



# Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind

## Citation

Korsgaard, Christine. 1998. Motivation, metaphysics, and the value of the self: A reply to Ginsborg, Guyer, and Schneewind. *Ethics* 109, no. 1: 49-66.

## Published Version

<http://dx.doi.org/10.1086/233873>

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3122464>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind

*Christine M. Korsgaard*

I would like to begin by thanking Hannah Ginsborg, Paul Guyer, and J. B. Schneewind for their fair and sympathetic criticisms. None of these philosophers is interested in opposing the Kantian approach to ethics. Instead, they have made some powerful challenges both to my reading of Kant and to my understanding of what a Kantian position in ethics involves and requires. So I take our shared task to be that of seeing what we can make of Kant's position, how attractive and defensible it may be rendered. I feel fortunate to have commentators who have taken up my work so exactly in the spirit in which I intended to offer it.

I am going to begin by laying out, as clearly as I can, my current understanding of Kant's moral psychology—the picture of what is involved in human motivation that lies behind Kant's ethics. Over the years I've become increasingly aware of how different this picture is from the vaguely empiricist picture which most of us trained in the analytic tradition take for granted, and of how much difference it makes to the correct interpretation of Kantian ideas. It will be especially necessary to have this picture before us in order to explain why I think that Hannah Ginsborg has to be wrong about Kant's attitude—or perhaps what should be Kant's attitude—about the free choice of nonmoral ends. Two developmental issues bear on the textual evidence for what I am about to say. First, I do not claim to have had this picture clearly before my own eyes in writing the essays in *Creating the Kingdom of Ends*, especially the earlier ones.<sup>1</sup> My own grasp of the picture has developed slowly over time. Second, I think that Kant himself only grasped the full import of his theory of free choice over time, and that the psychological picture I am about to lay before you is more firmly behind what Kant says in his later ethical writings than it is in the *Groundwork of the Metaphysics of Morals*.

The starting point for motivation is the incentive (*Triebfeder*). You are subject to an incentive when you are aware of the features of some

1. Christine M. Korsgaard, *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1996).

object that make the object attractive or appealing to you. Perhaps the object satisfies one of your needs, or perhaps because of human nature or your own particular temperament the object is one you are capable of enjoying—it interests you, it arouses the exercise of your faculties. The naturally sympathetic person in the notorious example of *Groundwork I*<sup>2</sup> has a capacity to enjoy helping people, so he experiences occasions on which he can help as providing an incentive to act.

Incentives operate on conscious beings causally, as a kind of attraction, but they do not by themselves cause actions. Incentives work in conjunction with what I will call principles, which determine the conscious being's responses to those incentives. A motive is an incentive which a conscious being acts on under the influence of a particular principle. In a nonhuman but conscious animal, the principles in question are instincts. An animal's instincts tell it to hunt when it is hungry, flee when it is afraid, fight when it is threatened, and so on. You can see from this description that incentives and principles exist in natural pairs: the fact that an animal has certain instincts explains why it is subject to the associated incentives. In this sense the animal's instincts play a double role in the account of its actions: they explain both why the animal is subject to certain incentives in the first place and how it responds to those incentives once they are present. The incentive is something presented to consciousness; the principle is something within the animal, something in its psychological structure, that determines how it responds to what is presented to consciousness; and it is the two together that produce the action and so constitute the motive. A motive, one might say, is an incentive operating under a certain principle, or viewed from the standpoint provided by that principle.<sup>3</sup>

Now of course in referring to instincts as a kind of principle, I mean to invoke an analogy between the work of instinct and the work of rational principles or principles of volition, for both of these determine how we respond to incentives. So now let us turn to the human case. A non-human animal is conscious, so its actions are determined in part by representations of things in the environment. But we are self-conscious, which means that we represent to ourselves not only things in our environment, but our own mental states: we think about them. Among other things this means we can think about the operation of incentives on our own minds. When you are aware of the operation of a natural incentive

2. Immanuel Kant, *The Groundwork of the Metaphysics of Morals*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1998), 4:398. Page references to Kant's works use the volume and page numbers of the German Academy edition, which are found in the margins of most translations; except in the case of the *Critique of Pure Reason*, which I cite in the usual way using the pagination of the A and B editions.

3. For another discussion of incentives and principles in Kant's moral psychology, see my *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), sec. 4 of the "Reply," esp. pp. 242–44.

on you—aware that something attracts you, or tempts you to action—you describe yourself as “inclined” to act in the way the incentive bids. So that’s what an inclination is: it is your consciousness that you are naturally attracted to acting a certain way.

As I argued in *The Sources of Normativity*, this same self-consciousness or reflective awareness of our own inner states and activities is the source of free will: precisely because you know that you are inclined to act in a certain way, you are now free to decide whether to act that way or not, and so are faced with a question. So it is in this space—the space of reflective distance from our incentives—that the issue of *reasons* arises: when you ask yourself whether to act as some incentive bids, you are asking whether your inclination constitutes a reason to act. This is why we need principles of volition or principles of reason, which tell us whether and when our incentives provide reasons to act.<sup>4</sup>

Notice here one way in which the Kantian picture is different from the familiar empiricist one. Acting from reason is not opposed to acting from inclination; reason and inclination have different functional roles in action. The faculty of inclination alerts us to incentives, or possible reasons, for action, while reason has the function of deciding whether to act as we are inclined to or not.<sup>5</sup> The opposite of reason here is instinct, and the problem of normativity emerges because self-consciousness liberates us from the government of instinct. Animals are heteronomous not because they act on inclinations but, rather, because their instincts determine how they respond to incentives.

