

ARTICLE

# Self-Knowledge and the Opacity Thesis in Kant's Doctrine of Virtue

Aaron Halper

Independent Scholar, US  
Email: [ahhalper@gmail.com](mailto:ahhalper@gmail.com)

## Abstract

Kant's moral philosophy both enjoins the acquisition of self-knowledge as a duty, and precludes certain forms of its acquisition via what has become known as the Opacity Thesis. This article looks at several recent attempts to solve this difficulty and argues that they are inadequate. I argue instead that the Opacity Thesis rules out only the knowledge that one has acted from genuine moral principles, but does not apply in cases of moral failure. The duty of moral self-knowledge applies therefore only to one's awareness of one's status as a moral being and to the knowledge of one's moral failings, both in particular actions and one's overall character failings, one's vices. This kind of knowledge is morally salutary as an aid to discovering one's individual moral weakness as well as the subjective ends for which one acts, and in this way for taking up the morally required end of treating human beings as human beings. In this way, moral self-knowledge can be understood as a necessary element of moral improvement, and I conclude by suggesting several ways to understand it thereby as genuinely primary among the duties to oneself.

**Keywords:** Kant; morality; virtue ethics; self-knowledge; opacity

Kant's Delphic Command to know oneself in the *Metaphysics of Morals* (*MM*, 6: 441)<sup>1</sup> appears to come into direct conflict with Kant's so-called Opacity Thesis, which limits the ability of moral agents to know their own maxims. As Owen Ware formulates it, according to the Opacity Thesis, 'the ground of my maxims lies beyond the reach of cognition. I am opaque to myself to the extent that I can never know my disposition immediately by way of introspection' (Ware 2009: 674). Kant thus simultaneously enjoins us to pursue moral self-knowledge as a duty while seemingly also denying us the possibility of carrying this out. The familiar Kantian thesis that 'ought implies can' implies that both of these cannot be true.

A number of solutions to this difficulty have been recently proposed, and I categorize and evaluate them in section 1. In section 2 I put forward my own solution, arguing that the Opacity Thesis only applies to genuine moral action, but not to moral failure. Moral self-knowledge, as I understand it, applies both to a recognition of oneself as a moral being under moral obligations, and to the knowledge of one's moral failings. It is this latter that Kant terms 'the hell of self-cognition' (*MM*, 6: 441).

In section 3 I discuss, in various ways, the utility of such self-knowledge for moral improvement. Finally, in section 4, I conclude by putting forward several suggestions as to why Kant describes moral self-knowledge as ‘the First Command of All Duties to Oneself’ (*MM*, 6: 441).

It may be helpful here to note Ware’s distinction between overt and covert restrictions on moral knowledge (Ware 2009: 674). Overt restrictions on moral knowledge are epistemic limits; covert restrictions stem from our own capacity for self-deception. I am arguing that there are overt restrictions that apply to cases of genuine moral action but that do not apply to non-moral action nor to moral failure. We would still be left, of course, with covert restrictions on knowledge of moral failings. Although from the point of view of the agent, covert restrictions likely constitute the greater threat, I focus my analysis throughout on overt restrictions not only because I think I have something important to say about them, but also because in my view properly understanding them may be helpful for dealing with the covert restrictions on self-knowledge. Indeed, I think this is so not only for Kant scholars working on relevant topics but also for moral agents who may be able to situate and contextualize their own moral self-deception in the context of the overt epistemological limits of moral self-knowledge.

### I. Two strategies

Kant describes moral self-knowledge as the first command of all duty to oneself: ‘This command is “*know* (scrutinize, fathom) yourself”, a familiar Delphic imperative that for Kant requires self-cognition

in terms of your moral perfection in relation to your duty. That is, know your heart — whether it is good or evil, whether the source of your actions is pure or impure, and what can be imputed to you as belonging originally to the *substance* of a human being or as derived (acquired or developed) and belonging to your moral *condition*.

Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one’s heart which are quite difficult to fathom, is the beginning of all human wisdom. For in the case of a human being, the ultimate wisdom, which consists in the harmony of a being’s will with its final end, requires him first to remove the obstacle within (an evil will actually present in him) and then to develop the original predisposition to a good will within him, which can never be lost. (Only the descent into the hell of self-cognition can pave the way to godliness.) (*MM*, 6: 441)

The question at issue here is how to understand this passage in harmony with the above-described Opacity Thesis, a thesis which Kant formulates in a number of different places, and of which the following passage is representative:

In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the

representation of one's duty. It is indeed sometimes the case that with the keenest self-examination we find nothing besides the moral ground of duty that could have been powerful enough to move us to this or that good action and to so great a sacrifice; but from this it cannot be inferred with certainty that no covert impulse of self-love, under the mere pretense of that idea, was not actually the real determining cause of the will; for we like to flatter ourselves by falsely attributing to ourselves a nobler motive, whereas in fact we can never, even by the most strenuous self-examination, get entirely behind our covert incentives, since, when moral worth is at issue, what counts is not actions, which one sees, but those inner principles of actions that one does not see. (*G*, 4: 407)

In recent years, two different kinds of strategies for overcoming this difficulty have been put forward in the literature. First, a number of scholars set out to weaken the duty of self-knowledge. Thus, Jeanine Grenberg puts forward what I will call the Aspirational View, according to which *full* self-knowledge is impossible and so we can have no duty to attain it. Rather, our duty is only constantly to *strive* to attain it (Grenberg 2005: 217–41). As Kant writes of the pursuit of moral perfection, 'It is a human being's duty to *strive* for this perfection, but not to *reach* it (in this life), and his compliance with this duty can, accordingly, consist only in continual progress' (*MM*, 6: 446). The view I present here can be understood as a version of the Aspirational View, for I seek to build on Grenberg's important work, in part by generalizing some of the points she makes in her work on Kantian humility. It must be noted, however, that a major obstacle to the Aspirational View is that it requires us to understand how it is possible to gain partial knowledge that one has acted from moral motivations, since it is hard to see how such knowledge can be divided. One has either acted from moral motivations, or one has not; and one either knows it or not. I will seek to overcome this difficulty by interpreting 'partial knowledge' as knowledge possible in some but not all situations, and I will clarify in which situations such knowledge is possible.

