Now there can exist no irritating juxtaposition of dissimilar personalities comparable to that which is possible aboard a great warship fully manned and at sea. There, every day among all ranks, almost every man comes into more or less of contact with almost every other man. Wholly there to avoid even the sight of an aggravating object one must needs give it Jonah's toss or jump overboard himself. Imagine how all this might eventually operate on some peculiar human creature the direct reverse of a saint!

I. INTRODUCTION

In the above passage, the narrator of Billy Budd, Sailor describes the wicked Claggart as being “the direct reverse of a saint.” A number of philosophers have been impressed with the thought that moral saints and moral monsters—or, evil people, to put it less sensationally—“mirror” one another, in a sense to be explained. Call this the mirror thesis. The project of this paper is to cash out the metaphorical suggestion that moral saints and evil persons mirror one another and to articulate the most plausible literal version of the mirror thesis. To anticipate, the most plausible version of the mirror thesis implies that evil persons mirror moral saints insofar as the characters of each are marked by similar aretaic properties: suffering from extremely vicious character traits—in a sense to be explained—suffices for being evil, whereas possessing extremely virtuous character traits similarly suffices for moral sainthood.2

II. THE MIRROR THESIS

Note initially the analytic truth that the evil person just is the morally worst sort of person: if someone could be a morally worse sort of person, then he is simply not evil, whatever else is wrong with him. Parity of reasoning suggests that a moral saint is the morally best sort of person, such that if someone could be a morally better sort of person, then she is not a moral saint, however praiseworthy she otherwise is.3 Thus, there is an obvious sense in which evil persons mirror moral saints: each is an instance of the highest degree of something. But to note merely that evil persons are of the morally worst sort is to employ a “thin” conception of evil personhood.4 The problem with thin conceptions of evil is that they are too austere. If “evil person” constitutes a genuine moral kind term, then there must be something else that all evil persons have in common with one another, some property or
properties they all possess in virtue of which they count as evil. A plausible version of the mirror thesis will be “thick,” a comparatively more substantive thesis about what evil persons are like.

Second, an adequate mirror thesis will not capture just any sense in which evil persons mirror moral saints, but rather the sense in which they perversely mirror them, something like the following:

**(MT):** a person is evil just in case she is a perverse mirror image of the moral saint

A number of evil-revivalists endorse something like (MT). We are told, for example, that the evil person is motivated “in the opposite direction” of the virtuous person and thus counts as his “mirror image,”7 that the moral saint is “the positive counterpart to the evil person,”6 that certain constitutive dispositions of evil persons and moral saints are the “exact reverse of one another,”7 that the evil person occupies a space “at the other end of the moral spectrum” from the moral saint,8 and that the sort of wickedness that marks evil persons “is the exact opposite” of the sort of moral goodness that characterizes moral saints.9 Hume too appears sympathetic with (MT):

A creature, absolutely malicious and spiteful, were there any such in nature, must be worse than indifferent to the images of vice and virtue. All his sentiments must be inverted, and directly opposite to those which prevail in the human species. Whatever contributes to the good of mankind as it crosses the constant bent of his wishes and desires, must produce uneasiness and disapprobation; and on the contrary, whatever is the source of disorder and misery in society, must, for the same reason, be regarded with pleasure and complacency.10

Hume’s “absolutely malicious and spiteful” creature is surely evil, and since moral saints will surely feel unease and disapprobation upon perceiving the suffering of others, the sentiments of Hume’s evil person perversely mirror those of moral saints.

Some will object to (MT) on the grounds that evil persons must be so radically different from even morally decent people that it is simply implausible to discern any similarity between the two. At least one conception of evil personhood identifies the evil person with the satanic person: someone with “a direct desire to do what is morally wrong for its own sake” who “pursu[es] evil under the aspect of evil.”11 But satanic accounts of evil are implausible, partly because they are unable to account for the more banal strains of evil personhood and partly because they demand that we attribute implausible motives to evil persons.12 Anyone who picks their least favorite Nazi as a paradigmatic example of an evil person will be forced to confront the fact that their Nazi probably did not intrinsically desire to do wrong for its own sake nor pursue evil under the aspect of evil. It may turn out that there are no actual evil people, but surely that is an empirical matter to be determined after the best analysis of evil personhood is settled upon; accounts of evil personhood that define evil out of existence by fiat are simply implausible.