But human beings act under the idea of freedom: when we are faced with an incentive, we *choose* whether to act on it or not. In place of instincts, then, we must have principles of reason or volition which determine how we deal with the incentives which present themselves to us. And Kant thinks that it follows from the fact that we act under the idea of freedom that the *fundamental* principle of volition is the categorical imperative. Since we are under our own government, rather than under the government of instinct, we make laws for ourselves. So when you are confronted with an inclination to act, the question you must ask yourself is whether you can will the maxim of acting on that inclination as a law.

But at this point—and here I am beginning to address Hannah Ginsborg’s argument—you might ask: Why do we even get this far? It is evident from what I have just said that Kant thinks that when we are conscious of an incentive we experience it as a sort of candidate reason to act. Kant supposes that when we are naturally inclined to do something, we

4. Korsgaard, *The Sources of Normativity*, 3.2.1–3.2.3, pp. 94–98.

5. In “Self-Constitution in the Ethics of Plato and Kant” (forthcoming in *The Journal of Ethics*), I argue that Kant joins Plato in advocating a “constitutional” model of the soul (that is, one in which the different parts of the soul are assigned different functions in the deliberative process) as opposed to the “combat” model of the soul (in which reason and passion fight for control) which is usually presupposed in empiricist discussions.

are going to treat that natural inclination as a reason as long as the resulting maxim can be willed as a law.<sup>6</sup> But why should we do that? Now this question presses especially on Kant as I read him, since as I read him he does not accept any form of substantive value realism. He does not think that the objects of our inclinations are good in themselves. We do not want things because we perceive that they are good; rather, our initial attractions to them are natural psychological impulses.

These initial attractions can be *explained* in terms of our instincts, which still operate in us even though they no longer govern us, and in terms of the modifications of instinctual attraction that result from certain operations of reason, such as comparison. Here I have in mind Kant's ideas about how specifically human desires arise from such rational operations, which he details in the essay "Conjectural Beginning of Human History" and which I discuss in "Kant's Formula of Humanity."<sup>7</sup> In Kant's example, Eve's temptation to taste the forbidden apple is prompted by its outward similarity to other fruits to which human beings are instinctively drawn. Ginsborg criticizes me for offering these views as evidence that reason guides us in the choice of nonmoral ends, claiming that the use of reason to compare objects instinctually desired with those which are similar but not instinctually desired is theoretical, not practical. Yet Kant himself describes the occasion as one on which reason, in the person of Eve, makes "the first attempt at a free choice" and he emphasizes that the event represents our liberation from the control of instinct.<sup>8</sup> I believe what Kant has in mind is something like this: the new incentives engendered by the comparison cannot be ones upon which Eve could act under instinctual principles, so she finds herself conscious

6. A complex question arises here as to whether on the sort of account I am putting forward it is correct to say that the inclination is treated as a reason or whether we should say that the incentive is treated as a reason. Suppose, for example, I have an inclination to walk in the woods, and the incentive is that I find it peaceful. If someone asks me why I am walking in the woods, I will mention the incentive, not the inclination: that is, I will say "because I find it peaceful" not "because I want to." This suggests that the incentive provides the reason. The difficulty with saying that the incentive provides the reason, however, is that it conceals something essential to the Kantian account as I understand it. As I go on to say in the text, I do not believe that Kant accepts any form of substantive value realism, and this means that I think he would reject the claim (and I would reject the claim) that "a peaceful state of mind" is an intrinsic value that one realizes by walking in the woods. Instead, the peacefulness gets its value from its relation to the walker and is ultimately valued by the walker because it is good for her or good in relation to her, and she values herself. The advantage of saying that the inclination provides the reason, although it is somewhat misleading, is that it keeps this relational *character* of human values more firmly in view, and that is why I have stayed with this formulation in the text.

7. See Korsgaard, *Creating the Kingdom of Ends*, pp. 110–14.

8. Kant, "Conjectural Beginning of Human History," trans. Emil L. Fackenheim in *Kant on History*, ed. Lewis White Beck, Library of Liberal Arts (Indianapolis: Bobbs-Merrill, 1963), 8:112.

of the need for new principles.<sup>9</sup> Since she is thereby aware not just of the outside world, but of the principles upon which she herself acts, this is a first moment of self-consciousness, which in turn brings freedom and practical reason in its wake. But Ginsborg is correct to emphasize that reason's role in generating these new incentives does not by itself make them objects of rational choice.

Nor does Kant think that the satisfaction of desire as such is good. So our tendency to treat our inclinations as candidate reasons cannot be explained by saying that either the objects of our inclinations or their satisfaction as such is good. And this might indeed make you wonder why we treat them as candidate reasons at all. You might think that, having turned the cold eye of reflection on the operation of incentives in us, and having in this way distanced ourselves from that operation, we see that they are just products of our natural psychology, with no value of their own, which we could just as well regard as so much psychic noise or as obscure and meaningless compulsions which we should try to suppress or ignore.