Ware seeks to limit the duty of self-knowledge by arguing that it does not require one to know the maxims of one's actions, but it involves a second-order claim regarding the sincerity of one's attempt to evaluate one's maxims. As Ware puts it, 'While the purity of my disposition, whether it is actually good or evil, is impossible to cognize directly, I am immediately conscious of whether I've examined my life conduct with due care' (Ware 2009: 695).<sup>2</sup> We are reminded here of the duty of honesty, which does not require that one always say what is true, but only that one always say what one believes to be true, i.e. that one be truthful.<sup>3</sup> Ware believes that because the Opacity Thesis refers only to first-order knowledge, such second-order knowledge does not run afoul of that thesis and is in fact attainable by an agent. A potential difficulty with Ware's view, as Ware is well aware, is that it is difficult to disentangle the question of 'whether I've examined my life conduct with due care' from the question of 'the purity of my disposition, whether it is actually good or evil'. It seems as though a property of a good disposition is that one who possesses it is able to examine her own actions with due care, so that if one has evidence that one has well examined her own actions, then one has evidence of the purity of one's disposition. This is a difficulty that can be well addressed, I submit, by my views of self-knowledge and of the Opacity Thesis.

Emer O'Hagan understands self-knowledge as concerned not with finding out 'what you care about', but as concerned merely with seeing 'what the moral law demands' in particular situations (O'Hagan 2009: 532).<sup>4</sup> In other words, for O'Hagan self-knowledge is just not concerned with discovering the maxims at the root of one's actions, but instead aims at using 'the theoretical foundation of Kantian ethics' as a 'kind of compass which allows the agent to locate herself in the moral landscape' (p. 533).

O'Hagan illustrates a failure of such self-knowledge with an example of an agent who in the abstract believes she has a duty to help others when they are struggling, but when she sees her colleague 'Ted as he tries to open his office door, balancing a load of books and steaming mug of coffee', notices that by helping him she will miss the elevator – 'and then suddenly I see that I have no such duty to Ted who, after all, would never go out of his way for me' (O'Hagan 2009: 525). O'Hagan points here to an important aspect of the failure of self-knowledge, namely, that it may not manifest to the agent herself as a failure of self-knowledge at all, but can instead take the form of a motivated insensitivity to the demands of the moral law in a particular circumstance. She thus rightly points out that recognizing the demands of the moral law in a particular circumstance is an important aspect of self-knowledge and that it is through such self-knowledge that one comes to realize that the right thing to do was to help Ted. But it is strange to cut this kind of self-knowledge off from a knowledge of what one cares about, as O'Hagan would have us do. It seems that the realization that one ought to have helped Ted contains as well the realization that one was being selfish, i.e. caring insufficiently about Ted's situation. At the very least, it seems only to be a small leap from the former to the latter.

O'Hagan herself thinks that the former alone cannot be the entirety of moral self-knowledge, and seems to think that there is also a role for self-evaluation to play in virtue, but she balks at attributing this to Kant, presumably because of the Opacity Thesis (O'Hagan 2009: 535ff.). But as I will show, the Opacity Thesis does not actually preclude such self-evaluation, and Kant's own language indicates strongly that he thinks it a vital part of self-knowledge.<sup>5</sup>

A second category of strategy seeks to weaken the limitations on moral self-knowledge. Anastasia Berg has recently attempted such a strategy. She argues that agents who genuinely perform moral action can indeed know that they have done so, and that therefore only a very weak version of the Opacity Thesis can be attributed to Kant, one which does not preclude an agent from knowing 'that the moral law is the ground of their will, when it indeed is' (Berg 2020: 5). This view is, however, explicitly contradicted by Kant who writes that 'in the case of any deed it remains hidden from the agent himself how much pure moral content there has been in his disposition' (*MM*, 6: 393; emphasis added). Berg argues for her view on the ground that moral action requires that one act from the representation of moral principles, and through a feeling of respect which requires 'immediate determination of the will by means of the law and consciousness of it' (*G*, 4: 404). Berg concludes from this that genuine moral action requires a consciousness that one has acted not merely in accordance with, but from, moral principles (p. 6), writing that 'the determination of the will by the moral law cannot therefore happen "behind the agent's back", so to speak' (p. 7). Arguments similar to Berg's have been advanced by Sven Bernecker and Dina Emundts in the context of putting forward what Ware has titled an Inferential view of self-knowledge.<sup>6</sup> As Bernecker phrases it, 'gehört es zum Begriff

der praktischen Rationalität, daß man weiß, warum man so handelt, wie man handelt' (It belongs to the concept of practical reason that one knows *why* one acts *the way* one acts': Bernecker 2006: 175; translation mine). See also Emundts' claim that 'The knowledge of one's own maxims seems to be unproblematic for Kant. This is so because an intentional action implies knowledge of one's maxims, and this knowledge is not part of an extra consideration or act' (Emundts 2017: 192).

A difficulty with this view is that it seems to consider the awareness of a feeling of respect to suffice, at least in ordinary cases, for one to be able to know that one has acted from the moral law. As Berg puts it, 'in the primary case for a moral agent to cognize the law with respect just is to know, in some sense, one's will as determined by the law' (Berg 2020: 6; see also Bernecker 2006: 176; Emundts 2017: 194). But it cannot be legitimate to infer from the fact that one has felt respect for the moral law while or before one has acted to the claim that one has acted *from* respect for the moral law. Not only do Kant's own formulations of the Opacity Thesis rule this out (e.g. the above cited G, 4: 407), but Kant holds that merely reflecting on the moral law or on deeds done in accord with and from the moral law suffices to awaken one's feeling of respect (see e.g. *CPrR*, 158ff.). It is thus perfectly plausible that one's actions and one's feeling of respect are causally independent. Nonetheless, as we will see, one's own actions are surely a useful part of the acquisition of moral self-knowledge.

