Kantian Review, 16, 1, 49–66 © Kantian Review, 2011
doi:10.1017/S1369415410000075

Embracing Kant's Formalism

BARBARA HERMAN
UCLA

Abstract

In response to critical discussions of my book, *Moral Literacy*, by Stephen Engstrom, Sally Sedgwick and Andrews Reath, I offer a defence of Kant's formalism that is not only friendly to my claims for the moral theory's sensitivity to a wide range of moral phenomena and practices at the ground level, but also consistent with Kant's high rationalist ambitions.

In a way, generations of Kant's critics have been right: understanding Kant's formalism is the key to understanding his ethics. Of course many of the critics think that the harder you look at the formalism, the less compelling the ethics. But that is to be seen, not presumed. Worries about Kant's formalism often blend with worries about his transcendental idealism, and in the case of the ethics, especially his views about the will. If the issues here are hard, some of the worries come too easily and really ought to be laid to rest. To take them on, I have thought, it would be best to develop as plain and unembellished an account of Kant's formalism as possible, one that is just sufficient for his announced purposes, though hopefully capable of being extended to meet some of ours. The idea would be to stay close to the theoretical ground, delaying the arrival of the familiar worries so that they show up, if at all, when we have resources to manage them.

For a variety of reasons, my own investigations of Kant's ethics have tended to be from the other side. Paradoxically, perhaps, I found the appeal of Kant's ethics in its provision of a theoretical framework and conceptual array through which one could understand moral character, practices and institutions. It has been worthwhile to let Kant be the teacher about basic notions of agency and will, character and choice, desire and deliberation. I also found that I could productively ask his austere texts to speak to a very wide range of practical issues, including abuses of power, extra-institutional justice, culturally demanded changes

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

of moral practice and perturbations of moral development. Not surprisingly, as I used the theory to work on the issues, I better understood what the theory contained. This is one of the virtues of approaching Kant's systematic philosophical account morality at the level of 'middle theory'.

It was always my assumption, tested occasionally, that one could do things the other way: go *from* the formalism *to* the extended engagement with moral phenomena. The three discussions of my work offered by Sally Sedgwick, Andrews Reath and Stephen Engstrom have turned that assumption into a temptation to offer just such an account. Trying to take in at one time Engstrom's rich characterization of the ambitions of middle theory and the value of close attention to moral phenomena, Reath's incisive exploration of the importance I give to the unifying work of obligatory ends and the completion project of practical reason, and Sedgwick's challenge that the very things I find most compelling in Kant's ethics might yet be incompatible with the formalism, I find I am wanting to knit the pieces together, from the side of the formalism. What follows is a sketch of how the paths converge.

1

It is close to the defining fact of Kant's ethics that some kind of principle is invariable and universal and binding. Exactly what it is, is a little less clear. We know its name – the moral law – and that the manner in which we encounter it is as a categorical imperative. Some of what we know is very abstract: the moral law directs us to ensure that our maxims have the form of universal law-giving. Some is concrete: we have duties and obligations that are, at the least, specific to our human nature, and perhaps more locally specified than that. In *Moral Literacy* (Herman 2008) I took on the project of exploring these possibilities, laying out a wealth of context-sensitive results at the level of moral requirement that I argued were consistent with Kant's one moral principle. I also offered what I regard as natural and attractive ways of thinking about moral action and volition that fit with this understanding of moral requirement and yet can still accommodate Kant's strictures about purity and rationality of motive. I did not offer a foundational argument for the fundamental moral principle, nor did I do much more than glance from time to time at the underlying Kantian commitments, especially the more metaphysical ones. My sense was, and is, that pressing hard on the practical side puts us in a better position to grasp the challenges faced by the theory because we know more about what it governs.

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

Sally Sedgwick challenges my approach from both sides. In the sphere of results, she suspects I attained the wealth of moral possibilities by letting Kant's one moral principle become vague and suggestive rather than sharp and rigorous. This not only undermines the invariable and universal commitment of the Kantian programme, she argues, but my embrace of moral pluralism is so extensive and so deep as to threaten the very idea of an objective basis for ethics. On the other side, my relative silence and occasionally voiced discomfort about foundational matters raises the spectre for her that I pick and choose the parts of the Kantian programme that are friendly to my normative ambitions, leaving the parts that don't work well for me in the shadows. Both concerns call for some clarification.