Kant explains the fact that we don't do that by appeal to a volitional tendency which he calls the principle of self-love. Kant defines self-love as "the propensity to make oneself as having subjective determining grounds of choice into the objective determining ground of the will in general."<sup>10</sup> Self-love is a primitive tendency that belongs to us as sensible or animal beings. Kant says "we find our nature as sensible beings so constituted that the matter of the faculty of desire (objects of inclination, whether of hope or fear) first forces itself upon us, and we find our pathologically determinable self, even though it is quite unfit to give universal law through its maxims, striving antecedently to make its claims primarily and originally valid."<sup>11</sup> Essentially, self-love is a disposition to treat our natural incentives as grounds of action, which has its roots in our animal nature, but which in a rational animal takes the form of a tendency to treat the incentives of inclination as reasons. Now, I realize that it may seem as if rather than explaining why we tend to treat inclinations as candidate reasons, I have only given the tendency a name, 'self-love', so let me emphasize what I think is explanatory and distinctive about the Kantian account. What is distinctive about Kant's account is this: according to Kant the reason we treat our inclinations as reasons is not that we think the objects of inclination are good in themselves or that we tend to take satisfaction as such to be good in itself. It is rather

9. This is a moment in human history that, according to Tamar Schapiro, is recapitulated in the development of each human child. The idea is elucidated in her unpublished paper, "What Is a Child?" to which I am indebted.

10. Kant, *The Critique of Practical Reason*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1997), 5:74.

11. *Ibid.*

that we take the objects of inclination to be good for us and we tend to take ourselves as good—I mean in the sense that we are important to ourselves. And Kant links this fact to our animal nature. Kant thinks that every animal has a primitive tendency to value itself, which is expressed by its tendency to act on its own natural incentives.<sup>12</sup>

Historically, Kant's idea here has roots in the Stoic and Natural Law tradition. According to thinkers in these traditions, nature has constructed every living thing so as to preserve itself. So much is true even of the biological mechanisms which maintain plants in existence. But consciousness makes a difference: consciousness enables the animal in a sense to enter into nature's plans for it, to actively try to preserve itself. The animal's tendency to work for its own self-preservation—that is, to perform *actions* that tend to preserve itself—is thought of as a kind of self-concern, a kind of love the animal has for itself. So our basic tendency to try to satisfy our inclinations comes not from the fact that we value their objects nor from the fact that we value satisfaction but from the fact that we value ourselves. Self-love, Kant says rather pleasantly, is a sort of benevolence toward the self: you want yourself to have the things you want because you care for yourself, in exactly the same way that you want those you love to have the things they want because you care for them.<sup>13</sup>

So self-love is what gives our inclinations standing—why they automatically count as candidate reasons. Recall that I said before that incentives and principles exist in natural pairs: self-love is the principle that pairs with our natural incentives and explains both why we are subject to certain incentives in the first place and how we respond to those incentives once they are present. But as I argued in *The Sources of Normativity*, the principle of self-love exists at a kind of parting of the ways.<sup>14</sup> If you follow the tendency of self-love unreflectively, not asking yourself why it matters that you should get what you want, your state tends to degenerate into the state Kant calls “self-conceit,” in which you act as if it mattered that you get what you want just because you are *you*. But if you identify yourself with your humanity or power of rational choice, the principle of treating humanity in your own person and that of any other as an end in itself more or less directly follows. But this means that, although Kant doesn't put it this way, there is a sense in which the moral principle is not a principle which stands over and above the tendency to self-love, checking and correcting and limiting it. Instead, it is a kind of definition of it, a filling out who the *me* is that is the object of my self-love. The thought leading to the Formula of Humanity identifies *me* with my *humanity*.

12. For more on this point about animal nature, see Korsgaard, *The Sources of Normativity*, 4.3.6–4.3.9, pp. 149–52.

13. Kant, *Critique of Practical Reason*, 5:73.

14. See Korsgaard, *The Sources of Normativity*, sec. 4 of the “Reply,” pp. 249–51.

Now let's return to the account of action. You experience a certain incentive, and you are inclined to act on it. What happens next? Strictly speaking what the will chooses is actions, so you must first do some instrumental reasoning about how to satisfy the inclination, and formulate a maxim. Let us say then that you are considering doing Act-A for the sake of End-E. Since self-consciousness makes you autonomous and so you must live under laws you make for yourself, the deliberative question you must ask is whether you could will that maxim as a law. So you ask yourself whether you could will the maxim of doing Act-A for the sake of End-E as a universal law. That's the categorical imperative test.

Now we will take two cases. First, suppose you cannot will the maxim as a universal law: it fails the test. In this case, autonomy demands that you do not do the action. Now Kant argues that in this kind of case the thought that you cannot will your maxim as a law has a complex effect on feeling, the effect Kant calls "respect" (*Achtung*). Kant tells us that the operation of moral thought checks the initial or primitive tendency of self-love, by thwarting the inclination, and it strikes down the tendency to self-conceit, by reminding the agent that his interests have no claims prior to the law. These effects are painful, and the latter even humiliating. At the same time, however, there is a positive side to the experience, a sense of release, of obstacle removed, of independence, as the agent puts inclination aside and steps forward, so to speak, in the confidence of true self-mastery. The complex mix of affect thus produced is respect, which is what serves as the incentive to moral action: the alternative action, the morally correct one, is presented to consciousness as worth doing.<sup>15</sup>

In the second case, the maxim passes the categorical imperative test and can be willed as a law. Although the incentive is natural, the principle under which the incentive is chosen is the moral principle. The maxim—which includes the action along with its end—is chosen as one fit for an autonomous being to act on. The end is chosen as one that is of value to me, but I now view *myself* as a human being and my end not merely as *mine* but rather as an object of human aspiration, that is, an end that is potentially shareable with others.<sup>16</sup>