Put another way, it is one thing to act while fully conscious of the relevant moral principles, and quite another to be conscious that those principles are indeed the source of one's action. For acting from the representation of the moral law and being aware that one is acting from the representation of the moral law are two different things, and Kant thinks we ought (and hence can) do the former while the latter he steadfastly denies. For self-opacity is not the result of an opacity concerning what the moral law requires (Berg 2020: 5) but an opacity about the causal connection between one's representation of the moral law and one's action. Berg argues that we must be aware of this causal connection because we are self-reflective beings in the sense that Kant points out in the B Deduction of the first *Critique*, so that 'for any representation *p*, I am able to make my representing *p* explicit to myself in *I think p*' (p. 6). But this point does not suffice, for even if we do act *in* consciousness of the moral law, this does not guarantee that we have acted *from* that consciousness.

A second difficulty with Berg's view is that it leads her to attribute to Kant a rejection of the very self-examination that the duty of self-knowledge enjoins. According to Berg, when an agent performs moral action, self-examination is unnecessary; when an agent performs immoral action, self-examination is pointless and pathological (Berg 2020: 14). Berg thus reads Kant's description of any attempt at self-examination as a 'descent into the hell of self-cognition' to be a warning against making such an attempt. But Berg omits the continuation of the passage, where Kant clarifies that such a descent, though hellish, nevertheless is the 'only' way that one can 'pave the way toward godliness' (*MM*, 6: 441). Kant clearly thinks that self-examination is both helpful and necessary for moral improvement, even if he does express misgivings about its hellishness. This is, I take it, the point of the duty of self-knowledge, but it requires an account that does not view self-examination as purely pathological.

## 2. An alternative approach

My own view combines aspects of both strategies. I follow the second strategy by arguing that the Opacity Thesis does not apply to cases where one has failed to perform genuine moral action, i.e. moral action performed from moral principles. This is pretty much the opposite of Berg's thesis. I follow as well Grenberg's Aspirational View because, for reasons that I make clear below, I do not consider this knowledge of one's moral failings to be completable. Instead, I understand it to be a part of the never-ending striving for moral perfection against the evil principle within oneself, a striving that is characteristic of the good life.

I argue that the Opacity Thesis does not apply to cases of moral failure for two reasons. First, Kant's formulations of the Opacity Thesis support this view. Consider the above cited statement of the Opacity Thesis at G, 4: 407: 'it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty'. Kant's point here is not that our own motivations are opaque to ourselves in every case, but only that we cannot be sure in the very specific case of genuine moral action. Indeed, I am not aware of any passage where Kant formulates a version of the Opacity Thesis without including this qualification limiting its scope to ruling out only the knowledge of whether one has acted from moral principles.<sup>7</sup> To my knowledge, in all of his formulations of the Opacity Thesis, Kant applies it only to limit our knowledge of whether genuine moral action has been performed, and never to limit our knowledge in cases where it has not.

Second, this weakening of the Opacity Thesis makes sense given what I take to be Kant's reasoning for putting it forward in the first place. Kant employs the Opacity Thesis because genuine moral action requires that one perform sensible actions from noumenal moral principles, but such causation is always opaque to the human mind, 'for there is no theory about the causal relation of the intelligible to the sensible' (MM, 6: 438; see also G, 4: 460ff.). Our inability to cognize this kind of causation necessarily implies our inability to subsume any particular action under this kind of causality. But the *failure* to perform genuine moral action requires no mysterious theory of causation from the noumenal to the phenomenal. As such, there is good reason to consider that the Opacity Thesis applies only to cases of genuine moral action.

In common with Berg's account, this view of the Opacity Thesis implies a certain epistemological one-sidedness to knowledge of our motivations (Berg 2020: 8–9). But where Berg thinks we can know our motivations only in the case of genuine moral action, I think we can know our motivations only in cases where the action is not genuinely moral. This asymmetry makes sense because Kant's claim that 'no certain example can be cited of the disposition to act from pure duty' (G, 4: 406) by no means implies that we cannot find certain examples of immoral or of non-moral actions. An implication of this asymmetry is that in a case where we have successfully performed genuine moral action, we cannot know that we have done so but must be left to wonder about it. After all, there are surely cases where we think we have acted morally and out of moral motivations, only to discover on reflection that we have not done so. Any action that we have performed that appears to ourselves to

have been genuinely moral could still turn out not to have been morally motivated; perhaps still more reflection is needed.

Although the Opacity Thesis thus rules out our ability to know our own moral successes, there are still two kinds of moral self-knowledge that it does not rule out, both of which Kant sums up in the pithy phrase: ‘know your heart – whether it is good or evil’ (*MM*, 6: 441). One can know one’s heart is good insofar as one is able to recognize that one is a ‘human being worthy of respect’ (*MM*, 6: 441) who has moral duties that one ought to discharge. One can know one’s heart is evil insofar as one is able to discover in what ways one has failed to discharge those very duties.<sup>8</sup> My account of the proper objects of the Opacity Thesis implies that the duty of self-knowledge must apply, abstractly, only to one’s status as a moral being who ought to fulfil the moral law and, concretely, to one’s specific moral failings.

Now, Ware argues that the former aspect of self-knowledge, concerning one’s awareness of one’s own moral status, cannot ‘meet the requirements of a duty’ but merely ‘functions negatively to counteract my self-loathing and misanthropic attitudes, and positively to shake off the lull of apathy’ (Ware 2009: 681). He argues that knowledge of one’s own status as a moral being cannot be a duty because it is too easy, being ‘intellectually obvious – even to scoundrels’ (p. 679). But Kant does not say that such self-awareness is easy for the scoundrel to attain, only that it is possible even for the most debased and hardened scoundrels to attain it (*G*, 4: 454). Since ought implies can, the fact that such knowledge is universally attainable is an argument in favour of, and not against, its attainment being a duty. In practice, however, I think Ware is correct that the duty of self-knowledge involves continual self-reflection on one’s own motivations in every concrete action, but I consider this self-reflection to be done properly only when it is done along with, or in the context of, an attention paid to one’s own status as a moral being, which can function as the sort of Kantian moral compass that O’Hagan describes.

