrated a concern for what *she takes* to be moral behaviour into her character but whether or not she has integrated a concern for *what is* moral behaviour into her character.

A moral saint then, is someone who has made a pursuit of morally worthwhile goals (so understood) a fundamental part of who they are. We can formalize this view in the following way:

MORAL INTEGRATION THESIS: A moral saint is someone who has made morally worthwhile concerns a fundamental part of her identity.

The claim that a saint is someone whose life is dedicated to moral projects seems to be a platitude about sainthood. It seems to make little sense to say that someone is a saint but that they are not particularly committed to any morally worthwhile projects. Indeed it is this dedication to moral projects that Wolf argues makes the life of a moral saint undesirable.[34]

This view of moral sainthood avoids the problem that both The Affective View and The Silencing View faced in accommodating the possibility of rational saints. The claim that moral goals must be a fundamental part of one's identity is fully

---

[33] (2003 p.77). Which is not to say that the evaluations of the agent's moral judgements are completely irrelevant to the moral assessment of the agent. Clearly it is preferable for an agent to only act for the right reasons but also to judge that those reasons are the right ones. However, this does not detract from the fact that judgements of moral worth are primarily concerned with the former.

[34] (1982). For a response to Wolf's argument see Carbonnell (2009).

compatible with the possibility of rational saints. The fact that one dedicates one's life to moral goals out of a dedication to moral principles rather than from moral sentiment does not make this dedication any less of a fundamental part of one's identity.

This claim about moral sainthood also has explanatory power. The reason why saints are able to act in ways that ordinary people are not is that they have managed to make moral concerns a key part of who they are. It should be unsurprising that such a person should find it easier to perform morally worthy acts. In fact this explanation appears to be supported by empirical psychological studies of moral saints. A common finding amongst those investigating the psychology of moral exemplars is that they make moral interests a central part of their identity. In a recently published study, Jerermy Frimer et al. found that exemplars were significantly more likely than the comparison group to have integrated their personal ambitions with their moral convictions. After ruling out various alternative explanations, the researchers concluded that, "These results are consistent with the claim that moral exemplars have achieved enlightened self-interest, whereby they best advance their own interests by advancing the interests of others."[35] Similarly, in their study of moral exemplars, Anne Colby and William Damon claimed the following:

> The exemplars have done so without devaluing their own personal goals. Nor do they disregard their own fulfilment or self-development, nor, broadly construed, their own self-interests. They do not seek martyrdom. *Rather than denying the self, they define it with a moral centre.*[36]

---

[35] (2011 p.160). For further psychological evidence for this claim see Frimer et al. (2009) and Frimer et al. (2012). See also Colby and Damon (1992).

[36] Colby and Damon (1992 p.300).

The point Colby and Damon are making is that we should not think of exemplars as acting against their own goals. Rather, exemplars make moral goals a key part of their identity. In their words, "They see their moral goals as a means of attaining their personal goals, and vice versa. This can only be possible when moral goals and personal goals are closely in synchrony, perhaps even identical."[37] According to Colby and Damon, the identification of self-interest with moral goals is not something that exemplars have innately. Rather, it is a gradual process by which the moral goals people strive towards become slowly harder to distinguish from the exemplars' self-interested goals.[38]

There are a number of worries we might have about this integration claim. We might worry that this claim requires that the moral saint would necessarily recognise a description of herself as someone whose self-interest and moral concerns coincide. Similarly, we might worry that this view is committed to saying that the saint who would be motivated by the thought that she would be advancing her own interests by acting in the interests of others. This seems worrying, as the saint may never stop to think what acts would be best from the self-interested point of view and if she did stop to think it is likely to be the concern for the well-being of others that would be her primary motivation. Finally, we might worry that in some situations, a concern for one's own well-being could not enter into the motivational set of the virtuous person,

---

[37] (1992 p.300).