\* \* \*

With that as background, let me turn first to Hannah Ginsborg's argument that on Kant's view nonmoral ends cannot be freely chosen. I hope it is clear from the description of action I have just given why in general I think that nonmoral ends are freely chosen. Ends are chosen as parts of maxims which in turn are chosen as laws the agent gives to herself. It is true that there is a sense in which action from duty is more

15. Kant, *Critique of Practical Reason*, 5:72–76.

16. For a slightly different route to this thought, see my "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values," in *Creating the Kingdom of Ends*, pp. 275–310.



completely autonomous, or anyway more completely self-generated or reason-generated, than action undertaken for the sake of satisfying an inclination. Action from duty is, in a way, more completely "top down" since the incentive is generated by the activity of reason rather than adopted from nature. The thought of the moral principle in this case generates its own incentive, whereas when we perform a permissible action the moral principle operates on a natural incentive. The moral incentive is, as Kant himself says, really just our consciousness of the moral law operating in us.<sup>17</sup> But it remains true that we do not act on incentives of inclination unless we choose to, so I see no sense in which action on inclination has to be unfree.

However, let me try to bring the issue between Hannah Ginsborg and myself into sharper focus by saying where in the picture that I have set before you it arises. Here it will help to distinguish two cases. I will first discuss a case where all that is in question is a single inclination, which we find the moral law does not oppose, and ask whether our action on that inclination is in any sense unfree. I will then discuss Ginsborg's other scenario, where we are choosing among the several ends of inclination and appeal to considerations of prudence in making the choice.

So far as the first case is concerned, one way to understand the view that action from inclination is unfree is to suppose that the inclination operates on us as a kind of force. If the moral law has nothing to say against acting on the inclination, then I cannot help but do so. But an incentive is not a causal force that is blocked by another causal force, the moral law, and which therefore operates freely as soon as that other force gets out of the way. It is simply a consideration that has a kind of automatic standing with a self-conscious animal prone to self-love.

But Ginsborg's view, as I understand it, is not that the incentive operates as a force; it is that we cannot help but have something like the principle of self-love, in the form of the principle of seeking our own happiness. So we should ask: Could a rational being resist the basic tendency to self-love and therefore deny that her inclinations do have standing? Although Kant seems to deny that this could happen in *Religion within the Limits of Reason Alone*,<sup>18</sup> I see no reason for him to do so, so let us suppose that you could. What would follow? You won't be pursuing any personal ends of your own, so what maxims will you consider acting on? We encounter the perfect duties, duties to tell the truth and keep promises and so forth, in the ordinary course of action, going about our business, but you now have no business. How about the imperfect duties? Presumably if you don't regard your own inclinations as having standing, you won't regard the inclinations of others as doing so either, and your duty to promote their happiness will disappear. So what are you going to

17. Kant, *Critique of Practical Reason*, 5:72.

18. Kant, *Religion within the Limits of Reason Alone*, 6:36.

do—the other imperfect duty, develop your talents and powers? Kant says that these are given to us for all sorts of possible ends, but now you have no other ends. The denial of self-love is a route to normative skepticism and emptiness, not to freedom from the control of inclination; unless human beings place value upon ourselves, there can be no reasons and values at all.<sup>19</sup>

The larger point here is that in Kant's theory our values are created from psychological materials, from the natural bases of interest and enjoyment, rather than from nothing. Here as elsewhere in Kant's theory, reason works by imposing form on a matter that it finds. In this case, reason creates values by imposing the form of law on psychologically generated incentives. And here as elsewhere in Kant's theory we are godlike in the ability to impose the form of law on the world; but we are not godlike to the extent that we do not ourselves generate the matter on which that form is imposed. Acting from respect for law is indeed as close as we come to the divine ideal of intellectual intuition—in that case we generate both form and matter, law and incentive, from our own activity. Ordinary permissible action is more like the operations of the understanding—we are genuinely active, not passive, although it is on a material that is given to us.

So far, I've just been talking about the standing of a single incentive of inclination. But Ginsborg brings up a different issue when she evokes the situation of choice between several ends of inclination. On Ginsborg's reading of Kant, the choice between two ends, when not settled on moral grounds, is determined by their contributions to happiness. Happiness, she thinks, is not a chosen end. Rather, as Kant says when he is introducing the idea that there is an imperative of prudence, it is an end that "we can safely presuppose [everyone] actually [does:] have, by a natural necessity."<sup>20</sup> As Ginsborg sees it, a person pursues happiness, like someone following a recipe, normatively guided by a sense of what will fit together well. And Ginsborg urges that one may be normatively guided in this way even if one hasn't chosen the end of happiness, just as one may be normatively guided by judgment when making an omelet, regardless of whether one chose to make the omelet or not. Ginsborg speculates, correctly, that I think there is a problem about all this, but it is not, as she also speculates, primarily a problem that arises from the indeterminacy of happiness.