My view of self-knowledge and of the Opacity Thesis can aid us, I suggest, in dealing with the difficulty I raised when discussing Ware’s view in the previous section. This was the difficulty that on Kant’s account our conscience seems to be able to give us unerring knowledge into ‘whether I have consciously examined my duties’ (Ware 2009: 692) even though I am still not able to assess the purity of my disposition. It is difficult to see how determining the former would not aid in ascertaining the latter. My suggestion is that conscience for Kant only comes into play in cases of moral failure, or at least of presumed moral failure, in intention or in deed. For conscience begins with an accusation against oneself (*MM*, 6: 438), that one has failed to live up to one’s moral principles, and proceeds to provide a ‘verdict’, either ‘acquitting or condemning’ (6: 440). Even in the case of acquittal, Kant is clear, one does not thereby provide oneself with a proof of the purity of one’s disposition, but one may only now rejoice ‘at having escaped the danger of being found punishable’ (*MM*, 6: 438). Even if conscience is infallible, its pronouncements cannot run afoul of the Opacity Thesis because conscience never pretends to establish that one has performed genuine moral action.

Thus understood, while the Opacity Thesis turns out to express no limitation on the scope of the duty of self-knowledge, it still expresses a limitation on our capacity for self-knowledge. For, as we just saw, we can never be sure that any of our actions



were performed out of moral motivation. As such, it turns out that all of our actions, past and present, are open for examination, just in case they turn out to be improperly motivated. Indeed, even our own attempts to perform such examinations must in principle be themselves open to being thus examined. We must especially suspect our best moments, for these we can be the least certain of, and they too may be no more than instances of self-deception. Here my view follows the above-discussed Aspirational View of Grenberg, according to which ‘the pursuit of the duty of self-knowledge is an ongoing and never-completed process’ (Grenberg 2005: 233).<sup>9</sup> When first describing the Aspirational View, I noted that its primary difficulty is that it must explain what it means to acquire self-knowledge partially, and this I believe I have done. Moral self-knowledge cannot take the form of a complete accounting of one’s moral character, but always concerns only the knowledge of one’s status as a moral being along with the piecemeal recognition of one’s specific moral failures and, as we will see below, specific vices. The fact that moral self-knowledge is an unending pursuit may be one reason that Kant describes its pursuit as a ‘descent into the hell of self-cognition’ (*MM*, 6: 441).

It is not unlikely that Kant uses this unattractive description of the pursuit of moral self-knowledge precisely because it requires uncovering our failings and bringing them into full view in an attempt to overcome the covert restrictions on such knowledge.<sup>10</sup> In this task, inferring our own character from our actions is only of help in limited cases.<sup>11</sup> For from our immoral actions we can safely infer that we acted on some kind of immoral motivation, but it may often be the case that we should be genuinely unsure which kind of immoral motivation we have acted upon. From our moral actions, any inference to our character must be on even less sure ground. The only means to knowing our heart is through examining it, through an active and deliberate practice of introspection. In this, we are bound to discover all manner of unsavoury and disgusting things about ourselves. What is more, the moral failings which we have to discover are precisely those which we have gone to the trouble of deceiving ourselves about, and are therefore precisely the moral failings which we do not wish to confront.<sup>12</sup> As O’Hagan points out so well, one’s own moral failings rarely show up to oneself as such; indeed, they often appear to oneself to be virtues. Realizing that one’s moral failings are indeed moral failings may be hellishly difficult and painful, but at least it does not represent an epistemic impossibility.

It may be worth noting here that the two components of moral self-knowledge that I outlined above are intimately related to one another. The moral self-knowledge that involves knowledge of one’s moral failings cannot be wholly separated from the knowledge that one is a moral being with inner obligations. For it is only by means of such general moral knowledge that one can judge oneself deficient in action or intention.

In this section I have presented and argued for a limitation on the scope of the Opacity Thesis and a corresponding limitation on our capacity for self-knowledge. With these points in hand, we can see that the Opacity Thesis and the duty of self-knowledge do not conflict. For the Opacity Thesis limits only our ability to know our moral successes, and the duty of self-knowledge concerns only our status as a moral being and our never-ending attempt to know our moral failings.



### 3. The worth of moral self-knowledge

In this section I address the question of the benefit of moral self-knowledge as I presented it in the previous section: in what ways is it morally salutary to know one's own failings and to recognize one's own status as a moral being? Now, one possible answer to this question can be ruled out immediately. The purpose of moral self-knowledge cannot be to denigrate oneself for one's misdeeds. For such a response to the discovery of one's failings, Kant assures us, is 'a self-contradiction' since it is only insofar as one recognizes that 'noble predisposition to the good in us, which makes the human being worthy of respect, that one can find one who acts contrary to it contemptible' (*MM*, 6: 441). The self-contradiction here, it is worth noting, is precisely a contradiction between the two forms of moral self-knowledge that I outlined above. Accordingly, Kant explicitly seeks to discourage self-loathing and self-punishment for one's past misdeeds, as encouraging 'secret hatred for virtue's command', though he does see a role for moral repentance 'with a view to improving' (6: 485). The point of knowing one's own failings cannot be self-rebuke, but must serve as an aid in moral self-improvement. But it is still unclear how such knowledge is necessary or even helpful for such self-improvement.

#### 3.1 Moral self-knowledge and virtue

In order to answer the above question regarding the utility of moral self-knowledge it may be helpful first to flesh out moral self-knowledge in terms of three different ways Kant discusses virtue in the Doctrine of Virtue: as moral strength, as moral health and as taking up ends that are also duties. I discuss each of these in turn in order to show that it is at least plausible that moral self-knowledge, as I have described it, can be understood to be beneficial in the relevant respects.