Historically, the first issue about Kant's formalism was that its moral principle was empty: a principle that could not, on its own, provide a standard of right and wrong for action. There is less worry nowadays among Kantians that this might be true. Some of this confidence comes from a shift in emphasis away from the formula of universal law to the formula of humanity; some comes from the recognition that fundamental categories of moral assessment – from property to promise to beneficence – are secured in the *Metaphysics of Morals*. While I doubt that the formula of humanity can replace the formula of universal law (because it depends on it), I agree that the *Metaphysics of Morals* is an essential piece of any account of Kantian moral requirements and a source of the terms of moral judgement.[1] But however one secures content, there remains the question of the place and point of the formalism.

I suspect that one impediment to working this out has been lack of clarity about the purposes of Kant's arguments in different places in his moral philosophy. It is, for example, essential to recognize that the *Groundwork of the Metaphysics of Morals* is not a text in normative ethics and its principles are not intended to be sufficient for moral judgement. Its announced purpose is to search for and establish the supreme principle of morality; its method of argument, Kant tells us, will forego the advantages of application.[2] Of course, the *Groundwork* uses examples, and more than once suggests that the formula of universal law can and should be used in appraisal.[3] But we make a mistake if we regard the examples as models of duty-derivation; their role in the text is clarificatory, to confirm at each stage of a very abstract argument about an *a priori* rational principle that we are in fact tracking morality. So if, as we believe, deceit is wrong and helping others is required, it had better be the case that violations of the duties can be expressed in

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

the terms set by the categorical imperative – that their maxims fail to be in conformity with universal law. And if there are different formulas of the categorical imperative, each must capture in its own way that flaw in willing. Setting aside entrenched habits of reading and teaching the *Groundwork*, there is no reason to suppose that the candidate supreme principle of morality is intended to be, *on its own*, generative of duties and obligations.

However, the *Groundwork* does provide the essential starting point for understanding the formalism: that the form of universal-law-giving is the condition of correct or good willing (and violation of the condition is what constitutes wrongful willing). I favour a relatively straight-up reading of this. If we begin with the idea of the will as practical reason – a capacity for practical reason*ing* to bring an agent to action – we can cast the moral law as the fundamental principle of practical reasoning, charting the correct route from some (true and appropriate) practical premise(s) to conclusions about action by way of a rule of practical inference. (It is in this sense that the moral law is the constitutive principle or law of the rational will's activity.) A maxim that conforms to the principles of sound reasoning is suited to 'give' universal law because it instantiates the universal form of a principle of practical reason (and in that sense the principle of the maxim applies universally, which is not to say that the maxim does).

This much would lead us to expect that actual instances of practical reasoning go wrong in either of two ways (assuming they are already instrumentally adequate). The first involves the overestimation of the justificatory reach of a premise (typically an end of self-interest). In the familiar deceitful promise example, the agent knows he is on morally difficult ground, but thinks his circumstances justify deceit.[4] Moral rules do have exceptions and he judges his case to be one of them. The universal law 'test' shows that there is not a self-consistent course of *reasoning* here (4: 442). If, as a matter of practical rationality, exigent self-interest did justify deceit in promising, any reasoner should be able to reason from exigent self-interest to the deceitful promise and then on to trust; but one cannot. A rational principle of trust in promises and a rational principle of deceitful promises based in self-interest are not compossible principles of practical reasoning, assuming both are fully general in their domain of application, as any candidate principle of practical rationality must be.[5] The failure is formal, a fault in reasoning, though what is reasoned about in a faulty or invalid way is a bit of material human activity.

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

The second way to go wrong is thinking that final ends – the premises of practical reasoning – are wholly discretionary or, what amounts to the same thing for Kant, under the authority of self-interest. The third and fourth *Groundwork* examples suggest that treating self-interest as a final end leads to a contradiction in willing because a principle of reasoning from self-interest will defeat itself (4: 423). There is a deeper explanation of this error about ends in the *Metaphysics of Morals*. There Kant argues that, since practical reasoning is from ends to action, if there are no ends given by practical reason itself, there can be no categorical imperative, because 'all ends would hold for practical reason only as means to other ends' (6: 385). To get us to the ends it is a duty to have, Kant offers a deduction from pure practical reason: 'what, in the relation of a person to herself and others, *can* be an end for practical reason *is* an end for practical reason' (6: 395).[6] This seems more obscure than it is. Suppose we asked, fully abstractly, what end or interest could practical reason qua practical reason have? 'The imposition of rational form in its domain of influence' has to be the answer. Strictly, this is the end that belongs to the formalism: we have a formal principle that applies to means (that maxims qualify for universal law-giving) and a principle that applies to ends (that action be given the aim of creating rational form). What that amounts to for us as human rational agents is that we should do what we can to manage our vulnerabilities as practical reasoners. We thereby get obligatory ends to self and others and the doctrine of virtue that comes with them.[7]

