Defective Actions and Tyrannical Souls: Korsgaard on Evil

PETER BRIAN ROSE-BARRY *Saginaw Valley State University*

ABSTRACT: Christine Korsgaard's Self-Constitution: Agency, Identity, and Integrity *is an impressive endeavour to synthesize the ethics of Plato and Kant in a comprehensive account of action and agency that locates the key to understanding both in self-constitution. A purportedly comprehensive account of action and agency will fail on its own terms if it cannot adequately account for some morally salient phenomenon. Korsgaard's account fails to adequately account for the possibility of evil actions and evil people. If self-constitution is key to action and agency, then we must abandon the Platonic and Kantian elements that Korsgaard endorses.*

RÉSUMÉ : L'ouvrage de Christine Korsgaard intitulé Self-Constitution: Agency, Identity, and Integrity *représente un remarquable travail de synthèse des éthiques platonicienne et kantienne qui, de manière exhaustive, rend compte de l'action et de l'agentivité sous le prisme de la constitution de soi. Toute tentative de saisie parfaite de l'action et de l'agentivité en termes propres ne saurait aboutir sans une prise en compte adéquate d'un quelconque phénomène moral saillant. L'approche de Korsgaard ne prend pas suffisamment en compte la possibilité du mal sous forme d'actions ou d'individus. Si la constitution de soi demeure centrale pour l'action et l'agentivité, alors il faudra mettre à l'écart les éléments platoniciens et kantiens que Korsgaard fait siens.*

Keywords: evil, Kantian ethics, Christine Korsgaard, Plato, self-constitution, agency

Christine Korsgaard's *Self-Constitution: Agency, Identity, and Integrity* is an impressive endeavour to synthesize the ethics of Plato and Kant, one that proposes a comprehensive account of action and agency and locates the key to

***Dialogue* 57 (2018), 29–46.**

doi:10.1017/S0012217317000567

understanding both in self-constitution.[1] Ultimately, I fear, it fails in its ambitious aims. A purportedly comprehensive account of action and agency will fail on its own terms if it cannot adequately account for some morally salient phenomenon. Korsgaard's account fails in this respect because Korsgaard is committed to denying the very possibility of evil agents and evil actions, notwithstanding her own attempts to account for each. If self-constitution is to play the role that Korsgaard suggests it does in agency and in action, then evil agents and evil actions simply cannot exist. The problem, I contend, is that Korsgaard cannot really afford self-constitution the weighty role that she affords it while retaining her Platonic and Kantian sympathies.

Making Something Out of Yourself

Any attempt to critique a work as wide ranging as Korsgaard's will omit some details. I will rest content with summarizing those aspects of her project that interest me and explaining where things go awry.

There are comparatively simple conceptions of action and agency and there are comparatively "fancy" accounts.[2] On a very simple conception, an action just is a bit of behaviour that is caused by an agent's reasons "in the right way"[3] and an agent just is the sort of creature who acts. The account of action and agency that Korsgaard defends includes fancier conceptions of each.

Action, for Korsgaard, is not simply behaviour caused in the right way by an agent's reasons. We are told variously that to act is to constitute oneself as the cause of an end (72, 79), that the function of action is self-constitution (xii, 82), and that action *is* self-constitution (25, 45, 82, 96). The project of explaining just what self-constitution requires is complicated and explication here takes up a fair bit of *Self-Constitution*, but briefly, self-constitution amounts to a kind of unification, the successful achievement of a kind of psychic unity (7). Achieving self-constitution "involves finding some roles and fulfilling them with integrity and dedication" (25). Elsewhere, Korsgaard speaks of "practical identities,"[4] those descriptions of ourselves under which we find our lives to be worth living and our activities to be worth undertaking. Many, if not most, of our practical identities will call for us to take on certain roles: the role of a philosopher, a parent, an activist, or whatever. An agent who achieves self-constitution is able to unify all of these roles; she sincerely affirms a set of practical identities simultaneously, finding meaning in all of their constitutive activities consistent with the demands and requirements that each place on her.

---

1 Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*. Subsequent references to this work are parenthetical.

2 Alfred Mele, *Motivation and Agency*, p. 216.

3 For the canonical version of this account, see Donald Davidson, *Essays on Actions and Events*, p. 79.

4 Christine Korsgaard, *The Sources of Normativity*, p. 101.

This can be tricky business given that some practical identities will conflict with others. Sometimes different practical identities will be logically incompatible: no one could consistently affirm a practical identity as an atheist and as a true believer, for example. More commonly, practical identities will conflict on pragmatic grounds given the contingencies of everyday life: sooner or later, the obligations endemic to one's role as a philosopher and those endemic to one's role as a parent are going to make irresolvable demands on one's time. But someone who successfully achieves self-constitution is able to overcome such conflicts and thereby succeeds at "integrating those roles into a single identity, into a coherent life" (25).

So long as the essence of action is self-constitution, action must be more than behaviour caused by an agent's reasons. We are told repeatedly that it is essential to the very concept that action is attributable to a person as a whole and not simply to forces at work inside her (xii, 18, 19, 45, 133). Since agents are not identical to their desires, even if some behaviour is caused by and therefore attributable to one of an agent's desires it is not thereby attributable to *her*. Korsgaard does not clearly deny that desires have *some* role to play in the production of action; she can't, given that some practical identities demand that the agent who adopts them has particular desires. A good parent *wants* her children to prosper and anyone who lacks a desire for her child's prosperity that is sufficiently robust to incline her to act has not fulfilled her role as a parent with integrity. But Korsgaard does deny that our actions, properly understood, can be understood merely as the causal product of our desires. Action, for Korsgaard, requires *agency* (45).