The first way Kant describes virtue is as possession of the 'capacity and considered resolve to withstand [the] strong but unjust opponent' which 'opposes the moral disposition *within us*' (*MM*, 6: 380).<sup>13</sup> According to this description, to be virtuous is to be able to constrain oneself to be moral in the face of competing sensible inclinations. This conception of virtue entails a corresponding view of vice according to which vices tend to undermine our ability to control and direct ourselves. The key notion here is of the unification of two conflicting selves: the noumenal self which '*imposes obligation*' (6: 417) upon oneself, and the 'same human being thought in terms of his *personality*' who is capable of being 'put under obligation' (6: 418). I understand this reference to 'personality' to be a reference not to ourselves merely as *homo phaenomenon*, for considered merely phenomenally the human being is not thought of as under moral obligation at all. The sought-for harmony here is rather between the *homo noumenon* who gives the moral law to himself and 'the human being as a sensible being endowed with reason' (6: 439, fn.). For the latter is made up, somehow, of both the noumenal and the phenomenal. To be virtuous in this sense is thus to align oneself as a combination of the noumenal and the phenomenal with the obligation imposed by oneself as *homo noumenon*.<sup>14</sup> Kant thus characterizes vices as 'hostile to our existence and proper functioning as humanly embodied rational beings' so that, because they sever the connection between these two notions of self, the 'vices of lying, avarice, and servility express disregard for one's dignity and subjection of one's choice to empirical desire' (Denis 2010: 177).

On this conception of virtue and vice, moral self-knowledge concerns the discovery of one's temptations and moral weaknesses, which we recognize as such by means of awareness of our status as beings who ought to do better. In uncovering particular moral failings, say that one has been stingy with one's money, one discovers tendencies towards particular kinds of moral failings, say miserliness. One thereby realizes the particular ways in which one finds oneself seemingly unable to control oneself, thus instances of failure to unify oneself with the moral law.

The second, related way Kant discusses virtue is in terms of moral health (*MM*, 6: 384) or moral integrity (6: 388). To be virtuous, on this conception, is to be a morally healthy being. Vice is an instantiation of an opposite of moral health, a kind of moral illness (6: 384). Moral self-knowledge concerns knowledge of one's moral health or lack thereof. As in the previous discussion, the morally healthy person is able to withstand temptations and to control herself, while the morally unhealthy person is not. The advantage of this conception of virtue over that in terms of the notion of strength is that it allows us better to gather up and sort out our particular moral failings. Thus one may recognize not only that one has acted in a miserly fashion and may be likely to do so in the future, but that one has a general problem with miserliness, a kind of 'moral disease' as it were, and thus that one is a miser. But just as a body is not its illness, so too the agent who discovers that she is a miser has only uncovered an 'illness' that she has, and should consider her very recognition of this vice within herself as a reminder of her status as a moral being.

The notion of moral health may also be broader than that of morality, a point that has been well developed by Jens Timmermann in his paper on indirect duties (Timmermann 2006). Indirect duties are not necessarily moral duties, but are duties that concern one's moral health. Timmermann cites the indirect duty of promoting one's own happiness in the *Groundwork*, a duty which is not moral and may even conflict with genuine moral duties. Nonetheless, one's own happiness is of moral concern because being unhappy makes one more likely to submit to temptations, so that 'making sure that you are happy is *indirectly* a good thing, morally, after all' (p. 298). Moral health is thus wider than morality, proscribing and prescribing a wider range of actions. This fits in well with Kant's extension of the notion of vice in the Doctrine of Virtue to encompass actions that are not directly immoral. For example, Kant prohibits the destruction of beautiful inanimate nature on the grounds that it weakens the disposition 'to love something (e.g., beautiful crystal formations, the indescribable beauty of plants) even apart from any intention to use it'; for although this disposition is 'not of itself moral' it still 'greatly promotes morality or at least prepares the way for it' (*MM*, 6: 443). It is not necessarily immoral to destroy beautiful things, but a concern for one's own moral health nonetheless prescribes against it, and this suffices for such action to be held vicious. Accordingly, a further advantage of the notion of moral health may be that consideration of such indirect duties allows for a kind of self-knowledge along the same lines as moral self-knowledge.

The third way that Kant discusses virtue is as concerned with ends that are also duties, 'ends we *ought* to set for ourselves' (*MM*, 6: 382). These ends are one's own moral perfection and the happiness of others. Duties to oneself, such as the duty of self-knowledge, concern only the former, and it is with these that I am here concerned. But considered as ends, one's moral perfection and the happiness of others

share a common feature, namely, that for an agent to take them up is 'to make the human being as such his end' (6: 395). According to this third way of considering virtue, to be virtuous is to take up the end of treating human beings as human beings. To be vicious is, correspondingly, not to take up this end but to take up instead ends that are beneath the inherent dignity of human beings.<sup>15</sup> Kant thus characterizes certain vices as failures to take up the end of humanity as such in favour of ends that are morally deficient. For example, with regard to money matters, Kant identifies two vices, prodigality which seeks 'to get and maintain all the means to good living with the intention of enjoyment' and stinginess which seeks 'to acquire as well as maintain all the means to good living, but with no intention of enjoyment (i.e., in such a way that one's end is only possession, not enjoyment)' (6: 432). These are both manifestations of replacing moral ends with non-moral ends, either possession or enjoyment, both of which are morally deficient compared to the end of humanity as such.

On this conception of virtue, moral self-knowledge concerns the discovery of one's true ends. Now, taking up a subjective end, such as humanity as such, appears at first glance to be readily accomplishable as if one could simply decide to treat human beings as human beings. But there is good reason to think that knowing one's true ends would entail knowing the relevant maxims one is operating from in the pursuit of those ends. If that is the case, then the Opacity Thesis would preclude us from knowing if we have indeed taken up the subjective end of humanity as such. We must instead always wonder if we have indeed adopted the end that we ought to have, or if instead we have adopted some other end. This point has been well argued by Mavis Biss who writes, 'We do not first adopt a moral end and then seek to accumulate instances of action on maxims containing that end. Rather, as we strive to make progress toward moral perfection we strive to develop a strong moral character, understood in terms of firm and deep commitment to moral ends' (Biss 2017: 629). Taking up the end of humanity as such, and thus becoming morally healthy, cannot be accomplished all at once; rather, it is something we must always strive to do. A vital piece of such striving is the cognitive task that is the pursuit of moral self-knowledge.<sup>16</sup> In uncovering one's particular failings, say that one has been stingy with one's money, one discovers thereby that one has not been acting from the end of humanity as such but from the end of possession. One thereby discovers the actual ends which one has been acting from, at least insofar as one has not been acting from moral ends.<sup>17</sup>