### III. Affective Versions of the Mirror Thesis

On some versions of (MT), evil persons mirror moral saints in virtue of possessing similar affective properties. Hillel Steiner offers an affective account of evil personhood insofar as he claims that, roughly, an evil person is evil in virtue of taking pleasure in his moral wrongdoing in a perversely similar manner to the moral saint.

Steiner endorses the “Negative Counterpart Thesis (NCT)” that “evil acts are simply the negative counterparts of supererogatory ones.”13 Just as supererogatory acts are morally permissible acts but not merely morally permissible, evil acts are morally forbidden acts but not merely morally forbidden ones. To account for the qualitative difference
between the two, Steiner suggests that “the basis of NCT symmetry appears to lie in the affective properties of evil and supererogatory acts. . . . Evil acts are wrong acts that are pleasurable for their doers, while supererogatory acts are right acts that are painful to perform.”14 So understood, evil acts are perverse reflections of supererogatory acts: while evil and supererogatory acts belong to different deontic categories, they are similar insofar as both are pleasurable for their agents. Thus, the qualitative difference between evil actions and merely wrong actions is that the former, but not the latter, have the property of being pleasurable.

Steiner’s NCT is a thesis about actions and not persons. However, Steiner suggests that supererogatory acts are “the actions of saints and heroes.”15 If persons count as moral saints in virtue of performing supererogatory actions, then Steiner’s discussion at least suggests the following:

\[(MT-1): \text{a person is evil just in case he is a perverse reflection of the moral saint such that he feels pleasure when performing actions that are morally wrong}\]

Since some philosophers have supposed that evil persons are defined in terms of a tendency or disposition to perform evil actions, (MT-1) merits consideration.16

As it stands, (MT-1) suggests a dubious account of supererogation. If Steiner is correct, then no one who fails to take pleasure in the performance of an act “beyond the call of duty” really supererogates. A Kantian firefighter who risks his life to save a stranger from a burning building but acts solely out of respect for the moral law, taking no pleasure in what he does, would not supererogate on Steiner’s analysis. But that seems wrong. Further, (MT-1) is too broad. If evil persons take pleasure in their morally wrong actions, then happily annoying one’s neighbor or stealing a candy bar suffices for being evil. So understood, evil personhood is purchased too cheap. Many of us sometimes take pleasure while knowing that we act wrongly; it is not that uncommon. But it is implausible to suppose that so very many of us are evil.

Whatever else is wrong with it, (MT-1) suffers from a still a deeper problem, one that becomes apparent when another affective account of evil personhood is considered. Like Steiner, Colin McGinn endorses an affective version of (MT) insofar as he identifies evil personhood with a kind of schadenfreude, or at least with the malicious enjoyment of another person’s misfortune. McGinn imagines two different kinds of beings, arguably identified with moral saints and evil persons:

Call them the G-beings and the E-beings. The G-beings are such that when another member of the species experiences pleasure, they too experience pleasure, while when another experiences pain they feel pain. . . . The E-beings, on the other hand, exemplify the opposite laws of social psychology: pleasure in one causes pain in another, and pain causes pleasure. . . . Their hedonic dispositions are the exact reverse of each other.17

McGinn’s E-beings “provide a model” for understanding the evil person: “one that derives pleasure from pain and pain from pleasure.”18 Since the hedonic dispositions of evil persons are the “direct reverse” of G-beings, then, supposing roughly that moral saints derive pleasure from the pleasure of others and pain from their pain, McGinn’s evil persons perversely mirror moral saints.

As McGinn recognizes, this analysis of evil character stands in need of refinement. It is unclear that a virtuous person could never feel pleasure at the suffering of another person; no less than Kant insists that everyone enjoys seeing the right good beating of a pest who annoys and vexes peace-loving folk.19

McGinn also makes the problematic assertion that an evil person “likes causing pain, for its own sake. . . . Pain is not merely a means to something logically independent of it.”20 Perhaps this appropriately characterizes
sadistic torture-killers, “exemplified” by Dennis Rader—also known as BTK, an acronym for “bind, torture, kill.” But whether torture-killers are motivated by an intrinsic desire to cause pain for its own sake is an open question. And in any event, it is simply implausible to suppose that more banal examples of evil persons are motivated by intrinsic desires to cause pain. If one thinks that the Nazis are evil, recall that Franz Stangl declared that he only acted as “a matter of survival” and that Hannah Arendt insisted that Adolf Eichmann “had no motives at all” aside from an interest in his personal advancement.