And agency, in turn, requires *unity.* If there is a pithy slogan to be found in *Self-Constitution*, it is this: "action requires agency, and agency requires unity" (45). Unless something pulls those desires together, there is no agent doing anything, only a "mere heap, of unrelated impulses" (76). Just as a collection of stones is not a wall unless they are pulled together and unified, a heap of desires is not a person unless something pulls them together and unifies them. And heaps of desires do not pull themselves together; *agents* do that by engaging in some kind of activity that pulls heaps of desires together into a unified whole and thereby makes something out of them. Not surprisingly, Korsgaard supposes that practical deliberation is the activity that will do the work of unification—or perhaps "reunification" (133)—and thereby transform a disparate heap into a unified whole. The upshot of this kind of practical deliberation, the kind that results in unification, is self-constitution.

However, not just any psychological activity, however deliberate, is sufficient for practical deliberation. In cases of particularistic willing, an agent makes claim to a reason for action that applies only to the particular case before her with no implications for other cases (73). But in such cases, Korsgaard argues, the agent fails to distinguish herself from the incentives that prompt her to deliberate in the first place (75). Given the particularistic nature of particularistic willing, the agent in question could wholly identify with one incentive on one

occasion, a different incentive on another, and so forth. At each different moment of particularistic willing, a different incentive would, at that particular moment, be her will. But then there is no unified whole, no agent distinct from the various desires that variously constitute her will (76). So, particularistic willing is an insufficient basis for action and agency just because it cannot unify.

But if particularistic willing is insufficient for action, then action requires universalizable willing: that is, action must involve a maxim that the agent takes to be universalizable (76). No surprise, Korsgaard concludes that Kant's Hypothetical Imperative (HI) and Categorical Imperative (CI) are just the principles of practical deliberation to guide agents here. We are told repeatedly that the HI and CI are constitutive of agency and action (48, 52, 58, 69, 72, 76, 80, 213).[5] Why? Because, Korsgaard explains, it is by conforming to the HI and the CI that an agent distinguishes herself from the incentives that move her (76, 81). The HI and the CI do not only have an essential role to play in the production of action; they are essential to agency too. Following Plato, Korsgaard holds that the kind of unity required for agency is the kind of unity that a city has in virtue of having a just constitution (157). And just as Plato contends that an unjust city cannot act, Korsgaard holds that "an unjust person cannot act at all" (152). Why? Because of the conception of agency that Korsgaard has defended. Again, agency requires unification and particularistic willing cannot result in unification. Absent some overriding principle to guide particular attempts at unification, the result will be an unorganized mess, not a unified whole. So, unification requires universalizable principles—that is, the HI and CI. But the HI and CI are moral principles, principles that guide a just person. If all this is right, then only just people can be agents and only the just can act since only they are guided by universalizable principles that can result in unification.

There is much that I am glossing over, but it is enough for present purposes to note that Korsgaard pretty clearly endorses the following two theses:

> K1): A person counts as an agent only if she is unified and she is unified only if she is guided by the principles of practical reason—that is, the HI and the CI.

> K2): A person's behaviour counts as action only if her behaviour unifies her as an agent and her behaviour unifies her as an agent only if it is guided by the principles of practical reason—that is, the HI and the CI.

These two theses are sufficient to generate the problems that I suggest above. In the next section, I explain why evil action is impossible if K2 is true. I then explain why evil agents are similarly impossible if K1) is true.

---

[5] Korsgaard also insists that there is just one principle of practical reason: the CI (81).

How to Act Badly

Korsgaard frequently describes HI and CI as constitutive principles of action and agency. In its most general sense, a constitutive principle is a standard that applies to an activity (28). A constitutive principle of some activity is that principle that describes the way in which an agent engaged in that activity directs or guides herself (29). An agent who hopes to engage in some activity *must* be guided by those principles to perform that activity *at all* (29, 31). To illustrate, consider one of Korsgaard's favourite examples: someone who hopes to build houses must look to the constitutive principles governing house building and someone who fails to even consider those principles just isn't building a house even if he successfully slaps bricks and mortar together (29-30). An object or activity that fails to meet the constitutive standards associated with it is "defective" (32).

It is important to note a crucial distinction in play: there is a difference between *defective* activity and what I will call 'deficient' action. Defective activity is, by definition, not action since the agents of defective activity are not guided by action's constitutive principles, just as activity that isn't guided by the constitutive principles of house building isn't really building houses. By contrast, deficient actions really are actions since they are guided by action's constitutive principles, but they are actions badly done. Korsgaard clearly recognizes that there is a category of action that isn't *faux* or counterfeit action in the way that defective activity is, but nonetheless falls short of action *par excellence*. Here is Korsgaard on house building again:

> Building a good house and building a house are not different activities: for both are activities guided by the teleological norms implicit in the idea of a house. Obviously, it doesn't follow that every house is a good house, although there is a puzzle about why not. It does, however, follow that building bad houses is not a different activity from building good ones. *It is the same activity, badly done.* (29)