In my view, then, moral self-knowledge for Kant is indeed a form of introspection, one that concerns the knowledge of one's moral failings and thereby one's vices. It provides insight into one's moral weaknesses, one's overall moral health and one's true subjective ends. An agent may come to realize that she is not simply relaxed, but lazy. Or not simply willing to let others live their own lives, but unsympathetic. She may come to realize that she is racist, or is cruel to her supposed inferiors, or what have you. She comes to these realizations by examining her specific moral failings, both as they manifest in vicious actions, but also as they manifest in certain insensitivities, as O'Hagan has described. Consider again O'Hagan's example of the agent who believes that she has a duty to help others until an occasion calls for it that involves a slight inconvenience to herself, at which point she convinces herself that this duty does not apply in this particular situation. Such an agent ought to reflect on this action, either after the fact or, better, immediately, and realize that her motivated

insensitivity in this particular situation implies that she had not fully made the end of humanity as such her own subjective end, that she is not as morally healthy as she would like, that she has a moral weakness. These are the sorts of realizations that constitute the 'hell of self-cognition'. To face up to them is to face the abyss within oneself and to realize that one has real shortcomings, which are in fact manifestations of one's propensity towards evil (*R*, 6: 31).

### 3.2 Moral self-knowledge and moral self-improvement

As Kant describes it, moral self-knowledge is a part of practical philosophy and not, like other forms of knowledge, for the sake of theoretical considerations. Moral self-knowledge is practical, and thus meant to be applied for the sake of improving one's character. The insight one gains through this introspection is, I take it, a condition for the possibility of its application for the sake of improving one's character. Insights into one's moral failings and vices are thus crucial components of the good life, one dedicated to the struggle against the evil principle within oneself.

For Kant, moral self-knowledge as I described it above is necessary for the pursuit of self-perfection in a moral sense. We can see this by noticing that Kant describes the aim of self-perfection as holiness (*MM*, 6: 446) and that it is, again, 'Only the descent into the hell of self-cognition [that] can pave the way to godliness' (6: 441; emphasis added). Putting these points together we find that moral self-knowledge is a means to moral self-perfection, and the inclusion of the word 'only' implies that moral self-knowledge is indeed a necessary condition for self-perfection. Biss wonderfully fleshes out the role of moral self-knowledge in striving after moral self-perfection 'as ongoing engagement in a process of contemplative activity, complexly related to self-knowledge and rooted in ordinary moral life', within which 'moral self-knowledge seems to combat both false narratives about one's own moral perfection and attendant careless action on bad maxims' (Biss 2015: 19, 7). More than that, though, Biss draws out the point that moral self-knowledge involves an attentiveness to our experiences of moral conflict, through which 'the agent strengthens to some extent his consciousness of the supremacy of the claims of reason to those of self-love. Moral contemplation brings the striving agent into a state of anxious aspiration, wherein he exercises his moral powers even in questioning their adequacy' (2015: 17).

Kant's point, again, is not that we should discover our moral failings for the sake of self-rebuke, but that the painful self-knowledge of the idiosyncratic obstacles within oneself that prevent one from genuine moral action is the first, necessary step to overcoming them. Moral self-knowledge is thus 'the beginning of all human wisdom' through which we can get closer to 'the ultimate wisdom which consists in the harmony of a being's will with its final end, requires him first to remove the obstacle within (an evil will actually present in him) and then to develop the original predisposition to a good will within him, which can never be lost' (*MM*, 6: 441). For moral self-knowledge concerns recognizing and thereby making it at least possible to begin to overcome the obstacles to virtue within oneself, and thus to begin taking up for oneself the ends that are also duties.

At the same time, Kant is not claiming that knowledge of one's misdeeds in any way suffices to erase them. This point has been well discussed by Richard Moran, who considers the example of a 'rakehell' who commits adultery and subsequently

oscillates, alternately feeling ashamed of himself and good about himself for thus censuring himself (Moran 2001: 174ff.). As Moran points out, the rakehell's oscillation between these two attitudes seems cheap, rather more like an attempt to avoid facing up to his own shame than a genuine moral commitment. In part, this is because the rakehell switches between the two feelings, rather than realizing that the feeling of confidence cannot override his shame because it logically 'depends on his understanding his original shame as perfectly appropriate and called for' (p. 177). Now, O'Hagan considers the rakehell's problem here to be that he is too successful at self-evaluation. Since the rakehell knows himself to be a rakehell, 'he excuses himself and treats his freedom (to continue to be a rakehell) as a kind of determination (*I'm just like that*)' (O'Hagan 2009: 533). O'Hagan's point is, again, that moral self-knowledge does not require such self-evaluation, but rather an evaluation of the requirements of morality. But Moran's analysis is spot-on here. The rakehell has not successfully evaluated himself at all, but contradicts himself in thought as well as in deed. The rakehell's failure to evaluate himself in no way implies that proper self-evaluation is unnecessary or unhelpful, but merely showcases a possible misunderstanding of the function of moral self-knowledge. The point of self-scrutiny is that one may thereby uncover features within oneself that one reviles, and simultaneously come upon the fact that one has within oneself an incomparable dignity on the basis of which one reviles those very features. An agent can realize that she has not taken up the end of humanity as such, but realize precisely thereby that she can take up that end, and that she is the kind of being for whom only such an end is fitting. It is this combination that keeps her both from excessive self-contempt and from excessive self-esteem (*MM*, 6: 441). In this way, the realization of one's vices makes possible both self-contempt on the basis of those vices as well as self-esteem on the basis of one's moral dignity, without contradiction. These allow the agent simultaneously to realize her own worth as a being who should be treated as an end in herself, while also keeping in mind her ever-present need to act from a full sense of that worth. One thereby comes closer to realizing for oneself 'the harmony' of one's 'will with [one's] final end' (*MM*, 6: 441).