In fact I think there is a whole knot of problems here. In the first place, I am completely mystified by Kant's claim that we "have" the end of happiness by a natural necessity. He might mean either that we necessarily *will* happiness, or that we necessarily *desire* it, but there are difficulties

19. See also Korsgaard, *The Sources of Normativity*, esp. 3.4.7–3.4.10, pp. 120–25, and 4.4.2, pp. 161–64.

20. Kant, *Groundwork*, 4:415.

either way. On the one hand, if he means that we *will* happiness by a natural necessity—that is, we choose it—his claim is contrary to his views about our essential freedom of the choice of ends. In the *Metaphysics of Morals* Kant says: “An end is the object of . . . choice. . . . Now, I can indeed be constrained by others to perform *actions* that are directed as means to an end, but I can never be constrained by others *to have an end*; only I can *make* something my end.”<sup>21</sup> Adopting an end is an internal action to which we cannot be compelled; hence it is an act of pure internal freedom. Of course there is no question of others forcing us to have happiness as an end, but Kant does seem to say here that our ends are objects of free choice, and on my understanding of his moral psychology, anyway, that’s what he must say. How are we to square the idea that we will happiness by a natural necessity with this freedom in the choice of willed ends? On the other hand, if all Kant means is that we can’t help but *desire* happiness, it is puzzling that he singles out a special sort of imperative, the imperative of prudence, which guides our pursuit of this desired end. For in the first place, there are many things, most notably the satisfaction of our physical needs, which we cannot help but desire, but Kant does not single out special imperatives for them. In the second place, mere desires for ends do not support hypothetical imperatives, which are based on the principle that whoever *wills* an end wills the means, and therefore cannot be derived from mere desires. And in the third place, the mere *desire* for happiness would only be one desire among others, which would have to compete for our attention with the more particular desires and ends. In fact, even if Kant did have an argument that we must *will* happiness as an end, it wouldn’t automatically follow that we should always rationally prefer it to more particular ends; nor does Kant give any argument at all to that effect. And if there were a principle of practical reason according to which it is rational both to *have* happiness as an end and to *prefer* it to every other end, that principle would in a way be in between Kant’s categories of hypothetical and categorical imperatives. Unlike a hypothetical imperative, it would assign us an end, but unlike a categorical one, it would hold only conditionally, since our pursuit of this end would have to give way to moral considerations.

Ginsborg’s view that when Kant says happiness is a naturally necessary end he means that it is not a freely chosen end enables Kant to escape some of these last complaints, but on the most intelligible reading of what he might mean by it, it is quite implausible. On the most intelligible reading, what Kant would mean is that once we do see clearly what happiness requires, that is always what we in fact prefer. Ginsborg’s own analogy between the pursuit of happiness and the making of an omelet

21. Kant, *The Metaphysics of Morals*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 1996), 6:381.

suggests this thought, for when we are making an omelet, it is largely true that we need only to know which of the available ingredients will make the omelet turn out well and we forthwith want to put them in. Most of us do not suffer from an obscure compulsion to put too much pepper in our omelets even when we know that that is going to spoil them. But wanting to put more pepper in your *life* than it can easily contain is quite another matter. I am rather inclined to side with Bishop Butler on this one, when he reminds us that “Men daily, hourly sacrifice the greatest known interest to fancy, inquisitiveness, love, or hatred, or any vagrant inclination.”<sup>22</sup> If what Kant means by saying that happiness is a naturally necessary end is that we always in fact prefer it, the evidence is simply against it. People sacrifice their happiness to love, to ambition, to political aspiration, and to simple temptation all the time. Of course someone might say that this must be because they really think this is what will make them happy, but the threat of making the claim that we desire happiness completely trivial then looms.

For all of these reasons, I find Kant’s claim that we have happiness as an end by a natural necessity and his claim that there is a rational imperative of prudence perplexing. Admittedly, one possible explanation of Kant’s view is suggested by the arguments I detailed earlier about the role of the principle of self-love in our moral psychology, and the biological roots of that principle. Those arguments do give a sense in which we might be said to value *ourselves* by a natural necessity, and this valuing of ourselves is certainly connected to having a general concern for, even commitment to, the satisfaction of our inclinations.<sup>23</sup> Kant may have had these views in mind when he claimed that we value happiness by a natural necessity. But a general commitment to satisfying one’s inclinations is not the same as a commitment to, say, maximizing the sum of their satisfaction over time, or whatever it is that the principle of prudence is supposed to direct. Some further argument would have to be given for that.

But in any case, my main reason for disagreeing with Ginsborg doesn’t lie here. It lies in my conviction that, on the Kantian story, self-conscious action and choice are essentially free, since self-conscious action takes place “under the idea of freedom.” Ginsborg understands this to mean only that we are aware of our freedom not to satisfy an inclination if it conflicts with duty. In support of this, Ginsborg gives a reading of the notoriously difficult argument about the “circle” in *Groundwork*

22. Joseph Butler, in the preface to the *Fifteen Sermons Preached at the Rolls Chapel*, reprinted in *Five Sermons Preached at the Rolls Chapel and A Dissertation upon the Nature of Virtue*, ed. Stephen Darwall (Indianapolis: Hackett Publishing Co., 1983), p. 21.

23. I don’t, however, take these arguments to show that we value ourselves inevitably or unfreely; I think that the freedom inherent in the standpoint of practical reason gives us the option of not valuing ourselves, although the price of taking that option is normative skepticism. For these ideas, see Korsgaard, *The Sources of Normativity*, 4.4.2, pp. 161–64.