No matter. McGinn’s claims about the desires of evil persons can be separated from a more plausible version of (MT) to which he appears sympathetic. What is crucial is that McGinn’s evil person takes pleasure in the suffering of others, not necessarily that he is the agent of it. So McGinn would presumably endorse, not (MT-1), but the following:

\[(MT-2): \text{a person is evil just in case he is a perverse reflection of the moral saint such that he feels pleasure when he perceives the suffering of others}\]

Unlike (MT-1), (MT-2) allows for the possibility of an evil voyeur who plays no causal role in bringing about the suffering in which he takes pleasure: imagine a twisted take on A Clockwork Orange, where a perpetually tranquilized sadist is surrounded by televisions displaying graphic images of violence, all the while relishing what he sees. Since such a person does not obviously fail to count as evil, (MT-2) is pro tanto plausible in a way that (MT-1) is not.

Still, both (MT-1) and (MT-2) are plagued by a fatal objection: if either is correct, then failing to feel pleasure exonerates otherwise morally horrifying persons from the charge of being evil. Suppose that Dennis Rader initially takes pleasure in the torture killings of his victims but over time becomes jaded and indifferent, although he continues to torture and kill as ruthlessly as before. If Rader was evil, he is not now if (MT-1) or (MT-2) is correct. Intuitions may vary, but it is far from clear that Rader is a morally better sort of person simply because he now fails to take pleasure in his crimes; a Rader who tortures, binds, and kills at the “hedonic zero” is not obviously a better person than the one who takes a bit of pleasure in his work.

If this is right, then the problem with affective accounts of evil personhood is that the affective property of feeling pleasure is, at best, a contingent property of evil personhood, not a constitutive or essential property. So the locus of evil personhood lies elsewhere.

IV. MORAL REASONS AND THE MIRROR THESIS

Eve Garrard too endorses something like (MT) when she claims that evil persons take up space “at the opposite end of the moral spectrum” from an especially virtuous person. In contrast to Steiner and McGinn, Garrard does not suppose that persons are evil in virtue of their affective properties but rather in virtue of a kind of moral insensitivity, a failure to be appropriately reactive to moral reasons.

Garrard’s moral saint is not merely continent and does not merely overcome temptation to act wrongly. Rather, “the tempting [nonmoral] considerations, which would in other contexts be genuine reasons for action, lose their reason-giving force in the presence of the reasons to do the right thing.” Nonmoral reasons are silenced in the presence of moral reasons, such that Garrard’s saints do not at all regard nonmoral reasons as reasons to act. Evil persons similarly, if perversely, silence moral reasons for action insofar as they are entirely impervious—blind and deaf—to the presence of significant reasons against his acting. It is not just that he allows less impor-
tant considerations, such as his own power or pleasure, to outweigh these more forceful considerations, e.g., the suffering and loss of life of others; rather he is completely insensitive to these features’ reason-giving force. For him, there is nothing to be outweighed. . . . [He] is unable to see that they are reasons for acting or refraining from action.26

Since moral reasons are significant reasons for action, Garrard’s evil person is completely insensitive to their reason-giving force and unable to see that they are reasons for acting better. So understood, evil persons and moral saints are similar insofar as a certain class of reasons for action is silenced for each in the presence of another class of reasons for action.