[38] See Colby and Damon (1992 Chapter 7) and Frimer and Walker (2009) for explanations of the development of moral commitment in exemplars. Elsewhere, (Archer and Ridge (2015)), I argue that this form of integration can plausibly be viewed as a form of moral depth. As I argue elsewhere, this form of commitment may explain why moral exemplars often judge supererogatory acts to be obligatory (See Archer and Ridge (2015)) or to be a moral necessity (See Archer (2015)).

as there will be no option available in which the saint could achieve any level of well-being.

We can respond to these worries by clarifying the exact nature of this integration. I am not claiming that the saint is someone who would necessarily recognise that her morally worthwhile concerns have been integrated with her self-interest, nor am I claiming that the saint is being motivated by a desire to advance her own self-interest. Rather, the claim I am making about saints is that the saint's morally worthwhile concerns do coincide with her self-interested concerns, even if this is not what is motivating her to act and even if she is unaware of this integration.

We might worry that if the saint is not being motivated by her self-interested concerns, then this integration is not playing a useful role in her moral life. After all, if these self-interested concerns are not what is primarily motivating the saint then we might wonder what the value of this integration is. The first answer is that even if this is not what is *primarily* motivating the saint, self-interested considerations may play an additional motivational role. The second, more important, response is that the self-interested concerns do not need to be playing a motivational role for this integration to be playing a useful role in the life of the saint. This important role this integration plays is in removing the conflict between the motivation to pursue morally worthwhile concerns and the motivation to pursue self-interested concerns. This is important because it is a common feature of the lives of many people that what they morally ought to do conflicts with their view of what would be best from a self-interested point of view. As H. A. Prichard puts the point:

> Any one who, stimulated by education, has come to feel the force of the
>
> various obligations in life, at some time or other comes to feel the irksomeness
>
> of carrying them out, and to recognise the sacrifice of interest involved.[39]

The useful role of the integration of morally worthwhile concerns then, comes from
the elimination of this conflict and the temptation to act immorally.

If the moral saint is someone who has integrated moral concerns into her identity then
what would the perverse mirror image of this be? Presumably, it would be someone
who has completely *eradicated* moral concerns from her identity. Just as the moral
saint is someone who makes morality a core part of their character, the moral monster
eradicates morality from her character. The following description of Rudolf Höss, the
commandant of the Auschwitz concentration camp from 1940 to 1943, by William
Styron captures this thought well: "Thus he became a mere servomechanism in which
a moral vacuum had been so successfully sucked clean of every molecule of real
qualm or scruple."[40]

Let's formulate this claim in the following way:

MORAL DETACHMENT THESIS: An evil person is someone who has detached
themselves from moral concerns.

This thesis does a good job of uniting the various different kinds of character we tend
to think of as evil. First, unlike The Affective Mirror Thesis, it can handle cases
where the evil person does not experience pleasure in the performance of evil acts, as

---

[39] (1912 p.21).
[40] (1979 p.179).

it does not require that evil people have such a response. Cases of banal evil are cases where seemingly normal people act in evil ways by failing to consider the effects of their action. This failure can be seen as one that displays a moral detachment. Second, unlike The Silencing Mirror Thesis, it can handle 'pure' cases of evil such as Milton's Satan and Dr. von Niemand who desire to do the bad because it is bad. These are people who have cut themselves off from a concern for morality and in addition have developed a desire to perform immoral acts. Third, people who are pathologically evil are those whose pathology prevents the development of a moral concern. In addition, unlike The Virtue/ Vice Mirror Thesis, this view does not rest upon the assumption that the evil person is one who is disposed to perform evil actions across a range of possible worlds. All that is required on this view is that the evil person has cut himself off from moral concerns in the actual world.

This view is also one that has explanatory power. The reason that evil people are able to perform acts that most ordinary people would judge themselves incapable of performing is that evil people have become detached from moral concerns while ordinary people have not.

This explanation is also one that is supported by evidence from empirical psychology. Just as moral saints go through a process of development during which their self-interested goals are increasingly hard to distinguish from moral concerns, so too the evil person tends to have gone through a similar process. Roy Baumeister describes the process of detaching oneself from moral concerns in his book *Evil: Inside Human Violence and Cruelty*.[41] Reflecting on a passage from David Stannard about an

---

[41] (1997).