The two obligatory ends tax us to be responsible for the ways we affect our own and others' rational and moral well-being.[8] Towards ourselves, for example, we are to be careful about activities we choose because we enjoy them that can lead to sloth, greed, sloppiness about what we believe and how we reason. Such effects are not themselves ends, nor likely intended; it is rather that, given our psychology, there are risks to practical rationality that arise naturally and won't be averted unless we attend to the effects of our pursuits on our rational functioning. From the point of view of practical reason, it cannot be a matter of indifference that our psychology, which is affected by what we do, is vulnerable to disabilities that can render us less able to respond to rational requirements. Towards others, we have parallel responsibilities. Human rational abilities are vulnerable to the effects of poverty, humiliation and sustained misdirection. Ignoring others' needs, demeaning their choices, making someone's life too hard *or* too easy can affect their ability to sustain or value rational activity. We should think that impartial practical reason cannot be indifferent to activity

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

that undermines our own *or others'* ability to engage in reasoning. But note, it is the *formal* requirement of good reasoning that launches obligatory ends specific to the human situation.

Andrews Reath notes that in *Moral Literacy* I make a case for the priority of the obligatory end of self-perfection. This will seem odd if one thinks that moral requirements are fundamentally other-oriented, and that moral concern for self is therefore and essentially instrumental. But such a thought confuses the point of obligatory ends with the objects of specific duties and obligations, many or most of which may be directed towards others. Obligatory ends do much of their work as we are doing other things, building a moral orientation into our activities, a framework for willing well and acting attentively. If for most of us, much of the time, the work of correctness in practical reasoning and judgement is work we have to do ourselves, giving priority to one's moral (and practical) development is not misplaced moral self-absorption, but a condition of correctness in volition generally.

But there is another point about the priority of the duty of self-perfection that I perhaps made more difficult than it needs to be. In 'Obligatory Ends' (Herman 2008: 262–3), I focused on the fact that our rational abilities (not our rational agency) are dependent on the rational abilities of others. While I may selectively wish that (some) others be rationally deficient, or under my control, I cannot on rational grounds be hostile to or indifferent to the rational abilities of others. My own rational abilities are enhanced and supported by the good reasoning of others. Reath wonders why we cannot argue directly for the obligatory end of others' well-being – from their equal moral standing, perhaps. I think we cannot because these arguments are part of working out what equal moral standing amounts to. However, we can get a more direct argument from the formalism. If the question is not, 'how should I, on rational grounds, regard you?' but 'through what ends will I carry out the project of bringing rational form to the human world?' then even though the ways in which I can be responsible for my own rational well-being are very different from the ways I can be responsible for yours, I have both ends.

Overall, obligatory ends play the role of basic practical categories that are to shape both our pursuit and our idea of happiness.[9] As final ends, they are usually not the only premise in practical reasoning, and often not the active one in choice, but they are always first premises. Our preferences and private ends gain standing only as they are consistent

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

with obligatory ends – they are subjectively of first interest to us, but formally subordinate. The effect of taking Kant's ethics beyond permissibility and anchoring practical deliberation in obligatory ends is to bring a very wide range of ordinary human activity inside morality and alters the place moral thinking has in a life. It alters how we should regard everything from baseball, to teaching, to music, to politics – activities that have their own sources and rules.[10] Even when our actions are not impermissible, there is a moral question about the ways we are affected by what we do: heightened competitiveness in sports, arrogance in teaching, derogation of the arts, cynicism about politics, may create a culture of persons less able to reason when they should, or to reason well. When we get things right, we can see in our ordinary actions that we are, in addition to promoting our ends, supporting the conditions for free and rational activity.

Because obligatory ends are regulative, guiding deliberation and judgement, when they give rise to duties of action, the duties tend to be imperfect in the following sense. Suppose I am so circumstanced that I can help people; I could help my friends or my neighbours or my relatives or the needy in Sudan. It is a matter of judgement to determine where my help is most needed, most effective and consonant with other ends to which I must also attend, some of which may be self-regarding. Having judged, I should act. The imperfection of the duty is *not* a permission not to help so long as I am willing to help on some other occasions.

So, about Kant's formalism: at its most abstract, we have a principle that is invariable and universal and binding – a law – that is the source for standards of good practical reasoning. One step down, we have the categorical imperative together with the obligatory ends articulating the *form* of good willing *for us*. As premises and principle of practical inference, they make it possible for the 'empty' moral law to yield determinate practical results when applied to the material of human lives.