But things are not quite so simple. Two different failures to respond appropriately to reasons should be distinguished: failures of receptivity and failures of reactivity. Failures of receptivity involve failed connections between reasons and an agent’s beliefs about them, while failures of reactivity involve a failure to appropriately translate beliefs about reasons into action.27 Agents who are completely insensitive to moral reasons could either fail to be receptive or reactive to them, such that Garrard could mean to defend either of the following:

(MT-3): a person is evil just in case he is a perverse reflection of the moral saint such that he fails to be receptive to moral reasons in the presence of nonmoral reasons

(MT-4): a person is evil just in case he is a perverse reflection of the moral saint such that he fails to be reactive to moral reasons in the presence of nonmoral reasons

Garrard seems to embrace (MT-3) when she suggests that evil persons cannot “see” moral reasons, that they suffer from an “inability to hear the victims’ screams as significant” and so forth.28 Theirs is “a profound cognitive inability to grasp the presence of reasons of the first importance.”29 So understood, evil persons cannot believe that there are moral reasons to act in the face of such reasons. The problem is that at least some evil persons deliberately engage in wrongdoing knowing full well that they do wrong, and that surely requires being receptive to moral reasons. If (MT-3) is true, then an agent who is receptive to moral reasons is not evil; indeed, deliberately and knowingly engaging in moral wrongdoing would exonerate a person from the charge of being evil because deliberately and knowingly engaging in moral wrongdoing requires reactivity to moral reasons that (MT-3) precludes.

More plausibly, Garrard means to defend (MT-4). Someone might be insensitive to the reason-giving force of another’s suffering if she is unaware of it, but also because she is aware of it and fails to be appropriately motivated to help. An insensitive person need not be oblivious; she may just be inappropriately motivated by that of which she is aware. So unlike the implausible (MT-3), (MT-4) allows that evil people can knowingly engage in moral wrongdoing. Still, (MT-4) is problematic for other reasons.

First, at least some putative evil persons seem to be reactive to moral reasons. At his infamous trial, Adolf Eichmann noted that he helped a half-Jewish cousin of his stepmother emigrate to Switzerland.30 Additionally, he recounted that after slapping the leader of Vienna’s Jewish community in the face—one of Eichmann’s “favorite Jews”—he repeatedly apologized in front of his staff and continued to be bothered by his behavior even after apologizing.31 Surely Eichmann’s motives and sincerity are in question, but if Eichmann could bring himself to act rightly for moral reasons on even a few occasions or regret his morally wrong actions, then he is reactive to moral reasons. Admittedly, Eichmann is a controversial example and intuitions will vary. But could not even a moral monster like Hitler act kindly towards Eva Braun and
do so for genuinely moral reasons? Not even once? Not ever?

Second, (MT-4) suggests a related thesis about moral sainthood that is far too demanding. If evil persons mirror moral saints and (MT-4) is true, then moral saints should fail to be reactive to nonmoral reasons in the presence of moral reasons for action. Such a thesis fails to square with the observation that any number of putative moral saints are not perfectly reactive to moral reasons: Martin Luther King was a philanderer; Mother Teresa confessed that for some period of time she lost her faith and on occasion declared suspect motives for caring for the hungry; Gandhi was estranged from his family and engaged in exploitive sexual relations with women to test his ascetic resolve; Socrates neglected his family for the sake of Athens; and Jimmy Carter, a deeply pious and religious man, admitted to having “lust in his heart.”32 But if our actual experience of putative moral saints suggests that they can be reactive to nonmoral reasons, it is unclear why evil persons could not similarly be reactive to moral reasons.

V. A CONSISTENCY VERSION OF THE MIRROR THESIS

Dan Haybron also appears to endorse some version of (MT) when he claims that the moral saint is “the positive counterpart to the evil person.” Haybron suggests that a person is evil only if her character is “thoroughly and consistently vicious,” where being thoroughly and consistently vicious is to lack significant moral virtues: to have “no good side.”33 Following Haybron, a deeply pious and religious man, admitted to having “lust in his heart.”32 But if our actual experience of putative moral saints suggests that they can be reactive to nonmoral reasons, it is unclear why evil persons could not similarly be reactive to moral reasons.