III<sup>24</sup> which tends to bring it closer to the Fact of Reason argument in the *Critique of Practical Reason*<sup>25</sup> than most interpreters have thought. Ginsborg's reading has textual merit, although I am still inclined to think that the argument of the *Critique of Practical Reason* represents a change of mind on Kant's part. My own favored version of the foundational argument, spelled out in "Morality as Freedom," contains elements drawn from both texts but leaves "acting under the idea of freedom" as the independent starting point of the argument.<sup>26</sup> But leaving the interpretive issue on one side, I believe that the Kantian thesis that we act under the idea of freedom is a deep insight about the standpoint of self-conscious choice and does not depend specifically on morality. It gives rise rather directly to the demand that all of our choices be autonomous. And although it has taken me a long time to see this, I now think it follows from this demand, rather surprisingly, that there is really only one principle of practical reason—the principle that we should choose our maxims as universal laws. A requirement of instrumental rationality still appears as a sort of subdivision of this principle, since your maxim cannot *be* a practical law if the act you propose to do will not actually tend to bring about the end you propose to achieve. But prudence does not seem to be, in this way, built into the very idea of a free rational action. Its status as a demand of rationality therefore remains obscure.

\* \* \*

Like Hannah Ginsborg, Paul Guyer focuses primarily on the line of argument I present in "Kant's Formula of Humanity" and "Morality as Freedom." Guyer's worry is less about whether I have managed to capture an authentically Kantian line of argument than about whether that line "does full justice to that aspect of Kant's approach to moral philosophy that renders his picture of moral reasoning most compelling."<sup>27</sup> This is because Guyer doubts whether the argument as I have reconstructed it "represents Kant's best effort"<sup>28</sup> to meet the condition, accepted by Kant himself, that moral philosophy should "rest on concepts and insights readily available to all."<sup>29</sup> As Guyer sees it, I have reconstructed a subtle and essentially metaphysical argument by drawing out the logical consequences of the abstract concept of a rational being, in particular of the fact that reason as such seeks the unconditioned. By applying this abstract concept of reason to the realm of value, I argue for the conclusion that human beings must see ourselves as value-conferring and must

24. Kant, *Groundwork*, 4:450.

25. Kant, *Critique of Practical Reason*, 5:29–31.

26. In Korsgaard, *Creating the Kingdom of Ends*, p. 175.

27. Paul Guyer, "The Value of Reason and the Value of Freedom," in this issue, p. 23.

28. *Ibid.*, p. 28.

29. *Ibid.*

therefore value humanity as an end in itself. According to Guyer, Kant's argument so understood is based on "philosophers' notions par excellence" and is not suited to make the demands of morality persuasive to "ordinary human beings." Guyer thinks that Kant and his interpreters do better when we focus on the immediately recognizable notion that our freedom gives us a special kind of dignity which we can live up to only through autonomous action.

I should be sorry to think this criticism correct, since I take it to be one of the hallmarks of Kant's philosophy, and one of its most attractive features, that Kant sees philosophical reasoning as absolutely continuous with ordinary reasoning. Philosophy is ordinary human reasoning rendered *persistent*. In a sense, I think this is the whole point of the doctrine that reason seeks the unconditioned. Kant's view, as I understand it, is that a person who starts out reasoning in some perfectly ordinary way—attempting to give a causal explanation of an event, or to give a rational justification of an action, or to distinguish a thing's accidental properties from what is really essential to it, or something of the sort—that such a person finds himself on a route that has no natural stopping place short of the unconditioned Ideas of Reason and the metaphysical perplexities to which they sometimes lead. The claim that reason seeks the unconditioned is not based on an analysis of the abstract concept of reason. It is more a claim about the plight of self-conscious beings who because we are self-conscious need reasons to believe and to act. When we go looking for those reasons we find ourselves—via a form of regress argument that is perfectly natural to any rational being—on a road that leads to the unconditioned, a road that threatens to have no satisfactory stopping point.

I therefore do not agree with Guyer when he characterizes Kant as holding that the dialectical inferences of reason were "invented by philosophers who were driven by reason but were also ignorant of its limits."<sup>30</sup> I take Kant at his word when he says, in the opening lines of the first *Critique*, that he is discussing the peculiar fate of *human reason*.<sup>31</sup> It is true, as Guyer says, that "Kant does not treat the principle of morality in the same way"<sup>32</sup> that he treats the dialectical problems of speculative metaphysics, but that is for a substantive philosophical reason. Practical reason represents the one domain in which we can *reach* the unconditioned: the moral law is human reason's great success. Kant's view is that we cannot find the kind of ultimate intelligibility sought by reason in the theoretical domain, but we can create it in the moral one. To put it another way, the kingdom of nature cannot be *known* to be a kingdom of ends, but we can do what is in our power to *make* it one.

30. *Ibid.*, p. 29.

31. Kant, *Critique of Pure Reason*, Avii.

32. Guyer, p. 29.

So I take Kant's argument for the Formula of Humanity to proceed not from the abstract concept of reason, but from the standpoint of the deliberating rational agent. Guyer thinks what I mean by an argument made from the practical point of view is "an exploration of the consequences of our *choosing* to conceive ourselves as rational beings."<sup>33</sup> But I don't mean that, for I don't think we have a choice about that. Our plight as self-conscious beings is that we find ourselves faced with the necessity of making choices and so in need of reasons to act. Kant's argument, as I understand it, is intended to start from that plight, to proceed from the standpoint of a being faced with the need for reasons and principles. This shows up most clearly in the foundational argument as given in the opening pages of *Groundwork III*. In that argument Kant sets up the *situation* of the free will and shows us how we can move from that situation to the moral principle: "As a will it needs a law, but as free it cannot be determined by a law outside itself, so what law is to determine its actions? It must be an autonomously given law." In the same way, I understand the argument for the Formula of Humanity to start from the deliberative situation of a person making a choice, and to show how that person is moved to the realization that he regards *himself* as the source of value. I don't think this reasoning need be conceived as a piece of abstruse metaphysics. Imagine it this way: "Why in the end does it matter that I achieve this end? Because it matters to me, and I matter. . . ." Certainly I do not mean to deny that Kant also appeals to the special dignity inherent in autonomy. But in the argument of the *Groundwork*, the notion of humanity as a source of value is prior, for the notion of autonomy is arrived at by combining the Formulas of Universal Law and Humanity:<sup>34</sup> it is the notion that we human beings legislate the good for ourselves.