Similar remarks characterize speaking English poorly (30), illogical thinking (32), leading the life of an unhealthy giraffe (37), and leading an unjust life (180). One engages in defective action, I stipulate, when one engages in activity *badly done*. But what is activity badly done? In the case of the shoddy house builder, what makes it the case that he is properly described as building houses, albeit badly? Korsgaard suggests that a shoddy builder may be doing one of two things:

> He may be guided by the norms [constitutive of house building], but carelessly, inattentively, choosing second-rate materials in a random way, sealing the corners imperfectly, adding insufficient insulation, and so on. But he may also, if he is dishonest, be doing this sort of thing quite consciously, say in order to save money. In that case, surely we can't say that he is trying to build a good house? No, but now I think we

> should follow Socrates' lead, and say that he is not trying to build a house at all, but rather a sort of plausible imitation of a house, one he can pass off as the real thing. What guides him is not the aim of producing a house, but the aim of producing something that will fetch the price of a house … (31)

Really, there is really only one house builder here: the first one, the shoddy-because-careless-or-inattentive-or-whatever builder. Since the first shoddy builder is guided by the principles constitutive of house building, albeit carelessly or inattentively or whatever, he really does engage in the activity of house building although he builds *deficiently*. The example of the first shoddy house builder suggests a general way of categorizing deficient action. Deficient actions really are actions since they are guided by principles of practical reason, but they are deficient given their agents are guided poorly by those principles—that is, guided carelessly or inattentively or whatever. Alternatively, deficient action is action in which agents are guided *badly* by the principles of practical reasoning. The first shoddy house builder builds badly and he builds deficiently, but he still builds. By contrast, the second shoddy builder is not guided by the principles constitutive of house building at all. For that reason, Korsgaard is committed to denying that he is really building houses (even if he is building *faux* houses, or something like that). Since the second sort of shoddy builder isn't guided by the principles constitutive of house building, he does not build badly or deficiently, but *defectively*—that is, he doesn't really build at all.

Again, it is clear given her remarks about shoddy house building that Korsgaard has the concept of deficient action in mind even if she does not invoke any specific terminology to refer to it; otherwise, there would be no difference between the two shoddy house builders. But, more importantly, the concept of defective activity is, as far as I can tell, the key to understanding how bad action is possible on Korsgaard's account. At times, Korsgaard identifies bad actions with defective actions (32). Defective actions are bad not because they are cruel or unjust or whatever, but because they "fail to constitute their agents as the unified authors of their actions" (25, 32). But recall K2): action demands unification and unification requires guidance by principles of practical reason. So, while things might come in degrees here (25), we are told that "your action must, in order to be a good one—one that serves its function—conform to these imperatives" (83)—that is, the HI and the CI. K2) implies that being guided by principles of practical reason is a necessary condition for behaviour to count as action *at all*: "if we don't follow them we just aren't acting" (32). Accordingly, defective actions cannot really count as bad actions because they are not actions at all. But it doesn't follow that Korsgaard cannot account of the possibility of bad action even if she is right to ask "How then is the shoddy builder even possible?" (30). She could instead identify bad actions with deficient actions—that is, actions that are guided by principles of practical reason, albeit badly.

This is not an *ad hoc* response. At least some of what we call 'bad action' is analogous to the first variety of shoddy house building noted above: deficient

action performed by an agent who is careless or inattentive or whatever. Some actions are bad because they count as instances of negligence; perhaps someone acts with good ends in mind or for a noble cause, but she is culpably inattentive to the consequences of her actions or wrongly fails to consider the full range of persons who will be affected, she is negligent and therefore she acts badly. The agent who acts negligently fails to do what a reasonable person would do in those circumstances and that is clearly a kind of moral failing. And, arguably, the general recipe for inexcusable action is "a steady refusal to attend both to the consequences of one's actions and to the principles involved."[6]

How plausible is it that *every* case of bad action is akin to shoddy-because-careless-or-inattentive house building? How plausible is it that *all* cases of bad action are cases of negligence, even grossly negligent action? As suggested above, I am actually inclined to think that a fair bit of moral wrongdoing can be construed as a kind of moral negligence, spawned by fairly ordinary moral vices like callousness, coldness, inattentiveness, laziness, indifference, and so forth. However, there is a significant range of moral wrongdoing that is not well understood as being akin to shoddy house building or as a variety of even gross negligence: that would be *evil* action.

On Evil Action

Before proceeding too much further, it will be helpful to say several things about the concept of evil action. There is, not surprisingly, some dispute about just what it is that makes an action evil. Some philosophers contend that evil actions tend to result in, not harm simpliciter, but *grave* harm, or, harm that is variously described as intolerable,[7] significant,[8] very serious,[9] both serious and excessive,[10] "life-wrecking,"[11] immense,[12] and so forth. Other philosophers reject a grave harm requirement insofar as they allow for the possibility of "small-scale evils"—that is, genuinely evil actions that do not result harm on a par with the harm caused by paradigmatically evil actions.[13] Some philosophers endorse comparatively "thin" conceptions of evildoing according to which evil actions just are culpable wrongs that cause or tend to cause sufficiently grave harm while other philosophers advocate comparatively "thick" conceptions of evildoing that imply that evil actions must be the product of some sufficiently

6 Mary Midgley, *Wickedness*, p. 64.
7 Claudia Card, *The Atrocity Paradigm*, p. 16.
8 Todd Calder, "Is Evil Just Very Wrong?," p. 188.
9 Matthew Kramer, *The Ethics of Capital Punishment*, p. 223.
10 John Kekes, *The Roots of Evil*, p. 2.
11 Paul Formosa, "A Conception of Evil," p. 228.
12 Marcus Singer, "The Concept of Evil," p. 193.
13 Eve Garrard, "The Nature of Evil," p. 45; and Adam Morton, *On Evil*, p. 60.

morally dubious psychological antecedents.[14] There is little chance of settling at present whether an adequate conception of evildoing must be thin or thick, whether it must incorporate a grave harm requirement, or whether some other alternative is live. However, at least two points are worth noting, both of which are going to be problematic for Korsgaard's account.