2

Looking at the formalism in this way, it is not difficult to see how Kantian ethics might provide an attractive option for thinking morally about the pluralism of ways people go about living. We know that some people lead solitary lives, some are intensely social; some are competitive, some seek the pleasures of cooperation. Entire cultures may prize one or the other, and develop aspects of morality differently as a result (think of conversational distance, where violations register as aggression or

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

indifference, depending on cultural views about personal space). However, openness to a range of possibilities for normal individual and cultural psychology does not imply that our conceptual hands are tied by the agents we happen to encounter or conjure up. Morality can say 'no' not just to actions but to patterns of development and even ways of life. Part of the purpose of morality is to change us, to train us to desires and interests that we wouldn't have had otherwise, and to alter our sense of the value of the things we already care about.

Suppose, as individuals, we start out with feelings of affection for others, desire to spend time with them, to share experiences, to have the pleasures of intimacy. We discover that there are things owed to others in such relationships, that we may not simply follow the path of feeling, that intimacy creates vulnerabilities, and so on. Recognizing this, we come to care about our friends in different and morally inflected ways, while still enjoying the pleasures of time together. There are also various norms of friendship. It can be casual; it can be more binding than family. Culture has a great deal to do with how this goes. Sedgwick worries that once one starts down this path, there is no stopping. Cultures are not just different in their norms and practices, they may have deeply different conceptions of the person, even different metaphysical views of the self, some that are surely incompatible with Kantian morality (or with the truth).

Here I have wanted to make a distinction between what is historically true about a practice's origins and the conditions for moral survival. A practice or a culture comes about as it does, with an origins story and supporting beliefs. For a practice to be *morally* justified, and so have a moral claim for survival (and so for our accommodation), it must be possible to reason to the practice's requirements from moral premises (among others): actions it requires must be consistent with general standards of permissibility and its values must be defensible on moral as well as its own grounds. So if a group has an ideal of adult maturity that discourages exercise of the ability to grasp and deliberate correctly about actions and practices generally, to that extent its practices are not justified; and if the ideal is inseparable from defining norms, the group may have no moral claim to survive.

To find a non-hysterical limit case, I conjured a culture that found non-physical pain difficult to accept as real pain (Herman 2008: 40). I thought I could understand well enough how it might emerge – warrior values honed in generations of honourable struggle to survive;

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

but, and this was the point I wanted to make, that a set of ideas has a compelling human history does not make it the case that the ideas support a morally acceptable way of living, though not just because it is different from our set of best practices. The qualification is important, because I also think we should be cautious in assuming that there is in principle a socially neutral conception of harm. If, as seems reasonable, human beings organize themselves in non-compossible sets of practices, and the practices are partly defined by attitudes and conceptions of self, then what may count as a harm *here* could be an opportunity *there* (lifetime employment a possible recent example) – though the view from *here* makes what is *there* seem unacceptable. But the perspective of such situated moral judgement is not original, not at ground level. What morality can require, and require universally, is that the notion of harm in play be consistent with the rational well-being of persons. The Kantian message about pluralism should be that it is an open question how much and what kinds of variety in human affairs should be tolerated; what are not open are the moral terms for answering it, case by case. There is no Kantian pluralism about moral first principles.

The invariance in Kantian morality's first principles does not require uniformity at the level of lives or of social practices. It is also not broadly permissive. Persons play the piano or garden or build highways because, some way or another, the activity suits them. A culture prizes individual achievement or cooperative endeavours, different models of parenting, as makes sense to them.[11] Whichever path is taken, morality enters two ways, assessing means *and* requiring that the activity as it works out in people's lives supports, and is taken to support, the values of the obligatory ends. The issue of moral possibility and so of goodness is whether an activity or practice can be reshaped *from within* by the moral values that then justify it. Only some of what individuals or cultures would like to do can survive this transformation.

3

One might have expected that somewhere in these arguments there would be appeal to the *value* or *worth* of persons – something always to be respected. I no longer find doing that helpful. It is not that I think we should not talk of the value or worth of persons, or of respect, but that these notions arrive on the scene as a result of, and not instead of, the work of the formalism. To say that a thing or person has value is to regard something about what it is as a source of direction for how we should act towards it. To say that rational nature has value is to say that

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

we are constrained by it – by its constitutive principle – in the ways we will and act. So what it is to respect the value of persons (as we engage with self and other) is to will in conformity with the principle of practical reasoning.