Guyer also thinks that when I claim that evil is in a particular way unintelligible, I am misinterpreting the Kantian thesis of the inscrutability of our fundamental choices.<sup>35</sup> He supposes I think there is a "logical puzzle how an evil act can be both free and yet undermine freedom."<sup>36</sup> This isn't what I had in mind. On Kant's view, there is a sense in which the choice of any fundamental maxim, good or evil, is inscrutable. Such a choice cannot of course be explained, since it is a free choice and therefore not subject to causal explanation. Nor can we exactly give a reason for it, since this is the fundamental choice that determines what the person is going to count as a reason and so no reason can be prior to it. But there is a further sense in which the choice of evil is unintelligible. The argument I reconstructed in "Morality as Freedom" is supposed to show that the moral law *is* the law of a free will. It is the nature of a free will that it chooses autonomously, and that is what the moral law tells us

33. *Ibid.*, p. 27.

34. Kant, *Groundwork*, 4:431.

35. See Korsgaard, *Creating the Kingdom of Ends*, pp. 171 ff.

36. Guyer, p. 35.

to do, so the moral law simply describes a free will's nature. To that extent a person with a free will does not need a reason to act on the moral law. But she does need a reason to deviate from the moral law. I think it helps here to keep the fact that Kant argues from a standpoint in mind here, for in a way the point I want to make is best captured by saying that we cannot imagine the free choice of evil *from the inside*, from the point of view of the one who makes that choice. It is exactly parallel to the problem of trying to imagine akratic choice from the inside.

In fact, I now believe that the point is a general one applying to all principles of reason and that what it illustrates is that the principles of reason are internal standards of thought and action. What I mean by this is that the principles of reason arise from the very nature of activities which they govern and represent standards that are internal to those activities. Thought that violates the principles of logic is defective not just because it fails to track an external truth but *as thought*—it is something less than thought. And heteronomous action is, in a similar way, something less than action. The problem of inscrutability is common to all forms of irrationality. Although someone can believe a contradiction, or will heteronomously, or fail to will the means to her ends, such a person is inscrutable to us: we cannot quite imagine her situation from the inside.<sup>37</sup>

\* \* \*

The first of J. B. Schneewind's criticisms is also concerned with the questions of the kind of value humanity has in Kant's ethics and what gives it that value. Schneewind notes, correctly, that in some of the essays in *Creating the Kingdom of Ends*, I talk about the value of humanity as if it were a form of intrinsic value, distinguishing the value of humanity in this way from the value of our particular ends. These, I claim, have only extrinsic value, since their value depends on our own desires and interests in them and is conferred on them by our own rational choices. More recently, especially in *The Sources of Normativity*, I have come to think of the value we place even on ourselves as also conferred. To that extent, I agree with Schneewind that there is a continuity between the value of humanity and the value of other things: they are all the result of our own acts of conferring value.

Schneewind, however, thinks that it follows from what Kant says in the second *Critique* about the precedence of the law over the good<sup>38</sup> that the value of humanity is in no way different from the value of any other

37. For these ideas, see my "The Normativity of Instrumental Reason," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Clarendon Press, 1997), pp. 215–54, esp. p. 249, where I argue that the instrumental principle is an internal standard; and "Self-Constitution in the Ethics of Plato and Kant" (forthcoming in *The Journal of Ethics*), where I argue that the categorical imperative is an internal standard.

38. Kant, *Critique of Practical Reason*, 5:58–66.



object of practical reason: all of the objects of practical reason are unconditionally valuable, because they get their value from the law. But I think that there is an absolutely essential difference. As the argument I gave earlier suggests, if we do not set a value on ourselves—if we deny the principle of self-love in its moralized form, the Formula of Humanity—then nothing else can have any value either. My argument for that point is not based on the traditional metaphysical argument that there is nothing in the effect that is not somehow in the cause, as Schneewind suggests. I am aware that in early papers I made it sound too much as if value were some sort of metaphysical substance that got transferred from us to our ends via the act of choice. As I've tried to clarify here, I don't think that: I think that self-love, valuing oneself, is essentially involved in the valuing of our ordinary ends. In this sense the value of humanity is the unconditioned condition of all other value, while the other ends of practical reason are not. I don't find the question of the relative precedence of the moral law and the value we place on ourselves intelligible, because I think that lawmaking is the activity by which we create values and that every act of lawmaking expresses the value we place on ourselves.