First, whether one thinks that the psychological antecedents must be especially morally dubious or merely sufficient to make an agent culpable for her wrongdoing, simple negligence is probably not sufficient to produce evildoing. To see this, consider the evolution of Claudia Card's influential conception of evil and evil action. Card has produced an impressive corpus on the subject of evil, and in an early volume she contends that:

> … an evil is harm that is (1) reasonably foreseeable (or appreciable) and (2) culpably inflicted (or tolerated, aggravated, or maintained), and that (3) deprives, or seriously risks depriving, others of the basics that are necessary to make a life possible and tolerable or decent (or to make a death decent).[15]

More pithily, "evils are understood as reasonably foreseeable intolerable harms produced by culpable wrongdoing."[16] However, an agent might be culpable for a piece of wrongdoing simply in virtue of being negligent and, arguably, the "moral gravity" of a negligent action is never enough to make an action evil.[17] Card herself has come to agree. Card has since revised her understanding of evil such that she understands evils to be reasonably foreseeable intolerable harms produced not by culpable wrongdoing but *inexcusable* wrongdoing, a revision that "clarifies and preserves culpability in the evil deeds of individuals."[18] Since negligence, at least sometimes, involves responsibility-diminishing ignorance that mitigates responsibility,[19] Card's revision ensures that negligence is an insufficient basis for evildoing.

Hopefully the problem for Korsgaard is clear. In the previous section, I argued that Korsgaard is committed to supposing that bad actions just are instances of deficient activity amounting to negligence. But evil actions are *not* merely deficient activity amounting to negligence. The problem is not that Korsgaard is committed to denying that evil actions are bad ones, although that would be a problem in itself. The problem is that Korsgaard has insufficient resources to explain why some acts are not merely wrong but evil. Accordingly, she is committed to either denying the existence of evil actions or rejecting the account of

14 Luke Russell, *Evil*, p. 69.

15 Card, *The Atrocity Paradigm*, p. 16.

16 Ibid., p. 3.

17 Kramer, *The Ethics of Capital Punishment*, p. 188.

18 Claudia Card, *Confronting Evils*, p. 16.

19 Ibid.

bad action expressed by K2), an option that amounts to moving well beyond the project of *Self-Constitution*. Supposing that Korsgaard does not intend to abandon her own account of bad action, the only available option is to deny that there are any evil actions.

The second point is related to the first. It is widely agreed that evil actions are qualitatively distinct from ordinary wrongs; this is sometimes held to be an intuition that is supposed to guide philosophical reflection about evil action.[20] But, if all bad actions are bad in virtue of being deficient, then all bad actions, including evil actions, are bad in virtue of the negligence of their agents. But the moral gravity of evil actions is wildly under-described by noting only that they are the product of negligence. It might be possible for Korsgaard to continue to adhere to a thin conception of evildoing but, in order to account for the qualitative distinctness of evil actions, to also endorse a grave harm requirement. However, that move is somewhat at odds with Korsgaard's Kantian and Platonic sympathies, and her project of locating the whole of ethics in facts about self-constitution would be imperiled if we must also appeal to consequentialist considerations to fully account for a morally salient phenomenon. She could alternatively decline to offer a conception of evil action at all and hold instead that there just are no evil actions, only bad actions understood as deficient activity. The defender of Korsgaard might insist that "evil-scepticism"[21] is a coherent position and note that some philosophers have argued that we ought to purge the concept of evil from our moral discourse altogether.[22] However, evil-scepticism, to the extent that it is plausible, is most plausible as a thesis about evil *people*; in its weaker versions, it entails only that relatively few actual people are evil while in its strongest version it implies that evil people are altogether impossible.[23] Understood as a thesis about evil *actions*, evil-scepticism is much, much less plausible if only because it is rather less contentious that there are actual evil actions.[24] True, Korsgaard need not be a sceptic about evil simpliciter; she is only committed to denying that there are evil actions, a result consistent with supposing that there is evil behaviour that doesn't quite qualify as action.[25] But, if it is terribly plausible to suppose that evil actions exist, then denying the existence of evil actions would seem to be a *reductio* of Korsgaard's view. Further, it is at least plausible to suppose that deliberate and intentional evildoing

---

20 Russell, *Evil*, p. 34.

21 I borrow this term from Luke Russell's "Evil-Revivalism Versus Evil-Skepticism," pp. 89-105.

22 See, for example, Richard Bernstein, *The Abuse of Evil*; and Philip Cole, *The Myth of Evil*.

23 Peter Brian Barry, *Evil and Moral Psychology*, p. 6.

24 Russell, *Evil*, p. 32.

25 My thanks to an anonymous reviewer for pointing this out.

is morally worse than mere negligent wrongdoing, however harmful. Since evil actions are the worst wrongs that people do,[26] then talk of evil behaviour is seriously misleading.