Haybron sometimes understands evil personhood not in terms of the consistent lack of moral virtue but in terms of the consistent absence of morally appropriate motivation. For example, he suggests that evil persons “are moved and motivated little or not at all, or even perversely, by morality and the good,”35 and that “to be evil is to be disposed to neither moved nor motivated (positively) by the good to a morally significant extent.”36 Using an apt phrase, Haybron claims that while being appropriately motivated by the good is to be “aligned with the good,” evil persons are “unaligned” with the good.37 These two descriptions of evil persons mesh. While it is a mistake to identify desires and virtues, there is surely some correlation between degree of virtue and strength of desire. As a general rule, the greater the degree of a virtuous character trait, the stronger the relevant desire to act accordingly, and the weaker the degree of a virtuous character trait, the weaker the relevant desire to act accordingly. And certainly if someone lacks any motivation to act rightly when the opportunity arises, he simply lacks moral virtue altogether.

Following Haybron, two different consistency accounts should be distinguished. On a weak version, only the absence of “morally significant” virtue is required for being evil, virtue that makes a significant difference to the overall moral quality of a person’s character. On a strong version, evil persons lack moral virtue, period.38 So Haybron could mean to endorse either of the following:

(MT-5): a person is evil just in case he is a perverse reflection of the moral saint such that he lacks significant moral virtue

(MT-6): a person is evil just in case he is a perverse reflection of the moral saint such that he lacks moral virtue altogether

(MT-5) allows that evil persons may be a slightly kind or generous, but not so much so that the overall moral quality of their characters improves. By contrast, (MT-6) implies
that they cannot be kind or generous at all. Perhaps Eichmann’s assertion that he helped a distant relative escape the Holocaust does count as evidence that he has some modest moral virtue; Eichmann might still be evil if (MT-5) is correct even supposing that he is not utterly without moral virtue, but he is simply not evil if (MT-6) is correct, whatever else is wrong with him.

Interestingly, by Haybron’s own lights, (MT-5) is probably false. If (MT-5) were true then mere ne’er-do-wells—individuals who genuinely care about matters of duty and the good of others but are never moved or motivated by such concerns—would be evil.39 Ne’er-do-wells apparently possess some moral virtue even though they are never sufficiently motivated to act rightly; hence the name.40 But a ne’er-do-well could surely be a worse sort of person; namely, one who is greatly motivated to act wrongly absent remorse or compunction. So moral theorists inclined towards consistency accounts ought to embrace (MT-6) rather than (MT-5), if they accept either.

But (MT-6) fares little better than its sibling. There are several reasons to question (MT-6), not the least of which is that it is too austere an account of evil personhood: it says far too little about what evil persons are like aside from pointing out what they lack. First, it is not clear that individuals who consistently lack moral virtue must be evil.41 In the midst of discussing the “noble savage,” Rousseau remarks that “men in the state of nature, having no kind of moral relationships between them, or any known duties, could be neither good nor evil, and that they could have neither virtues nor vices,” and he concludes that they “are not wicked precisely because they do not know what it is to be good.”42 One problem is that (MT-6) implies otherwise—that creatures lacking moral concepts altogether are evil absent consideration of any conditions that would exempt them from moral responsibility. Second, while it is probably a mistake to suppose that evil persons must act wrongly, it is not implausible to suppose that they are disposed to act wrongly and that part of their evil character consists in their being so disposed. (MT-6) cannot explain why this should be so, since it only implies that evil persons lack moral virtue and not that they possess moral vices or morally dubious intentions and desires.

Thus far, no plausible version of the mirror thesis has been located. A different strategy is needed. Since (MT) implies that evil persons are perverse mirror images of moral saints, identifying the most plausible account of moral sainthood should be revealing. To that end, consider what the morally best sort of person is like.

VI. AN ASIDE ON MORAL SAINTHOOD

According to one scholar, recent philosophical discussions of saints suggest that even if the ideal of a life in which every action is morally right is not appealing, real-life moral saints do not fit this ideal, since they have both nonmoral aspirations and projects as well as ordinary human flaws.33 This suggestion is relevant to evaluating Susan Wolf’s important and influential discussion of moral sainthood. Of present concern is not Wolf’s argument that moral sainthood does not constitute an ideal of human well-being, but rather the account of moral sainthood that she implicitly invokes.

Wolf claims variously that the moral saint is “a person whose every action is as morally good as possible,” that the moral saint is “a person . . . who is as morally worthy as can be,” and that “A necessary condition of 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.”44 These remarks suggest three different theses concerning moral sainthood:

(W1): an agent, a, is a moral saint only if, for every action, ψ, that a performs, ψ is as morally good as possible
(W2): an agent, $a$, is a moral saint only if $a$ is as morally worthy as can be.