Talk of value does add something: it re-represents moral principle in terms connected to ends, and reminds us that it is *the good* that is the object of a will determined by practical reason.[12] This in turn shifts the aspect of the moral law to which we attend: less about law, more about universality, or the equality of persons qua rational beings. If the reasoning in my maxim is sound, if I have reasoned from correct final ends to action by practically valid steps, then my reasoning is, formally, reasoning that is good reasoning for any reasoner.[13] The ideal of possible co-reasoning lets us see more deeply into what is wrong with certain kinds of action. When I make a promise to you, my (apparent) maxim functions as a premise of your reasoning (which might go: 'Since BH, needing money, asks to borrow from me and promises to repay, trusting her word, I will loan the money'). But of course you don't capture my maxim; my (actual) maxim could not serve as the premise of your reasoning (you cannot reason: 'Since BH, needing money, asks to borrow and falsely promises to repay, trusting her word, I will loan her the money'). In the spirit of co-reasoning, we could say that my deceit makes *me,* not good reasoning (yours and mine), the final authority for *your* action.

Now to treat a person as an end in herself is to accept that the rational nature (the humanity) of the other is an equal and equally final source of justificational authority for what you will.[14] When we make use of another person's reasoning and choice, we ought to be able to regard our (actual) maxim as one that could be *co-willed* by each of us: the principle of action must be one that each can judge to be valid on the same rational grounds, and our maxims, our principles of action, should acknowledge and accept the conditions of co-willing. And if, as I would put it, the possibility of co-willing (or co-reasoning) is what the standard of universal form requires, we must support the conditions that make others good reasoners. The value of persons involves no more and no less.[15]

This high rational way of thinking about the value of persons actually gives some purchase on the moral significance of pain in Kant's ethics. It is a natural fact that living things that feel pain don't like it, and normally avoid it if they can. It is also a fact that most living things can

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

manage that dislike and tolerate fair amounts of pain in order to accomplish their ends. Pain is sometimes thought to be the central case of something that should matter to morality primitively: 'Because it hurts' is at least *prima facie* a reason not to do it. But where pain fits in the Kantian scheme has not been so clear.

Given a focus on correctness and independence of deliberation and judgement, pain (as well as pleasure) will enter the space of moral concern as things that impede or undermine the work of the rational (or reason-involving) capacities. Pain burdens activity, making it harder to think clearly and act effectively.[16] Pleasure prompts us to go on; pain makes it harder to. In the extreme, pleasure can make it difficult not to go on and pain forces us to stop. In both cases, we have good reason to do one or the other. If we do not independently prize the Stoic option, if there is reason to allow the vulnerabilities of choice that expose us to both pleasures and pains, then, given what pleasures and pains are like for us, morality has to take them into account, though not always in a way that matches our preferences.[17] Offering pleasure in some circumstances can short-circuit and so impede thinking as much as pain slows and dulls it. We could say: there is a moral interest in avoiding impeding pains and pleasures, in ourselves and others, just as we have a moral interest in avoiding making and opening ourselves to threats: they are not consistent with the conditions of good reasoning.

4

I want to turn now to a different set of questions, about motive and will. If the formalism tells us that a person acting morally is one whose actions follow from good reasoning, morality, at least on this account of it, requires that reasoning can affect what we do. To what are we committed in saying this? Following Kant, I would think that we are committed to some idea of rational causality. One question is: how much further must we go? Kant holds that we cannot regard rational causality, the causality of the rational will, as part of the temporal sequence of things in experience; that the idea of an autonomous will requires transcendental freedom, independent of but also connected to the material causality of nature. I see no reason to be shy about discomfort here. Kant is rather bold in his own discomfort, arguing that, in the end, it is enough to show that an idea we have to have for practical purposes is not impossible, and even helps make sense of other ideas we have to have about the understanding. The commitment we cannot forego is Kant's core insight in the practical domain: that the will is a

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

rational, norm-constituted causal power. The practical investigations that I have made explore what follows from the idea that the moral law, an *a priori* principle of practical reason, is the constitutive principle of such a will's causality. Does Kant have the best or the only account of the possibility conditions of such a principle or capacity? I don't see that his core insight is compromised if the answer is 'no'. To sustain the core insight we need a way to maintain the idea of reason as a distinctive ability that has a normative and effective role with respect to what we think and do. But there are many reasons to want to have that.

The core insight does require a particular understanding of the motives of a rational being, and so of us. Rather than an element of a system's hydraulics, a motive is a determination of the will – that is, a piece of practical reasoning. Here is the picture that is *not* right: I recognize a morally salient fact; I figure out what I ought to do, given that fact; but what happens depends on the presence (or absence) of a sufficiently strong motivational source to get me from recognition to action. Rather, on Kant's view, the practical cognition is itself sufficient for action. Of course means and opportunity need to be secured, and there can be impedances to and failures of reasoning. But there is no extra motivational element beyond the agent's reasoning.[18]

We, however, are so constituted that, although we come to action by way of practical reasoning, we are prone to errors that make our representation of the moral law's implications inaccurate or incomplete, subject to various kinds of distortion or *mis*representation. The erring agent, accepting the standard that his principles of action should be fully general, yet pressed from the side of his wants or needs or happiness, offers a friendly modification of principle that he counts as satisfying the standard. The effects of non-rational desires on willed action are thus not direct, as if they provided an independent and competing source of motive or agency; their effects are on actual reasoning, hindering the work of practical cognition.