Note that the current objections are importantly different than another objection that has been advanced against Korsgaard. Insofar as Korsgaard attempts to locate the metaphysical foundations of ethics in facts about what is constitutive of agency, she is plausibly regarded as an advocate of "constitutivism."[27] Different constitutivists appeal to different alleged facts about what is constitutive of agency; Korsgaard's own favoured version appeals to the supposed fact that action constitutively aims at self-constitution. A familiar and persistent challenge to constitutivism is predicated on the assumption that constitutivism's foundational standards can be normative only if they can be violated.[28] But if some standard cannot but be met, then it is difficult to see how it could be normative. Thus, the challenge for constitutivists is to explain how we can fail to meet those constitutive standards and still act, for example—a challenge that Korsgaard herself recognizes insofar as she worries that her view seems to imply that only good actions are actions (160). I am not presently raising *that* objection since I am willing to allow that Korsgaard can understand bad actions as bad in virtue of being deficient and thus that the constitutive standards of action to which she appeals are normative. Rather, the current complaint is that, even if Korsgaard can overcome this familiar challenge to constitutivism, she is committed to denying the very possibility of evil actions, an implausible result.

However, things are still worse. I contend that Korsgaard is not only committed to denying that evil actions exist; she is also committed to denying that evil people exist.

On Evil People

There is something of a consensus that it is a mistake to identify evil *people* and evil*doers*.[29] More plausibly, evil people are rightly regarded as evil in virtue of suffering from especially morally depraved characters, characters that distinguish them from the rest of us: just as evil actions are plausibly regarded

---

26 Card, *The Atrocity Paradigm*, p. 28.

27 I borrow this way of putting things from Matthew Silverstein in his "Teleology and Normativity," pp. 214-240.

28 This constraint is discussed in great detail in Douglas Lavin, "Practical Reason and the Possibility of Error," pp. 424-457.

29 Barry, *Evil and Moral Psychology*, p. 93; Card, *The Atrocity Paradigm*, p. 22; Formosa, "A Conception of Evil," p. 233; Eve Garrard, "Evil as an Explanatory Concept," p. 321; Daniel Haybron, "Moral Monsters and Saints," p. 279; Adam Morton, *On Evil*, p. 65; Luke Russell, "Dispositional Accounts of Evil Personhood," p. 232; Marcus Singer, "The Concept of Evil," p. 190.

as being the worst wrongs that people do,[30] evil people are plausibly regarded as being the morally worst sort of people.[31] But just as it appears that Korsgaard is committed to denying that there could be evil action, she is similarly committed to denying that there could be evil agents. Recall that, while K2) makes unification and guidance by principles of practical reason a prerequisite for action, K1) makes unification and guidance by principles of practical reason a prerequisite for agency. The problem is not that Korsgaard cannot allow for the possibility of bad agents. She can: just as the shoddy builder is a bad builder in virtue of being guided carelessly or inattentively or whatever by the principles constitutive of house building, the bad agent is bad in virtue of being guided carelessly or inattentively or whatever by the principles constitutive of practical reasoning. And, as noted above, Korsgaard can perfectly well make sense of the thought that bad agents suffer from morally flawed characters, given that evil people will necessarily be callous[32] and will consistently refuse to consider and attend to the consequences of their actions. But someone who is merely callous could be a much, much worse sort of person—say, someone who isn't merely badly guided by the principles constitutive of practical reasoning, but guided *perversely* by them or unguided by them *altogether*. To put it another way, the evil person is not evil merely in virtue of being negligent, even in virtue of being plagued by a strongly fixed disposition to act negligently, and someone who is merely callous could be a much, much worse sort of person and is thus wrongly regarded as evil, whatever else is wrong with her. She could also, or instead, be terribly cruel or unjust or malicious or otherwise suffer from some strain of moral vice that is at least as bad as callousness if not far, far worse. But, if the bad person has to be understood simply as being bad simply in virtue of being negligent, it is difficult to see how Korsgaard could make room for persons being evil *qua* being vicious in some other respect. More forcefully, if agents are not plausibly regarded as being evil in virtue of being merely negligent and if K1) implies that being guided by the principles of practical reason is necessary to be an agent at all, then there could be no evil agents.

Korsgaard could rest content with denying the possibility of evil agents, much in the way that Kant denied the possibility of a devilish person. Again: evil-scepticism about evil people is a live position. Interestingly, Korsgaard does seem to allow for the possibility of evil people. At one point, Korsgaard distinguishes two conceptions of the evil person. On one conception, the evil person is powerless and pathetic, lacking any standards and entirely without integrity. This is the sort of person who would sell out a friend for cheap, who would

30 Card, *The Atrocity Paradigm*, p. 28.

31 Barry, *Evil and Moral Psychology*, p. 16.

32 For this reason, it is odd that Raymond Critch objects that Korsgaard cannot make sense of people who are evil in virtue of their callousness. See Raymond Critch, "Principled Tyranny," pp. 277-287.

pursue any occurrent appetite at the expense of her own wellbeing, and is well exemplified by a multitude of losers like the junkie and the drunk in the gutter and the sociopathic hothead (170). This sort of person, unlike the dedicated and virtuous person, lacks unification. So understood, evil is weakness, the *privation* of strength and conviction, and Korsgaard dubs the corresponding conception the "privative conception of evil" (170). Privative conceptions of evil have a long and storied history in philosophical thought but it is probably a mistake to suppose that we can adequately conceptualize evil personhood solely in privative terms. Privative conceptions tell us what the evil person is *not* like—she isn't strong, she lacks standards, and so forth—but far too little about what she *is* like. If being evil is a matter of having an especially dubious sort of moral character, then any conception of evil personhood had better say something intelligible about what evil character *does* amount to.