(W3): an agent, $a$, is a moral saint only if $a$'s life is dominated by a commitment to improving the welfare of others.

As has been observed, these theses are not equivalent. (W1) implies that just once settling for satisfying rather than optimizing moral goodness disqualifies an agent from moral sainthood; not so with respect to (W2) and (W3). Albert Schweitzer kept and played a piano during the midst of his humanitarian activities, and thus it is probably not the case that every one of his actions is as morally good as possible. Still, Schweitzer may be as morally worthy as can be—at least, as worthy as human beings can be—and his life is surely still dominated by a commitment to improve the welfare of others. So (W1) does not entail either (W2) or (W3).

More importantly, (W1) is not as plausible as either (W2) or (W3). Rhetorically, does anyone count as a moral saint if Schweitzer does not? For any putative moral saint, imagine a counterpart whose every action is as morally good as possible: imagine a King who never cheated, a Gandhi who never exploited, a Schweitzer who never entertained himself by playing the piano, and so forth. If (W1) is correct, these putative saints must be comparatively worse persons than their counterparts. But arguably, these putative moral saints are such good people that these slight variations yield no change in the overall moral quality of their characters. As such, although these putative moral saints suffer from moral vices that their counterparts do not, it is false that their counterparts are morally better persons. But if those counterparts are not better persons than these putative moral saints, then (W1) is false.

(W2) and (W3) fare comparatively better. As long as (W2) is constrained by what is psychologically possible for human beings, it will not be vulnerable to the sort of counterexamples that plague (W1). And as suggested above, (W3) can tolerate the possibility that a moral saint will both take time for herself and live a life dominated by concern for others. But consider other implications of (W3). While the lives of King and Mother Teresa and Schweitzer are dominated by a commitment to improving the welfare of others, their commitment is surely paired with significant moral virtues. Wolf seems to agree when she asserts that a moral saint “will have the standard moral virtues to a non-standard degree” and have them to “an extreme degree.” So, by Wolf’s own lights, if a person’s life is dominated by a commitment to improving the welfare of others then she will possess extreme moral virtues, extreme in the sense that they will be held to significant degrees.

Importantly, however, the virtues of a saint are probably extreme in another sense, not in the sense of being held to a significant degree but in the sense of being among the very best of their kind. Meekness and kindness are perhaps both moral virtues, but surely kindness and justice are virtues of a comparatively better sort. Persons who merely possess comparatively lesser virtues are probably not saints, even if those virtues are possessed to significant degrees; however meek someone is, a better sort of person would be just as meek but also courageous, for example. So moral saints do not only possess virtues that are extreme in the first sense. But neither do they only possess virtues that are extreme in the second sense; someone who is significantly just is a better sort of person than someone who is only slightly just. Since moral saints are plausibly identified with the morally best sort of person, moral saints must possess extreme virtue in both senses—that is, the virtues of the moral saint include the morally best sort of virtues possessed to significant degrees.

Reflection on (W2) and (W3) suggests that moral saints possess extreme virtues.
However, an even stronger thesis seems warranted:

(W4): an agent, a, is a moral saint just in case a possesses extremely virtuous states of character.

(W4) implies that possession of extreme moral virtue is not just necessary but sufficient for moral sainthood. But this is not implausible. Explaining why calls for some reflection on the moral virtues and their relationships to one another.

On one view of things, every moral virtue is systematically related to every other moral virtue. The thesis that a person can possess a moral virtue only if she possess every other moral virtue as well—the so-called “unity of the virtues” thesis—is familiar enough and has an enviable pedigree. But while the unity of the virtues thesis continues to find adherents, it is often enough rejected by contemporary philosophers. But one need not endorse the unity thesis to suppose that the virtues are not systematically disunited to one another. Contemporary work in virtue theory suggests that virtues are multitrack dispositional states—that they dispose their agents to act in certain ways in certain circumstances, to act for certain reasons with certain feelings, and so forth. Some of the dispositions constitutive of particular virtues will be shared; both generosity and justice dispose their agents to help others in certain circumstances; both honesty and magnanimity dispose their agents to tell the truth to others; both kindness and friendliness involve caring for others and taking their well-being to be a reason for action, and so forth. In broad strokes, at least some particular virtues are consistently correlated—and hence, not entirely disunited—with others because they share constitutive dispositional states.