American cavalry major describing shooting Native American children as target

practice in Colarado in 1864, Roy Baumeister says the following:

> The men were not talking about how awful it was to kill children, nor were
>
> they acting with demented glee or sadistic pleasure at the killings. *It is an*
>
> *attitude devoid of moral reflection or profound emotion*: a matter of fact
>
> attitude. […] This is an attitude more common of experienced killers than of
>
> novices. Typically, first killing is psychologically difficult and upsetting, even
>
> traumatic. Somehow, though, people can get used to killing, so that it
>
> produces less and less reaction.[42]

Andre Gide gives a description of this process in his novel *The Immoralist:*

> I soon came to understand that the things that are reputed worst (lying, to
>
> mention only one) are only difficult to do as long as one has never done them;
>
> but that they become – and very quickly too – easy, pleasant, and agreeable to
>
> do over again, and soon even natural. So then, as is always the case when one
>
> overcomes an initial disgust, I ended by taking pleasure in my dissimulation
>
> itself, by protracting it, as if afforded opportunity for the play of my
>
> undiscovered faculties. And every day my life grew richer and fuller, as I
>
> advanced towards a riper, more delicious happiness.[43]

How does someone manage to overcome the psychological difficulties involved in

killing? Of course some people may never develop this moral concern[44] but for most

---

[42] (1997 p.9-10).

[43] (1960/ 1902 p.59).

[44] We might think that psychopaths are people who have never developed moral concern. It is important though to distinguish this from the claim that psychopathology can be wholly attributed to biological factors. See R. D. Hare (1993

people performing acts of this sort will involve a great deal of psychological trauma. According to Baumeister there are a number of ways in which people are able to avoid the negative psychological effects that would normally accompany acting immorally. One way of doing so is to focus on the low level practical details of one's actions rather thinking about the broader implications of one's actions.[45] Gitta Sereny's study of Franz Stangl, the commandant of the Nazi concentration camp at Trblinka, highlights this strategy. Sereny asks Stangl how he managed to get used to his work. Stangl replies by saying "There were hundreds of ways to take one's mind off it; I used them all […] of course, thoughts came. But I forced them away. I made myself concentrate on work, work and work again."[46] There is some empirical evidence to suggest that this is what criminals typically do when committing crimes.[47] Another way of avoiding one's typical moral reactions is to see one's victims as inhuman.[48] Hitler's minister of armaments, Albert Speer, points to the effectiveness of this method when he said of the Jews, "If I had continued to see them as human beings, I would not have remained a Nazi. I did not hate them. I was indifferent to them."[49] Importantly for our purposes, these coping mechanisms are all ways of detaching oneself from one's moral concerns.

3. Objections and Replies

Ch. 10) for a defence of the view that psychopathology results from a (poorly understood) combination of social and biological factors.

[45] (1997 p.269, 336-340).

[46] (1995 p.200).

[47] See Wegner and Vallacher (1986). More worryingly, Adam and Balfour (2009) claim that it is in the nature of bureaucracies to encourage a focus on these low level concerns, making possible widespread administrative evil.

[48] As pointed out by Bandura (2002 pp.108 -109). Bandura discusses a number of other ways in which people morally detach themselves that I do not have the space to explore here.

[49] (1970 p.315).

There are, though, a number of objections that might be raised against this account of evil. First, it might be objected that many evil people have not fully cut themselves off from moral concerns. Eichmann, for example, helped his half-Jewish relative emigrate to Switzerland.[50] Similarly, many saints have not fully integrated moral concerns into their character. Martin Luther King, for example, has been accused of being a philanderer. Does this mean that on the account I have offered we cannot class Eichmann as evil or King as saintly?