The advocate of the privative conception might argue that her conception does implicitly suggest a positive account of evil character: insofar as the evil person lacks strength and conviction, she not only lacks moral virtues like justice and courage, but she suffers from some pretty obvious moral vices as well, especially those involving a lack of strength and conviction: for example, vices having to do with weakness and softness. And perhaps some putative evil people are weak and soft in just this way. Apparently, Charles Manson's wickedness sprung at least partly from being rejected by the music industry, Adolf Hitler's wickedness sprung at least partly from his failure as an artist, Jeffrey Dahmer's wickedness sprung at least partly from his inability to come to grips with his sexual orientation, and so forth. Perhaps, but surely we do not tend to regard Manson and Hitler and Dahmer as evil because they are akratic or incontinent, lest we be unable to distinguish them and their wrongdoings from a benevolent ne'er-do-well who is chronically unable to pull it together. Vices associated with weakness and softness are the *wrong* vices to appeal to if one is interested in illuminating evil character. So, and this is the second problem, privative conceptions of evil personhood are unable to characterize those evil persons who have strength and conviction in abundance and it is for this reason that Korsgaard rightly rejects the privative conception (170).

An alternative conception of evil personhood understands evil as kind of power, a conception that Korsgaard dubs the "positive conception of evil" (171). To explain what the evil person is like on this conception, Korsgaard returns to a discussion of souls. Insofar as the evil person is the morally worst sort of person, evil people must have the morally worst sort of soul, and, while there are any number of flawed souls—including timocratic, oligarchic, and democratic souls (175)—the worst sort of soul is the *tyrannical* soul. To explain, Korsgaard invokes the Platonic thought that an unjust city may be governed albeit governed by bad laws (161-162). Both the evil soul and the unjust city are unified, in *some* sense of the term, since they are governed. Just as the unjust city is not governed by just laws, an evil soul is not guided by moral principles

like the CI and the HI, but she *is guided* such that she is not the quivering, indecisive mess suggested by the privative conception. So understood, evil people really are agents because they chose "in accordance with the exercise of a principle by which the agent rules himself and under whose rule he is—in a sense—constitutionally unified" (175). They remain *evil* because the principle that governs them "is not reason's own" (175) but some lesser principle that lacks the moral status of the CI and the HI. In particular, the tyrannical soul "is governed by some nightmarish erotic desire," one that "subordinates the entire soul to its purposes leaving the person an absolute slave to a single dominating obsession" (169-170). This way of conceiving of evil people has some intuitive purchase insofar as it does capture part of what is so horrifying about some putative evil people: if only the ruthless agent of genocide was less unified! Were that the torture-murderer less resolved! There is no shortage of examples of putative evil people who are powerful and principled and thus the positive conception has some intuitive purchase.

Thus, Korsgaard seems to allow that there can be evil people, understood as being evil in virtue of suffering from a tyrannical soul. If so, then she is not an evil-sceptic—at least, not a sceptic about evil people.[33] Perhaps, but Korsgaard is committed to remaining an evil-sceptic—if not an evil-sceptic about evil *people*, she remains an evil-sceptic about evil *agents*. For Korsgaard is clear on this point: "The tyrannized soul can never separate himself from *one* of his impulses, and so consolidates himself into a mere a [*sic*] force of nature, an object, a thing" (173). If that's right, then an evil person possessed of a tyrannical soul just isn't going to meet the minimal conditions for agency demanded by K2). He is neither autonomous nor free (173). So even if, strictly speaking, Korsgaard does not rule out the possibility of evil people, she does rule out the possibility of evil agency.

This result is problematic on its own for at least two reasons. First, it is commonly thought that evil people deserve our strongest moral condemnation[34] which in turn suggests that evil people are especially blameworthy. But evil people are not agents if they are unguided by the principles constitutive of practical reason, given K1). And, if they are not agents, then they are not morally responsible agents and therefore they are not apt targets for the range of negative reactive attitudes that are appropriate for those persons unresponsive to our demands to show good will and refrain from showing ill will. Arguably, this result should be taken as a *reductio* of Korsgaard's view: it appears that, on Korsgaard's account, *evil is its own exempting* condition. If extraordinary

[33] Here, too, thanks to an anonymous reviewer for pushing me to clarify just what sort of evil-scepticism Korsgaard does endorse.

[34] Joel Feinberg, *Problems at the Roots of Law*, p. 129; Kekes, *The Roots of Evil*, p. 58; Russell, "Dispositional Accounts of Evil Personhood," p. 232; Singer, "The Concept of Evil," p. 190.

viciousness is not sufficient for being the most damnable sort of person, then what is? If being evil cannot make someone "blameworthy," then what can?[35] To be sure, some bit of behaviour need not be an action for us to hold someone morally responsible for it, especially when someone is responsible for the fact that her behaviour doesn't count as an action—say, because she failed to maintain her integrity and pull herself together. It is an interesting feature of Korsgaard's view that responsibility is, in general, responsibility for omission (175). But it is entirely unclear how or why we should hold persons morally responsible for their omissions if they don't qualify as agents in the first place. Thus, I submit, Korsgaard cannot make sense of the thought that evil people are especially deserving of moral condemnation.