Supposing that the virtues are not entirely disunited, it seems clear that possessing extreme virtue probably entails possessing a great many other moral and nonmoral virtues as well; especially generous persons will also probably have to be resolved, dedicated, intelligent, unselfish, and so forth. So the possession of extremely virtuous states of character will tend to ensure that their possessor has a reasonably well-rounded set of virtues and that she probably is not lacking anything that we expect the morally best sort of person to have.

At least two arguments support (W4). First, (W4) helps to distinguish moral saints from (merely) morally decent people. Morally decent people may be somewhat compassionate and loyal and courageous, but not to the degree that a moral saint is: King was extremely courageous, not just somewhat so; Schweitzer was extremely compassionate, not just somewhat so; and so forth. Second, (W4) explains why some moral saints are rightly regarded as such despite their moral failings. If they did not possess extremely virtuous character traits, it is unlikely that their comparatively slight moral failings would be more than compensated for such that we still regard them as moral saints.

There is good reason, then, to suppose that (W4) is correct. And that yields important results for an account of evil personhood.

VI. AN EXTREMITY VERSION OF THE MIRROR THESIS

If possession of extreme moral virtue suffices for moral sainthood, then the following thesis may well seem plausible:

(E): an agent, a, is an evil person just in case a possesses extremely vicious states of character.

In the same way moral virtue can be extreme in at least two senses, moral vices might be extreme if possessed to significant degrees or if they are among the morally worst of vices or both. And just as (W4) is most plausible when the virtues of saints are understood to be extreme in both senses, (E) is most plausible when the vices of evil people are understood
to be extreme in both senses. Thus, (E) implies that suffering from the morally worst of vices to significant degrees suffices for being evil.\(^52\)

It is probably not news to be told that evil people suffer from moral vices. But (E) makes a stronger claim, asserting that evil people must suffer from the morally worst vices to significant degrees, not just any vices to any degree. More controversial is (E)'s implication that possession of extreme vice suffices for being evil. But this implication, too, is not implausible. If virtues are properly understood as multitrack dispositional states, vices should also be understood as such. And just as particular constitutive dispositions are tied up with multiple virtues, particular constitutive dispositions are tied up with multiple vices. So just as it is plausible that the virtues are not entirely disunited, it is plausible to expect some limited unity of the vices as well: people who are malicious will probably also be boorish; people who are vain will probably also be boorish buffoons; people who are mean will probably also be stingy and petty and cruel, and so forth. So just as possession of extreme virtue suggests a generally exceptionally laudable character lacking nothing that should be expected from a moral saint, possession of extreme vice suggests a generally exceptionally lamentable character lacking nothing that should be expected of an evil person.

Aside from what I take to be its intuitive plausibility, at least four arguments support (E). First, evil persons should be distinguished from merely indecent persons. Morally indecent people surely suffer from some vices, but their indecency might be a result of comparatively minor moral vices like pettiness or stinginess and probably not a result of more significant moral vices like maliciousness or malevolence. But perhaps some people are indecent in virtue of suffering from much more grave moral vices, such as cruelty or callousness, for example. Still, those merely indecent people surely do not suffer from those grave vices to the significant degree that evil persons do. Many ordinary people suffer from some moral vice, after all; the difference is that their moral vices are modest or tepid in comparison to the vices of evil people and thus, merely indecent people are not evil. (E) explains why this is so.