Fortunately there is no need for my account to be prevented from classifying Eichmann and King in these ways. Evil and sainthood are clearly concepts that come in degrees. Eichmann was evil but not as evil as Hitler or Milton's Satan. It is no problem then that some evil people are not fully cut off from moral concerns nor that some good people have not fully integrated moral concerns into their character. This just shows that they are less evil or saintly than they could be.[51]

A more worrying problem is that some of the clearest cases of evil people are also people who believed that they were doing the morally right thing. Take, for example, The Junta that seized control of Argentina in 1976. They subjected those they suspected of being political opponents to a brutal regime of torture and murdered an estimated eleven to fifteen thousand people.[52] However, the reasons they gave for doing so were moral ones. As one member of The Junta, General Videla put the point, "The armed forces have assumed the direction of the state in fulfillment of an obligation from which they cannot back away […] This decision is aimed at ending

---

[50] Arendt (1963 p.70).

[51] Some might object that my account is *committed* to classing King as a saint and object to this result. However, my account is not committed to this classification. Rather, on my account the question of whether or not King should count as a saint is a question of the extent of the integration of his moral concerns.

[52] Kekes (2005 p.86).

misrule, corruption and the scourge of subversion."[53] In the words of Stanley Loomis, "There is no crime, no murder, no massacre that cannot be justified provided it be committed in the name of an Ideal."[54] The problem this raises for The Moral Detachment Thesis is that it seems that these cases of idealistic evil cannot be described as people who have detached themselves from moral concerns. Rather, it is their moral *commitments* that are the problem.

This objection raises an important worry for the view of evil that I have proposed. Certainly, any view of evil that fails to class these actions as evil will not be a plausible one. However, the right response to this objection is to clarify The Moral Detachment Thesis, rather than reject it. The reason the objection has force is that there is a crucial ambiguity in the claim that the evil person is one who has 'detached herself from moral concerns'. One way of reading this is that the evil person has detached herself from *her own judgements* about morality. Clearly, if we interpret the claim in this way then it is vulnerable to the above objection. However, another way of reading this claim is that the evil person is one who has detached herself from concerns that *are in fact* moral. If we interpret the claim in this way then the previous objection has no force. Despite the fact that the members of The Junta may have remained attached to their own judgements of right and wrong they had plausibly cut themselves of from any concerns that can accurately be described as moral.

It is worth noting that this interpretation is also the most plausible way of interpreting The Moral Integration Thesis. As I made clear in my initial presentation of this thesis, what is important for moral sainthood is that the saint is committed to morally worthwhile concerns read *de re* not *de dicto*. The same is true for the morally evil

---

[53] This quote is from Rosenberg (1991 p.123).
[54] (1964 p.403).

person. Someone is morally evil if they have detached themselves from moral concerns, read *de re*. Such a person might, nevertheless be attached to concerns she herself takes to be moral.

We might worry that once we have clarified the view in this way that it becomes somewhat uninformative. On this account, to describe someone as evil is to say that they have detached themselves from a concern for what is morally important. We might wonder how such a description could ever help us understand what led someone to act a certain way.

We can see why this thought is misguided by applying this account of evil to an example. Consider Herman Melville's short story Billy Budd. Billy Budd is a foretopman of angelic character working on H.M.S. Indomitable. He is falsely accused of trying to start a mutiny by the evil master-at-arms Claggart. Due to a speech impediment Budd fails to respond to the allegation. In frustration he strikes Claggart, who falls, hits his head and dies. This presents a moral dilemma for Captain Vere, Captain of the Indomitable, which Melville outlines in the following:

> In the legal view, the apparent victim of the tragedy was he who had sought to victimise a man blameless; and the indisputable deed of the latter, navally regarded, constituted the most heinous of military crimes. Yet more. The essential right and wrong in the matter, the clearer that might be, so much the worse for the responsibility of a loyal sea commander, inasmuch as he was authorised to determine the matter on that primitive legal basis.[55]

---

[55] (1959 pp. 253-4).