5

Now it is one thing to be talking about Kant's formalism and the standards of correctness that are drawn from the moral law, and quite another to give an account, within the terms of the theory, of what real people are to do, guided by morality. It seems right both to tax the theory with the needs of practice, and to keep the practice of moral judgement and action open as much as possible to the elements of the

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

theory. If we start from the side of practice, we begin with human beings, real agents, who are not to be understood in abstraction from time and place. They come burdened with psychology and history, beholden, it will seem to them, to a variety of evaluative conceptions, any number of which can turn out to be mistaken, yet responsible for getting things right. One of the things a moral theory will have to provide is tools for sound moral thinking, a kind of middle theory, where high abstraction can be drawn down into the kinds of practices and institutions within which real persons act. We need questions to ask – 'who benefits?' is always a good one – and warning signs to watch for. If, for example, the justification for a restrictive dress code is that it protects women from sexual predation, we should want to know why women should bear the burden, and how it comes about that the predatory impulses remain outside sanctions. The questions and the warning signs devolve from an ideal of respect for persons, one that leads us to prize and protect the conditions necessary for each of us to be an autonomous judge of the correctness of ends and action. What this amounts to has to be worked out, over and over again, as different issues confront us.

An agent judging and acting depends on an array of skills and practices and appropriate sensitivities to morally salient facts. They are acquired over a lifetime, formed socially, and both open to and resistant to change. The resistance comes with the need for an agent to be confident in her choices and actions. We will be burdening ourselves and others with what we do and choose, not merely in exceptional circumstances, but routinely when we make or break promises, spend resources on some rather than others, or when we blame, or ask for forgiveness.[19] The challenge for us as moral agents is to be both confident in what we do *while* taking seriously that we may not have it all right. I do not mean that I should harbour doubt about the correctness of responding to your request for the time by telling you what it is, or doubt that promises impose burdens on both parties (of performance for one party and of release, if necessary, for the other). But I might want to be prepared to think more about what I am obliged to do as a teacher when I have students whose social background prepares them poorly for college work: something that is not in my job description may be something that qua teacher I nonetheless ought to do, on my own, or with others with whom I share an institution (not everything we ought to do can we do alone). We should be open to asking in these times what obligations those of us with secure jobs might have to the increasing numbers of our co-citizens who do not. Maybe we should

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

always have asked that question – that is to be seen; the question certainly presses now, and if we are competent moral agents, will begin to learn how to think about it.

It was in thinking about practical questions of this sort, as well as the conditions of more intimate successes and failures, that I was drawn to talking about *moral literacy*: a capacity to read and respond to the basic elements of a moral world. I find the notion attractive because of the mix of competence and performance: there is a skill and it is in use. I tend to use the literacy metaphor in its active mode – less possession of a technical skill or body of knowledge than engagement with the skill or knowledge and its subject. (As opposed to a notion that I also use of minimal literacy – just enough competence to get you going and to build on.) To be literate in reading Shakespeare is not just to get it word by word, but to engage, to find certain questions both intelligible and necessary: we learn to ask, 'what about his way of letting himself be loved made Othello so vulnerable to jealousy?' So, I think, to be morally literate is to know when to ask: how did we get here? How was it that taking care and trying to do the right thing still left everything in shambles?

On the active side of moral literacy, we will want to think about various kinds and levels of moral competence. Domains of competence can be limited. What is built on the foundation of minimal competence will often have gaps and vulnerabilities. But the competence is what makes one a moral agent, a subject of moral predicates, rightly to be held responsible for what one does, even in the regions of one's gaps and vulnerabilities.

Stephen Engstrom wonders why moral literacy should be unlike many other skills, a kind of mastery that can be used either for good or for evil. How it is that there is something about *moral* literacy that blocks this, somehow implicating the presence of the moral motive? He asks: given the array of behaviour-directing institutions and systems of social sanction within which we live, why do *they* not provide a kind of moral literacy that gives the vicious person skills for conforming when necessary, and exploiting morality when it is to his advantage?