Relatedly, and this is the second problem for Korsgaard's proposal, she cannot make sense of the equally plausible thought that evil people are the morally worst sort of people. Someone possessed of a tyrannical soul is arguably worse than someone possessed of an oligarchic or democratic soul, but lots of merely bad people will similarly suffer from a tyrannical soul yet fail to count as evil. For example, someone singularly possessed of a desire to bind, torture, and kill young women (like Dennis Rader, the BTK killer) is unable to separate himself from one of his impulses, and so consolidates himself into a mere force of nature, but so does an author singularly dedicated to finishing his manuscript or the artist who dedicates herself to her craft. The all-too-dedicated author and artist probably suffer from some serious vices; they arguably suffer from "wickedness of exclusion," given that they too subordinate all other moral concerns to their commitment to their particular project.[36] But, while flawed and vicious, such people are not plausibly regarded as being the worst sort of person—that is, as *evil*. If so, then Korsgaard cannot make sense of the thought that the evil person is qualitatively different and distinct from merely bad people, a result that is fatal to a theory of evil personhood.

So, even if Korsgaard can allow that there are evil people who fall short of being agents, the theory of evil personhood she proposes is seriously flawed insofar as it cannot make sense of the plausible theses that evil people are especially blameworthy and qualitatively worse than merely bad people. Korsgaard at times pursues one final strategy that demands attention. At times, Korsgaard seems to allow that the evil person is unified "in a sense" (175), perhaps in the sense that the full-blown agent is. As noted above, someone possessed of a tyrannical soul can choose "in accordance with the exercise of a principle" and is "in a sense—constitutionally unified" (175). Additionally, Korsgaard suggests that action "is an idea that admits of degrees" and that "an action chosen in a way that more successfully unifies and integrates its agents

[35] I borrow this way of putting things from Michael McKenna in his "The Limits of Evil and the Role of Moral Address," pp. 127-128.

[36] Stanley Benn, "Wickedness," p. 797.

is more authentically, more fully, an action, than one that does not" (25). But surely not in the sense of 'unified' that is operative in K1). To flesh out the Platonic analogy, Korsgaard slides between two different principles—the aforementioned K1) and:

> K1*): A person counts as an agent only if she is unified and she is unified only if she is guided by *some* principles.

K1*) allows Korsgaard to explain how tyrannical souls can act. They can act because they are governed by some principle of choice, just not reason's own. Fair enough, but endorsing K1*) requires abandoning any comprehensive conception of action and agency that builds on Kantian and Platonic themes. Once we allow that *mere* organization of the soul is needed for agency, and not aristocratic organization in particular, we can no longer agree with Plato that an unjust person is incapable of agency (159). Similarly, if we allow that other principles besides the HI and the CI can guide and unify a will, we allow that heteronomous action really is possible, *contra* Kant. Explaining evil agency remains out of the question so long as Korsgaard clings to K1) but while K1*) is more useful in that regard it lacks the venerable pedigree that Korsgaard finds desirable.

It is, in any case, pretty clear that Korsgaard doesn't really mean to entertain the plausibility of K1*) or anything like it: we are told that the "bad person is determined from the outside," a mere "conduit for forces working in him and through him" and thus that he is "enslaved" (161); we are told that "bad souls are mere heaps" with no "definite criteria of identity" (164) much in the way that a defective house is just a heap of mortar and bricks; we are told that the tyrant is not a force even if his desire is (171); we are told that the tyrant doesn't really choose "an act for the sake of an end" since there is but one end dictated by his obsessive desire that he is going to pursue no matter what (172); we are told that the tyrannical soul can never separate himself from his obsessive desire "and so consolidates himself into a mere a [*sic*] force of nature, an object, a thing" (173); we are told that persons lacking an aristocratic soul "fall apart" (180); and so forth. It is hard to see how evil people could be agents if they are mere conduits, mere heaps, mere objects.

Note too that, even if actions can come in degrees such that they can unify and constitute their agents to greater or lesser degrees, it is unclear that this conception of action will allow Korsgaard to endorse the popular and plausible thought that evil people are qualitatively worse than merely bad people. For if actions can unify and constitute their agent to a greater or lesser degree, then agency would be a quantitative function, not a qualitative one. But, again, evil people are not just somewhat worse than merely bad people and their characters aren't flawed to a greater degree: rather, evil people are an altogether worse sort of person than the merely bad person. Allowing that action and agency can come in degrees will not help here.

Conclusion

I contend that Korsgaard is committed to denying the very possibility of evil agents and evil actions. In an interesting passage, Korsgaard reflects on her own assertion that there is something puzzling about evil:

> … I say that evil is very puzzling "in one way," because in another way it isn't hard to understand evil at all. Speaking a little roughly: when we take up the first-person point of view of the agent, and imagine her making the choice with her eyes wide open, we cannot fathom how she could choose to be anything less than a unified, free, and effective agent. On the other hand, looking at her conduct from a third-person point of view, say, as the object of social scientific explanation, we may be able to see quite clearly why she did what she did. (164)

There might be something puzzling about evil even if we take up the third-person point of view: even if one person's evilness is explained by appealing to her environment, to her upbringing, to her genetics, and so forth, there remain a large number of people who are the product of similar stuff and don't wind up being evil. But no matter. To my ear, there is something puzzling about what Korsgaard says about the first-person point of view. I find it odd to imagine an agent who, with eyes wide open, chooses to be something other than unified and free and effective. Famously, of course, Satan in John Milton's *Paradise Lost* utters his infamous imperative "Evil, be thou my good"[37] which at least suggests the kind of voluntarism about evil character that Korsgaard has in mind here. But Milton's Satan is a literary fiction, after all, and whether any putative evil person in the actual world ever made such a conscious and effective choice is an open question.