According to naval law, Budd is guilty of murder and must be hanged. Nevertheless, Captain Vere is strongly moved by compassion to let Budd escape this punishment. This is not simply a conflict between military law and moral law but one internal to morality.[56] Captain Vere is, after all, a loyal sea commander and as such is morally committed to naval law. Vere decides that his conscience must yield to naval law and hangs Budd.

To see why my account of evil can play an explanatory role we should compare an alternative version of the case, featuring a different captain, Captain Coldhearted, who is completely detached from moral concerns. Unlike Captain Vere, Captain Coldhearted is glad of the opportunity to send Billy Budd to death and feels no remorse for doing so. We are now in a position to see why The Moral Detachment Thesis is capable of both playing an explanatory role and distinguishing evil acts from those that are simply wrong. To say that this captain performed his act because he was evil is to say that his actions can be at least partially be explained from his complete detachment from the moral considerations that counted against doing so. This would not be an appropriate explanation for Captain Vere's actions, as he acted *despite* having legitimate moral concerns that pushed him to act differently. To cite evil in the explanation for Captain Coldhearted's act is to say something that is capable of helping us to understand why he acted as he did and distinguishing this act from one that is merely wrong.

Concluding Remarks

In this paper I have defended a new version of The Mirror Thesis, the view that the morally evil person is the perverse mirror image of the moral saint. I started, in §1, by

---

[56] This point is made by Winch (1972 p.158).

looking examining the existing accounts of The Mirror Thesis and explaining their shortcomings. I then, in §2, outlined my own account. According to this account a moral saint is someone who has made moral concerns a fundamental part of their identity. The morally evil person, on the other hand, is someone who has detached themselves from moral concerns. I pointed to a number of advantages for this account over its rivals. First, it does a good job of explaining what it is that the various different kinds of character we tend to think of as evil have in common. Second, it is an account that it capable of playing an explanatory role. Finally, it is supported by evidence from empirical psychology. I finished, in §3, by considering the objection that this account prevents those who believe that they are doing the morally right thing from counting as evil. This would be problematic as some of the clearest cases of evil people are also people who believed that they were doing the morally right thing. However, I argued that both The Moral Attachment Thesis and The Moral Detachment Thesis are concerned with whether or not an agent has integrated a concern for *what is* moral behaviour into her character rather than a claim about whether or not she has integrated a concern for what *she takes* to be moral behaviour into her character.[57]

Bibliography

Adams, G.B. And Balfour, D.L. *Unmasking Administrative Evil* (London: Sage Publications).

Archer, A. (2015) "Saints, Heroes and Moral Necessity," *Royal Institute of Philosophy Supplementary Volume*. Vol. 77 pp.105 – 124.

Archer, A. and Ridge, M. (2015) "The Heroism Paradox: Another Paradox of Supererogation," (Co-Authored with Mike Ridge) *Philosophical Studies* (Vol. 172 No.6 pp.1575 – 1592).

Arendt, Hannah (1963) *Eichmann In Jerusalem: A Report on The Banality of Evil* (London: Penguin).

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry Into Moral Agency.* (Oxford: Oxford University Press).

Augustine, Saint (1961/ 398) *Confessions* Translated by R. S. Pine-Coffin (London: Penguin).

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, *31*(2), 101-119.

Barry, Peter Brian (2009). "Moral Saints, Moral Monsters, and the Mirror Thesis." *American Philosophical Quarterly* 46 (2):163 - 176.

Barry, Peter Brian (2011). "In Defense Of The Mirror Thesis". *Philosophical Studies* 155 (2):199-205.

Baumeister, R. F. (1997) *Evil: Inside Human Violence and Cruelty* (New York: W.H. Freeman and Company).

Carbonell, V. (2009). "What Moral Saints Look Like." *Canadian Journal of Philosophy* 39 (3):pp. 371-398.

Cesarani, David (2007), *Becoming Eichmann: Rethinking the Life, Crimes, and Trial of a "Desk Murderer."* (London: De Capo Press).

Clendinnen, I. (1999) *Reading the Holocaust* (Cambridge: Cambridge University Press).

Colby, A. and Damon, W. (1992). *Some Do Care*. Free Press.