#### 4. Conclusion: moral self-knowledge as the first duty to oneself

In this article I have laid out an account of the duty of moral self-knowledge according to which it does not contradict the Opacity Thesis. I have shown how Kant's claims about moral self-opacity still allow for an agent to attain a great deal of insight into her own character in a way that is useful for moral development. I want to conclude with some suggestions concerning Kant's mysterious claim that the duty of self-knowledge is the First Command of all duties to oneself (*MM*, 6: 441). As such, it must have some relation of priority to the duties to oneself that come after it. But in what sense is this duty prior to the other duties to oneself? Now, Kant's discussion of duties to oneself contains only scant references to virtues but quite a broad discussion of vices. This is already an indication that Kant regards moral failure as of broadly significant concern for the individual moral agent who must seek to avoid vice.

Lara Denis argues persuasively that duties to oneself, and perfect duties to oneself in particular, have a primacy over duties to others in a number of senses. They have: 1) a rational primacy in that they follow 'most immediately from the basic rational

requirement that one's willing not contradict itself'; 2) a practical primacy in that their violation 'undermine[s] one's recognition of oneself as a bearer of humanity and subject of duty more profoundly, pervasively, and easily than do disrespect and maltreatment from others'; 3) a hierarchical primacy in that 'the matter of all moral requirements can be thought of as duties to oneself in a broad or formal sense'; 4) a logical primacy in that 'if there were no duties to oneself [then] we would not be bound to any duty'; 5) and a constitutional primacy in that 'duties to oneself (in a broad or formal sense) help constitute all duties' (Denis 2010: 182ff.). Now, Kant's description of moral self-knowledge as 'the First Command of all Duties to Oneself' implies that the duty of self-knowledge has some kind of primacy over the other duties to oneself. But in what sense is moral self-knowledge primary?

I want to suggest that the view of moral self-knowledge that I have presented may help us better understand its primacy over other duties to oneself, and in particular in the various senses of primacy that Denis identifies that duties to oneself have over duties to others. 1) Moral self-knowledge may have a rational primacy in that the identification of one's moral failings and the recognition of the status of one's moral health is of the highest concern to a rational but embodied being concerned to discharge her duty but vividly aware that she does not always do so. For such an agent has and must always have the moral status of her actions and character as a primary concern. 2) Moral self-knowledge may have a practical primacy in that a lack of concern with one's vices necessarily demonstrates a fundamental lack of concern with the ways that one has failed to perform genuine moral action, and thereby a lack of concern with oneself as a moral being. 3) Moral self-knowledge may have a hierarchical primacy in that, with regard to duties to oneself, every moral failing and every vice, of any sort whatsoever, comes under the purview of moral self-knowledge as it manifests as a particular accession to temptation and a particular end which has been acted from over and above that of humanity as such. 4) Moral self-knowledge may have a logical primacy in that, if we could not know of a particular moral failing, then we would not have been able to avoid it except through chance, and hence could not have been obligated to avoid it at all. 5) Finally, Kant himself seems to indicate that moral self-knowledge may have a constitutional primacy, using it to explain humility, confidence, impartiality and sincerity as 'duties to oneself that follow directly from this first command to cognize oneself' (*MM*, 6: 441–2). The implication may be that other duties to oneself follow in some fashion from this first command, and at least in some cases might even be understood as a particular determination or specification of the duty of moral self-knowledge.

**Acknowledgements.** A shortened version of this article was presented at the Eastern APA in 2022. I am grateful to my commentators, Sherry Kao and Mavis Biss, whose thoughtful engagement with my work has helped this article a great deal. I also wish to thank an anonymous reviewer at the *European Journal of Philosophy*, and anonymous reviewers at the *Kantian Review*, for the same reason.

## Notes

1 Citations of Kant's texts are from the Cambridge edition of his works. I have made use of the following standard abbreviations: *Metaphysics of Morals* (*MM*), *Critique of Practical Reason* (*CPrR*), *Groundwork of the Metaphysics of Morals* (*G*) and *Religion within the Boundaries of Mere Reason* (*R*).

2 A similar claim has recently been argued by Knappik and Mayr 2019.

3 See Mahon 2003.

4 A similar view has also been put forward by Melissa Merritt, who views self-knowledge as concerned not with introspection but with developing a skill for ‘taking the appropriate interest in one’s own capacity to think about how things are and what is worth doing ... by considering the objective questions about how things are and what is worth doing in the right spirit’ (Merritt 2018: 181).

5 A quite different attempt at weakening the duty of self-knowledge has also been recently put forward by Vivek Radhakrishnan, who argues that one can only have knowledge of one’s phenomenal self and never of one’s noumenal self; accordingly, the duty of self-knowledge must be a duty only to know one’s phenomenal self. This view unfortunately takes a limitation of theoretical philosophy and imports it wholesale into moral philosophy, despite Kant’s insistence that this limitation does not apply to moral philosophy (see especially *CPpR*, 5: 6–7). In the moral realm, agents are not restricted to empirical knowledge, but are aware of themselves as possessing moral capacities and intentions. What is more, according to Radhakrishnan, the contents of self-knowledge concern a knowledge about ‘one’s own moral disposition[, which] involves attribution of a moral value to one’s choice of maxims in the practical realm’ (Radhakrishnan 2019: 32). But this clearly involves a knowledge of oneself as noumenally free, and cannot be taken to be knowledge of pure phenomenality.