Second, unlike (MT-6), (E) does not imply that evil persons lack moral virtue altogether; indeed, (E) allows that evil persons possess some degree of moral virtue, albeit rather minimal degrees of it lest they be morally redeemable in a way that evil people are surely not. But this implication speaks in favor of (E). While Kant does not doubt that some human beings are evil, he does doubt human beings can be “devilish”—that our reason could be “entirely exempt from the moral law.”\(^53\) On Kant’s view, supposing that evil people are entirely beyond the reach of the moral law amounts to defining evil people out of existence by demanding more than what is possible for creatures like us. In a slightly different vein, no small number of social psychologists interested in evil have stressed that the agents of the morally worst crimes are, in other respects, similar to common, morally decent folk.\(^54\) In particular, Erich Fromm echoes Kant when he cautions that if one supposes that evil persons must be “utterly devoid of any kindness, of any good intentions . . . one will not discover an evil man.”\(^55\) Here, too, a theory of evil that makes unrealistic demands for evil personhood is implausible insofar as it defines evil personhood out of existence. (E) avoids these implausible results.

Third, if even evil persons can possess some minimal degree of moral virtue, (E) explains why they are nonetheless rightly regarded as evil: their vices are so extreme that their overall moral character is not improved by the fact that they are slightly virtuous. The evil person’s extreme vices “poison the well,” as it were. If evil persons did not pos-
sess extremely vicious character traits, it is unlikely that their moral failings would dwarf whatever minimal virtue they possess such that we still rightly regard them as evil.

Finally, (E) explains a popular thesis among philosophers interested in evil persons and evil actions, that being evil is qualitatively different from being morally wrong or morally bad and even being very, very morally wrong or bad.56 This seems plausible and (E) explains why this should be so. According to (E), an evil person suffers from extreme moral vices, both in the sense that they are possessed to significant degrees and in the sense that they are among the morally worst sort of vices. Now, perhaps the degree to which a vice is possessed can be quantified such that the difference between a vice possessed to a significant degree and a vice possessed to a more modest degree is a quantitative difference. However, the difference between a vice that is among the morally worst of its kind, on the one hand, and a vice that is not, on the other, is not a quantitative difference, but a qualitative one. No surprise, then, if being evil is to be a qualitatively morally worse sort of person than being a bad or even a very, very bad person: being evil is partly a function of suffering from vices that are qualitatively worse than other vices.

If (W4) is a correct account of moral sainthood and (E) is a correct account of evil personhood, then the following should seem plausible:

\[(MT-7): \text{a person is evil just in case he is a perverse reflection of the moral saint such that he suffers from extremely vicious character traits}\]

If (W4) and (E) are correct, then there is some relevant similarity between evil persons and moral saints: both possess extreme traits of character constitutive being the morally worst and best sort of person, as (MT-7) suggests. But if (MT-7) is true, then (MT) is true: evil persons are the perverse mirror images of moral saints. As such, the mirror thesis is vindicated.

To be sure, the mirror thesis is hardly adequate as an account of moral sainthood; roughly, it characterizes the structure of an evil person’s character but not its content. A full-blown account of evil personhood would also say something about what particular vices, if any, from which an evil person must suffer. Further, the mirror thesis does not explicitly suggest any particular relation between evil personhood and evil action, whatever relation that should be. But any account of evil personhood must start somewhere, and the mirror thesis represents a plausible place to start. What began as a metaphorical thesis about evil persons is grounded in a literal substantive thesis about evil character.

\[\text{Saginaw Valley State University}\]

\[\text{NOTES}\]


2. To anticipate, philosophical interest in evil has traditionally been interest in the so-called problem of evil: the problem of reconciling (or not) the apparent incompatibility of the existence of a supremely loving and caring and wise God with the existence of gratuitous suffering. That traditional problem of evil is different from the present problem of understanding what evil persons are like.


18. Ibid., p. 62.

21. According to psychiatrist Michael Stone, the “most evil” persons are “psychopathic torture-murderers with torture their primary motive,” more evil, apparently, than torture murderers with some other primary motive. See http://www.freerepublic.com/focus/f-news/1347361/posts.
26. Ibid., p. 330. Like Steiner, Garrard is officially providing an account of evil action, not of evil personhood. But Garrard’s account is equally suggestive of an account of evil personhood.


31. Ibid., pp. 46–47.


35. Ibid., p. 63.

36. Ibid., p. 70.

37. Ibid., p. 71.

38. Ibid., p. 70.

39. Ibid.

40. Ibid.


46. Ibid., p. 396.


51. Even Haybron claims that it is the “profound compassion” of the moral saint that impresses us (Haybron, “Moral Monsters and Saints,” p. 274).