Let us look a bit at the cases. Among the figures Engstrom has in mind is Hume's sensible knave. The knave has a problem with the rules of justice (there is no reason of self-interest to prefer vice over virtue apart from this, says Hume). He gets it that having the institutions of justice is generally beneficial; what he does not 'get' is why things would not be

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

even better if justice reigned and he took advantage of some crease or loophole where 'an act of iniquity or infidelity will make a considerable addition to his fortune' (Hume 1777: 232). Hume thinks such a person is a dupe, often a victim of his own cunning, and someone who has traded the great values of enjoyment of character and society for 'worthless toys and gewgaws … feverish, empty amusements'. Hume does not see in self-interest a general threat to virtue because the self in Hume's theory is thoroughly socialized and people are dependent on others for their good opinion of themselves. However, justice is the part of morality that works by way of public rules and institutions, and can function without the allegiance of persons – indeed it is designed to do just that. Because the rules do the work of judgement for us, *they*, not the knave's moral literacy, create space for his iniquity. But because the sensible knave is not otherwise vicious, he will be at odds with himself when his iniquity, which is episodic, conflicts with his benevolence, which is deep-seated. Or so Hume hopes.

Beyond Hume's world, there are other dark figures. There is the moral simulator, someone who is thoroughgoingly selfish but responsive to external sanctions that lead him to track morality. He looks like one of us, but he is in fact a vicious character. Engstrom wonders whether his responsiveness to sanctions depends on his being morally literate. That literacy might then enable vice (or at least protect it)? I think not. If, as I think they must be, external sanctions are directed at insuring rule-following, an agent sensitive to sanctions will be tuned in to the rules, but not to the values that they embody. As a result, his grasp of what morality requires will not outstrip the rules. Where they run out, in the interstices where judgement is required, where one might need to make an exception, he has nothing to go on. And if he is in this way evaluatively blind, he is not morally literate, though he knows about the advantages of appearing to be moral. He is in this way like the paler figure in Kant's *Groundwork* who acts only 'according to duty', whose maxims lack moral content. And that suggests that the viciousness may be a red herring – it is the *simulation*, however motivated, that is the problem.

More disturbing is the moral predator, whose understanding of morality, especially the morality of others, is what allows him to take advantage of them. He too knows some things about morality, and he makes his nasty way by exploiting moral facts. A con artist; Bernie Madoff, perhaps. He is close cousin to Kant's doctor whose skill makes him such a good poisoner (where we do not think the skill resists its use for that purpose).

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

As I see it, moral literacy, or the moral literacy that belongs with a Kantian account of the moral agent, is a competence that cannot be separated from the moral motive. While its possession is consistent with episodic or local immorality, it is not like a skill or body of knowledge equally available for good or evil use. The question may arise because of the way the moral motive is being understood – if, for example, one thought that the competency of moral literacy was on the side of belief, and the role of the moral motive is to make it real. But in Kant's theory, the motive is the determining ground of volition or practical cognition. It provides the premise from which reasoning begins, and so brings value into effective judgement. In less Kantian terms, the morally literate agent knows not just *that* something ought to be done, but *why*. She may not have a deep understanding of the *why*, but she recognizes that politeness or kindness or refraining from violent self-assertion is a matter of appropriate moral regard, that it flows or follows from respect for persons, or an acknowledgement of their equal standing and claim to a particular kind of attention in one's reasoning to action. And she knows that reasoning in this way has final authority for her action. Whether or not such a person acts well in the instance, if she has this knowledge, she cannot lack the motive. *Knowledge that* can be had from rules, and so it can be simulated in external behaviour; knowledge *why* requires a grasp of value, which, in Kantian theory, is a matter of motive. So I do not think we should count the moral predator as morally literate, though he or she may well have knowledge of morality. Theoretical knowledge, we might say.[20]

6

At this point we have come full circle. By reading Kant's formalism as a matter of practical cognition, a way of bringing rational form to agents' aims and actions through the requirement that they reason to what they do from moral ends by way of the rule of the categorical imperative, we come to see that and how the formalism pervades the moral theory. Although moral, philosophical and interpretative questions remain, it should be clear that the formalism is a source of the theory's power, not a problem to be avoided.

Email: herman@humnet.ucla.edu

Notes

1 I discuss the relation between the two formulas in 'The Difference that Ends Make', in Wuerth (2010: 92–115).

2 In the preface (4: 392) he tells us that the establishment of the supreme principle is a purpose to be kept apart from other moral investigations, and warns that the facility

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

with which a principle can be used is not only not a guarantee of correctness, it distracts from the rigour needed to establish the principle. Citations to Kant's works are to the Prussian Academy volumes and page numbers.