6 See Ware 2009: §2.1.

7 Consider again Kant’s claim that ‘In the case of any deed it remains hidden from the agent himself *how much pure moral content* there has been in his disposition’ (*MM*, 6: 393; emphasis added). As well, his claim that ‘The depths of the human heart are unfathomable. Who knows himself well enough to say, *when he feels the incentive to fulfill his duty*, whether it proceeds entirely from the representation of the law or whether there are not many other sensible impulses contributing to it that look to one’s advantage (or to avoiding what is detrimental) and that, in other circumstances, could just as well serve vice?’ (6: 447; emphasis added). And in the *Religion*, Kant claims that it is the ‘the transformation of the disposition of an evil human being into the disposition of a good human being’ whose assurance ‘cannot of course be attained by the human being naturally, neither via immediate consciousness nor via the evidence of the life he has hitherto led, for the depths of his own heart (the subjective first ground of his maxims) are to him inscrutable’ (*R*, 6: 51). Perhaps the most difficult formulation of the Opacity Thesis, as pointed out to me by Sherry Kao, occurs at *R*, 6: 63, where Kant states that ‘Indeed, even a human being’s inner experience of himself does not allow him so to fathom the depths of his heart as to be able to attain, through self-observation, an entirely reliable cognition of the basis of the maxims which he professes, and of their purity and stability.’ But even here Kant is concerned with ruling out the knowledge of the ‘purity and stability’ of our maxims, and the context makes it clear that ‘the maxims which he professes’ are presumed to be moral maxims. In all of these passages Kant is clear that what is unknown to us is not the ground of all of our maxims, but only whether we have indeed acted from our moral maxims.

8 A similarly dual conception of self-knowledge can also be found in Grenberg, although again she is primarily concerned with discussing this self-knowledge in terms of humility (Grenberg 2005: 221).

9 Laura Papish also views the pursuit of self-knowledge as a never-ending task, though she views this primarily as a result of the way that self-knowledge is a self-interpretation which necessarily changes and deepens the subject investigating herself (Papish 2018: 165ff.). Although this is a quite different account from the one I offer here, the two seem to me broadly compatible.

10 For a similar account, see Potter 2002: 386ff.

11 This is the so-called Inferential view, about which see note 7.

12 See Longuenesse 2017: 208ff.

13 Kant offers substantially similar definitions twice at *MM*, 6: 394, again at 6: 405, and again at 6: 440. For further discussion of this definition of virtue, see especially Baxley (2010: 51).

14 In contrast, Lara Denis characterizes the distinction at work here as one between *homo noumenon* and *homo phaenomenon* simply (Denis 2010: 172).

15 I must stress here the word ‘instead’. It is not the case that taking up non-moral ends is necessarily vicious, but only the taking up of such ends in place of, or in preference to, moral ends. Indeed, in the case of imperfect duties, which allow for latitude in how and when they are carried out, in many circumstances one may permissibly perform non-moral actions instead of moral actions.

16 Merritt also highlights the cognitive aspect of Kantian virtue: ‘the motto of the Kantian reflective ideal is *self-determination through understanding*’ (Merritt 2018: 208).



17 An anonymous reviewer helpfully points out that the discovery of one's moral deficiencies may be less useful as a guide to one's character in the case of imperfect duties, which allow for latitude in their execution, than in the case of perfect duties, which do not. For example, one cannot conclude that one has not taken up the end of giving charity by the fact that rather than giving charity one has spent some money on something else. For taking up the end of giving charity does not mean that one has to spend all one can on charity. But there is still room for moral self-knowledge here, and of an important kind. For one might come to realize that one has spent one's money because one does not properly prioritize the moral end of giving charity, rather than because one is exercising one's latitude to discharge this imperfect duty. What is more, one may discover that one is *always* spending money on things and not giving to charity, in which case one must conclude that one has indeed failed to take up the end of charity.

## References

- Baxley, A. M. (2010) *Kant's Theory of Virtue: The Value of Autocracy*. Cambridge: Cambridge University Press.
- Berg, A. (2020) 'Kant on Moral Self-Opacity'. *European Journal of Philosophy*, 28(3), 567–85.
- Bernecker, S. (2006) 'Kant zur moralischen Selbsterkenntnis'. *Kant-Studien*, 97, 163–83.
- Biss, M. (2015) 'Kantian Moral Striving'. *Kantian Review*, 20(1), 1–23.
- (2017) 'Avoiding Vice and Pursuing Virtue: Kant on Perfect Duties and "Prudential Latitude"'. *Pacific Philosophical Quarterly*, 98(4), 618–35.
- Denis, L. (2010) 'Freedom, Primacy, and Perfect Duty to Oneself'. In L. Denis (ed.), *Kant's Metaphysics of Morals: A Critical Guide* (Cambridge: Cambridge University Press), chapter 9.
- Emundts, D. (2017) 'Kant's Ideal of Self-Knowledge'. In U. Renz (ed.), *Self-Knowledge: A History* (Oxford: Oxford University Press), 183–98.
- Grenberg, J. (2005) *Kant and the Ethics of Humility: A Story of Dependence, Corruption and Virtue*. Cambridge: Cambridge University Press.
- Knappik, F. and E. Mayr (2019), "'An Erring Conscience is an Absurdity": The Later Kant on Certainty, Moral Judgment and the Infallibility of Conscience'. *Archiv für Geschichte der Philosophie*, 101(1), 92–134.
- Longuenesse, B. (2017) *I, Me, Mine: Back to Kant, and Back Again*. Oxford: Oxford University Press.
- Mahon, J. (2003) 'Kant on Lies, Candour and Reticence'. *Kantian Review*, 7(1), 102–33.
- Merritt, M. (2018) *Kant on Reflection and Virtue*. Cambridge: Cambridge University Press.
- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- O'Hagan, E. (2009) 'Moral Self-Knowledge in Kantian Ethics'. *Ethical Theory and Moral Practice*, 2(5), 525–37.
- Papish, L. (2018) *Kant on Evil, Self-Deception, and Moral Reform*. Oxford: Oxford University Press.
- Potter, N. (2002) 'Duties to Oneself, Motivational Internalism, and Self-Deception in Kant's Ethics'. In M. Timmons (ed.), *Kant's Metaphysics of Morals: Interpretative Essays* (Oxford: Oxford University Press), chapter 16.
- Radhakrishnan, V. K. (2019) 'The Duty of Knowing Oneself as one Appears: A Response to Kant's Problem of Moral Self-Knowledge'. *Problemas*, 96, 23–35.
- Timmermann, J. (2006) 'Kant on Conscience, "Indirect" Duty, and Moral Error'. *International Philosophical Quarterly*, 46(3), 293–308.
- Ware, O. (2009) 'The Duty of Self-Knowledge'. *Philosophy and Phenomenological Research*, 79(3), 671–98.