Schneewind's real interest in attacking the idea that value must spring from an unconditional source, however, is a more general doubt about the role of the unconditional in Kant's ethics: that is, about whether a Kantian need be committed to acting on unconditional reasons or an unconditional law. Schneewind says that Kant has "a complex account of reason's tendency to seek the unconditioned in theoretical knowledge" and wonders what I think about that; and he adds that Kant had "an additional reason to search for an unconditional rational principle in morals," namely, the view that it is inconsistent with human dignity for us to simply obey divine commands in a spirit of humble obedience.<sup>39</sup>

As I have already suggested in my response to Guyer, I think that the unconditioned plays an essential role in Kant's philosophy throughout, for I think that Kant derives the ideas of reason from the idea of the unconditional: the ideas of God, Freedom, and Immortality, which are essential to both theoretical and practical thought, are arrived at by seeking the unconditioned in regresses created by the disjunctive, hypothetical, and categorical syllogisms, respectively. I think that in the same way the idea of obligation, embodied in the Categorical Imperative, is an idea of reason arrived at by a justificatory regress. I do not think Schneewind is correct in describing it as a mere "theoretical fact" that reason seeks the unconditioned: it is an essential feature of Kant's conception of reason.

So let me address the additional reason for being concerned with

39. J. B. Schneewind, "Korsgaard and the Unconditional in Morality," in this issue, p. 41.

the unconditioned which Schneewind finds in Kant's views about human dignity and the issue of divine command. I have already mentioned a sense in which I think we do achieve a sort of divine status in acting from the motive of duty in Kant's system: action motivated by respect for law is pure activity, like intellectual intuition. But more importantly, I think that on a Kantian view it is not only inconsistent with human dignity for us to follow moral laws given to us by God; it is strictly speaking impossible for this to be all that morality comes to. Given that we must act under the idea of freedom, our obedience to the laws of God must depend on our acceptance of a prior law, namely, the law that "I ought to conform to the laws of God." According to a familiar argument against voluntarism, this law cannot itself intelligibly be one we follow because it is commanded by God. So it must instead be something we acknowledge to be a law *in itself*. And that in turn must mean, according to Kant, that it has the intrinsic form of a law. Since the categorical imperative tells us to conform to maxims which have the intrinsic form of a law, that would have to mean that the law of obedience to God follows from the Categorical Imperative and that in turn means it must be willed autonomously by the agent. So a system of theological morality cannot get off the ground unless the maxim of obedience to God is autonomously willed by the agent in her own sinful pride. In other words, Schneewind is absolutely right about the nature of my claim: I am saying that if you are truly a rational agent, you must accept Kantian morality. These claims are related to those I made at the end of my response to Paul Guyer: the Categorical Imperative is an internal standard of free rational action as such.

Schneewind suggests as opposed to this that we must start somewhere, with something not unconditioned but simply accepted, and he provides a battery of arguments for thinking so, of which I will mention only a few. He reminds us of Rawls's theory of reflective equilibrium, which tests proposed moral principles by their coherence with our most considered moral convictions, and suggests that Kant's appeal to the spectators' approval of the Good Will functions analogously as a test on Kant's principle. But Kant appeals to the spectators' approval of the Good Will in the first section of the *Groundwork*, when he is professedly presenting the material analytically—that is, he is analyzing ordinary rational knowledge of morality—and Kant's doctrine in general is that conclusions reached by the analytic method are ultimately to be vindicated by synthetic arguments, such as those in *Groundwork III*. I see no reason for Kant to accept the idea of reflective equilibrium except as an aid to making sure that we have correctly identified the principles we hope to justify, but then—another story—I see no reason for Rawls to do so either.

Schneewind also invokes Leibniz's view that we must accept some reasons for our actions, including the action of considering reasons, without giving them critical consideration. I hope the description of

Kant's moral psychology which I gave earlier makes it clear why I disagree with that: I think that the need to *consider* reasons is forced upon us by our situation as self-conscious agents and, therefore, that we do not need a reason for doing that. We can do it well or badly, but that's another story.

Finally, Schneewind also mentions the idea that an agent must start from the basic presupposition that desires and motivating commitments give her reasons unless there are reasons why not: otherwise deliberation will be endless. As will be obvious from what I have said about the principle of self-love, there is a sense in which I agree with this. But the "reason why not" that may block an inclination from serving as a reason need not be some other reason. It can instead be the fact that the maxim of acting on the inclination cannot serve as a universal law. On a Kantian account, it is not true that what blocks the force of a proposed reason is always some competing reason. A consideration may fail to have the *form* of a reason. And the difficulty with denying normative standing to our natural impulses in general is not, as Schneewind suggests, that if we do that deliberation will take forever: it is that if we deny normative standing to our natural impulses there will be no source of *content* for our reasons at all.

I assume that Schneewind would not really disagree, since he agrees with me that we should find the reasons we can share, and those will be ones that can serve as laws. But he thinks they need only serve as laws for the group of people actually affected by the action, and also that we may need some sort of Habermasian open discussion to determine what the shareable reasons are. I am tempted by this to think that Schneewind's real objection to searching for the unconditional is less that he thinks it an inappropriate standard than that he thinks it is a tough one to meet, and that Kant didn't succeed in showing us how to do it. Perhaps this is because Schneewind thinks that the argument Kant uses to show that the Categorical Imperative is the fundamental law of our agency doesn't work, or perhaps it is because he thinks the Categorical Imperative doesn't always give us a determinate answer to the question about what we have reason to do. I am in sympathy with the second of those worries, but not with the first. In my own view, the argument that Kant uses to show us that the Categorical Imperative is the law of free action is successful, and combined with the argument that we must act under the idea of freedom it shows that nothing short of a principle that can be autonomously willed can justify an action. To all of the fans of the embedded, the pragmatic, the contextual, and so on, who are always insisting that *justifications must come to an end somewhere*, Kant would answer that justifications can come to an end only with a law you yourself will, one you'd be prepared to will for everyone, because justifications must come to an end with you—with the dictate of your own mind. And in this, I stand with Kant.