3 The 'should' here is one source of the worries about Kant's formalism, but it should not be. That it is comes from lack of attention to what appraisal involves. For Kant, I will argue, appraisal looks at an agent's deployment in her maxims of moral categories like promise and lie; it is not a first argument for what matters morally.

4 The maxim that represents the bit of reasoning would be something like 'when in need, one may promise deceitfully'.

5 The deliberative device of the formula of the law of nature makes the compossibility condition vivid: laws of nature have to be in systematic harmony. Kant emphasizes this in the *Critique of Practical Reason*'s examples.

6 A categorical imperative cannot be based on any purpose based in an agent's sensible nature. Any end of practical reason is, because it is an end of practical reason, an end it is a duty to have.

7 This is right if we restrict focus to individual ends and actions. Collectively, or omnilaterally, we must also give rational form to a social world through active membership in a republican form of government.

8 Kant calls them the obligatory ends of self-perfection and the happiness of others. I prefer 'rational and moral well-being' rather than happiness, because the content of the duties that fall under the obligatory end of others' happiness is less about supporting their ends than about meeting what Kant calls their 'true needs' and avoiding acting in ways, or with attitudes, that undermine persons' confidence in their equality and dignity.

9 The remaining practical categories are drawn from the conditions of life in society – property and contract, family and law. They set limits on action and provide terms for how we should think about what we do.

10 Reath wonders whether there cannot be realms of objective value between the moral and the mere desire-satisfaction. Surely other kinds of activity – aesthetic, affiliative – can have non-arbitrary form, and something like objective standards of their own. The Kantian response has to be: only one standard can be unconditional. A little more helpfully one might note that the question conflates two senses of objective that it is essential to keep separate. The non-arbitrary, meaning: not a matter of preference or person-to-person variability; and the objective: meaning, determined by the object. The beautiful can be a species value without being an objective value (we can demand agreement in aesthetic judgement among all persons because they share a certain structure of mind). Only the moral offers an object, the Good, which is or contains the standard of correctness in judgement for all rational beings.

11 Parallel with individual duties, a society or culture may have to have satisfactory norms in some categories – parenting and the general tending of children is an obvious example.

12 This is the point of the important chapter 2 of the *Critique of Practical Reason*, On the Concept of the Object of Pure Practical Reason. The will willing well is its own formal object.

13 In this it is rather like the ideal of scientific reasoning: good experiments can be repeated; their success does not depend on their being performed by this or that person. A maxim that meets the conditions of universal form can be part of the sound practical reasoning of any agent.

14 To treat humanity in *yourself* this way is to bind your choice to the final authority of rational nature.

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press

15 That is, moral requirements do not end if my maxims are ones that *could* be willed by others. I may need to take steps so that others can see what my maxim is, or to insure that the manner of my action does not intimidate. Conditions of co-reasoning won't be satisfied if some of the parties are too hungry to think straight.

16 If it was not a burden we might not mind it. But then, if we did not mind it, pain would not perform its warning role about the condition of the body under stress. The injunction to 'play through the pain' can be an intelligent response to pains that we have learned are not dangerous, or a signal of indifference to the well-being of the active person. Both acknowledge pain's impeding effect.

17 There are complicated issues here having to do with sensible pleasure and pain and the possibility of enjoyments that belong to the exercise of our rational faculties. Kant had doubts that reason was the superior route to the former; there are interesting arguments in Kant about the existence and place of the latter. Were it the case that rational activity was not itself enjoyable, the prospects for us as rational beings might be even dimmer than they are.

18 This is one of the reasons why overdetermination of motive should never have been an issue in discussions of moral worth: desires and inclinations are a different order of thing, their presence or absence no more relevant to the agent's reasoning than a scientist's pleasure in the result of a well-formed experiment taints the result. This point is developed in Herman (2006).

19 Some of the purpose of having rules and practices is to make it easier for an agent to know what to do.

20 My interest has been in the deformities and perturbations of more or less normal moral character. They may not be the right point of entry for thinking about the many different darker forms of vice and viciousness. If we followed Kant, we might want to look at the affects and the passions, elements of our psychology that Kant describes as competing with or shutting out the sovereignty of reason. They would seem to pose a very different threat to morality than the characters of the knave or scoundrel.

References

Herman, Barbara (2008) *Moral Literacy*. Cambridge, MA: Harvard University Press.

—— (2006) 'Reasoning to Obligation'. *Inquiry*, 49, 44–61.

Hume, David (1777) *An Enquiry Concerning the Principles of Morals*. Repr. Oxford: Oxford University Press, 1975.

Wuerth, Julian (ed.) (2010) *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*. Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S1369415410000075 Published online by Cambridge University Press