Cole, P. (2006). *The Myth of Evil: Demonizing the Enemy*. (Edinburgh: Edinburgh University Press)

Ferry, M. (2013) "Does Morality Demand Our Very Best? Moral Prescriptions and the Line of Duty." *Philosophical Studies* 165 (2):573-589.

Frimer, J. A., & Walker, L. J. (2009). "Reconciling The Self and morality: An Empirical Model of Moral Centrality Development." *Developmental Psychology*, 45, 1669–1681.

Frimer, J. A., Walker, L. J., Dunlop, W. L., Lee, B. H., & Riches, A. (2011). "The Integration of Agency and Communion In Moral Personality: Evidence of Enlightened Self-interest*." Journal of Personality and Social Psychology*, 101, 149–163.

Frimer, J. A., Walker, L. J., Lee, B. H., Riches, A., & Dunlop, W. L. (2012). "Hierarchical Integration of Agency and Communion: A Study of Influential Moral Figures." *Journal of Personality*, 80(4), 1117-1145.

Garrard, Eve (1998). "The Nature of Evil." *Philosophical Explorations* 1 (1):43 – 60.

Garrard, Eve (2002). "Evil as an Explanatory Concept." *The Monist* 85 (2):320-336.

Garrard, Eve & McNaughton, David (2012). "Speak No Evil?" *Midwest Studies in Philosophy* 36 (1):1-17.

Gide, André (1960/ 1902) *The Immoralist* Translated by Dorothy Bussy (Harmondsworth: Penguin).

Hare, R. D. (1993) *Without Conscience: The Disturbing World of The Psychopaths Among Us* (London: Guilford Press).

Horgan, T. and Timmons, M. (2010). "Untying a Knot from the Inside Out: Reflections on the 'Paradox' of Supererogation." *Social Philosophy and Policy*, 27: 29–63.

Kekes, John (2005). The Roots of Evil. Cornell University Press.

Loomiss, Stanley (1964) *Paris in the terror* (Philedelphia: Lippincott).

McGinn, C. (1997) *Ethics, Evil and Fiction* (New York: Oxford University Press).

Melville, H. "Billy Budd, Foretopman." *Four Short Novels* (New York: Bantam Books, 1959).

Milton (1674) *Paradise Lost* (London: Simmons).

Morton, Adam (2004) *On Evil* (London: Routledge).

Neiman, Susan (2003) *Evil in Modern Thought: An Alternative History of Philosophy* (Melbourne:Scribe).

Oliner, S. P., & Oliner, P. M. (1988). *The Altruistic Personality: Rescuers of Jews In Nazi Germany* (New York: Free Press).

Prichard, Harold Arthur (1912) "Does moral philosophy rest on a mistake?" *Mind* Vol. 21: 21-37

Rosenberg, T. (1991) *Children of Cain* (New York: Morrow).

Russell, L. (2014). *Evil: A Philosophical Investigation* (Oxford: Oxford University Press).

Sereny, Gita (1995) *Into that Darkness* (London: Pimlico).

Shafer-Landau, Russ (2004) *Whatever Happened to Good and Evil* (New York: Oxford University Press).

Speer, A. (1970) *Inside The Third Reich* (London: Weidenfield and Nicolson).

Steiner, Hillel (2002) "Calibrating Evil." *Monist* Vol. 85

Styron, W. (1979/2004) *Sophie's Choice* (London: Vintage).

Wegner, D. M. and Vallacher, R. R. (1986) "Action Identification". R. M. Sorrectino and E.T. Higgins (Eds.) *Handbook of Cognition and Motivation* (New York: Guilford Press pp.550-582).

Weil, S. (1952/ 2002) *Gravity and Grace* Translated by Emma Crawford (London: Routledge).

Winch, P. (1972) 'The Universalizability of Moral Judgments.' In P. Winch *Ethics and Action* (London: Routledge).

Wolf, Susan. (1982) "Moral Saints." *The Journal of Philosophy* 79.8 : 419-439.