There is a more compelling reason to doubt that evil demands making the deliberate choice to be evil that Korsgaard cannot fathom. Again, if evil people are the morally worst sort of people, then we should expect that they have the morally worst sort of characters—that is, deeply flawed characters plagued by especially grave moral vices. But consistent with the virtue ethics tradition, extremely grave vices should emerge only after some period of habituation and repetition and development has passed. Just as especially morally valuable virtues are not acquired and possessed to significant degrees simply by choosing to adopt them, extreme vices are similarly not acquired and possessed to significant degrees simply by choosing to adopt them. Note that Satan bids farewell to remorse *prior* to uttering his infamous imperative; at the moment that he utters it, he has acquired a character ravaged by envy and pride and despair and ambition.[38] He *already* suffers from a deeply and thoroughly depraved character. Satan's infamous imperative is not *the cause* of his evil but its *expression*.

---

37 Milton, *Paradise Lost*, p. 221.

38 Barry, *Evil and Moral Psychology*, p. 53.

It is somewhat odd that the prospects for a comprehensive account of action and agency turn on that account's ability to make sense of evil agents and evil people; they are exceptional instances of action and agency, after all. But, if an account of action and agency is going to be comprehensive, then it needs work even at the margins. My contention is that Korsgaard's bold proposal fails at the margins: it fails to explain how evil people could be agents and how evil-doing could amount to action.

References

Barry, Peter Brian
2011 *Evil and Moral Psychology*. New York: Routledge.

Benn, Stanley
1985 Wickedness. *Ethics* 95 (4): 795–810.

Bernstein, Richard
2005 *The Abuse of Evil: The Corruption of Politics and Religion Since 9/11*. Malden: Polity.

Calder, Todd
2013 "Is Evil Just Very Wrong?" *Philosophical Studies* 163 (1): 177–196.

Card, Claudia
2002 *The Atrocity Paradigm: A Theory of Evil*. Oxford: Oxford University Press.

Card, Claudia
2010 *Confronting Evils: Terrorism, Torture, Genocide*. Cambridge: Cambridge University Press.

Cole, Philip
2006 *The Myth of Evil*. Edinburgh: Edinburgh University Press.

Critch, Raymond
2013 "Principled Tyranny: Can Korsgaard Explain Evil Action?" *International Journal of Philosophical Studies* 21 (2): 277–287.

Davidson, Donald
1980 *Essays on Actions and Events*. Oxford: Oxford University Press.

Feinberg, Joel
2003 *Problems at the Roots of Law: Essays in Legal and Political Theory*. Oxford: Oxford University Press.

Formosa, Paul
2008 "A Conception of Evil." *Journal of Value Inquiry* 42 (2): 217–239.

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

Garrard, Eve
1998 "The Nature of Evil." *Philosophical Explorations: An International Journal of the Philosophy of Mind and Action* 1 (1): 43–60.

Haybron, Daniel
2002 "Moral Monsters and Saints." *The Monist* 85 (2): 260–284.

Kekes, John
2005 *The Roots of Evil*. Ithaca: Cornell University Press.

Korsgaard, Christine
2009 *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford University Press.
Korsgaard, Christine
1996 *The Sources of Normativity*. Cambridge: Cambridge University Press.
Kramer, Matthew
2011 *The Ethics of Capital Punishment: A Philosophical Investigation of Evil and Its Consequences*. Oxford: Oxford University Press.
Lavin, Douglas
2004 "Practical Reason and the Possibility of Error." *Ethics* 114 (3): 424–457.
McKenna, Michael
1998 "The Limits of Evil and the Role of Moral Address: A Defense of Strawsonian Compatibilism." *Journal of Ethics* 2 (2): 123–142.
Mele, Alfred
2003 *Motivation and Agency*. Oxford: Oxford University Press.
Midgley, Mary
2001 *Wickedness*. New York: Routledge.
Milton, John
1998 *Paradise Lost*, 2nd edition, Alastair Fowler, ed. New York: Longman.
Morton, Adam
2004 *On Evil*. New York: Routledge.
Russell, Luke
2006 "Evil-Revivalism Versus Evil-Skepticism." *Journal of Value Inquiry* 40 (1): 89–105.
Russell, Luke
2010 "Dispositional Accounts of Evil Personhood." *Philosophical Studies* 149 (2): 231–250.
Russell, Luke
2014 *Evil: A Philosophical Investigation*. Oxford: Oxford University Press.
Silverstein, Matthew
2016 "Teleology and Normativity." *Oxford Studies in Metaethics* Volume 11, Russ Shafer-Landau, ed. Oxford: Oxford University Press: 214–240.
Singer, Marcus
2004 "The Concept of Evil." *Philosophy* 79 (2): 185–214.