DOI: 10.1111/phpr.70041

Philosophy and Phenomenological Research

ORIGINAL ARTICLE

# The Kantian's Desire

**T. A. Pendlebury**

University of Chicago, Chicago, Illinois, USA

**Correspondence**
T. A. Pendlebury, University of Chicago, Chicago, Illinois, USA.
Email: tapendlebury@uchicago.edu
t.a.pendlebury@gmail.com

**Abstract**

Christine Korsgaard's oft-quoted image of the reflective agent "backing up" from her desires has been criticized on the ground that it depicts desires as items external to practical reason which can, therefore, have no normative bearing on her rational activity. I argue that her critics have failed to recognize the possibility of a different explanation of the deliberative import of desires: that even if their origin is external to reason, reason confers deliberate import upon them. I develop this alternative as a reading of the moral psychology of Kant, whose work informs that of both Korsgaard and her critics, and consider the question whether Korsgaard embraces it too.

## 1 | INTRODUCTION

In recent years, many philosophers otherwise sympathetic with her general approach have taken issue with the image of "backing up" which Christine Korsgaard uses to illustrate her conception of rational mindedness:

> [O]ur capacity to turn our attention to our own mental activities is also a capacity to distance ourselves from them, and to call them into question. [. . . ] I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? [. . . ] The reflective mind cannot settle for. . . desire, not as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward. (Korsgaard, 1996, p. 93)

These critics object to the image because, they say, it suggests that I relate to my desires as alien, brute impulses having nothing to do with my own rational mindedness, unfit to serve as even "candidate, provisional, or presumptive reasons" (Ebels-Duggan, 2023, p. 166). Many of them, like Ebels-Duggan (2023), Boyle (2016), and Moran (2022), also suggest that Korsgaard does not really mean to depict them in this way. Indeed, Fix (2022) goes so far as to claim that the point of the passage is that "to 'step back' is to reflect on the exercise of practical reason to make it explicit and articulate" (p. 656). Thus, he suggests, the passage does not really concern the relation between two things—reason and desires, things other than reason—at all.

In so arguing, these critics distinguish between two different conceptions of desire. The one they oppose has the characteristics of what Tamar Schapiro has called a "brute force" view of inclination, according to which "when I am inclined, but not thereby determined, to $\varphi$, I am relating to a brute force that has its source outside of me" (Schapiro, 2021, p. 43).[1] The view they prefer, which I shall call *transformativism*, is one on which my desires are themselves expressions of my rational activity, akin or equivalent to judgments of practical reason but with certain affective features.[2] Among the reasons given for preferring the latter to the former is that it makes it possible to explain how desires could have any import for practical reason and, more specifically, could be things that you could "act on" rather than merely respond to, as—in Schapiro's (2021) words—features of your circumstances (p. 49). Unless desires are products of reason, many of these authors appear to believe, there is, as Boyle (2016) puts it, an "interaction problem that renders it mysterious how a rational agent's desires can provide her with reasons to act" (p. 538).

I agree with the transformativists that Korsgaard may have a problem. But if she does, I think they have misdiagnosed it. They think that it lies in her conception of desires: that they are not themselves borne of reason. I think that it lies in her conception of practical reason, or the rational will. Once that problem is fixed, the problem with desires is fixed automatically. To explain what I mean by this, I shall develop a reading of the moral psychology of the historical antecedent both of Korsgaard's view and of those of her critics: Kant. On this reading, Kant shares Korsgaard's conception of desires except insofar as he has a different conception of practical reason, one which explains the possibility of acting on desires in a non-transformativist way. Having described the explanatory structure of Kant's account, I consider whether a more generic version of it might be available to Korsgaard. I shall suggest that the answer is not obvious and that this may be part of why she has been targeted by the transformativists.

And, finally, I shall suggest that transformativism is more rationalistic in spirit than Kant's philosophy. The relation of our sensitive or affective nature to our intellectual or rational nature is a perennial topic of philosophical reflection. Both in the treatment of Kant and in contemporary systematic discussion, it seems often to be assumed either that these things are independently intelligible, so that we are two things welded together; or else that our sensibility is a branch or department of reason, so that we are psychically homogeneous through and through. Kant's actual position, as I understand it, is a compelling third alternative, whose upshot is that while exercises of our sensibility—being, as he puts it, blind (A51/B75)—need not be informed by our reason, our reason is in part constituted by its relation to our sensibility, this being part what makes it finite reason, which in turn determines sensibility's functional profile in us.

I proceed as follows: Section 2 introduces level one of what I call Kant's *two-level* explanatory strategy. Section 3 concerns the interaction of this specification with another fundamental

[1] I mean the same thing by (countable) "desire" and "inclination." Since the latter is the usual translation of Kant's "*Neigung*," I use it more often than "desire."

[2] This label is derived from the title of Boyle (2016).

element of his moral psychology: the government of the finite will by the moral law.[3] Section 4 considers level two, and Section 5 returns to its relation to morality, considered this time at level two. In Section 6, I turn to Korsgaard's moral psychology, which I consider in light of Kant's. Section 7 concludes with a final reflection on the difference between Kant's view, as I read it, and the rationalist presuppositions of transformativism.

A note before I begin. This paper is about both Kant and Korsgaard because the moral psychological discussion to which it responds is about, inter alia, both Kant and Korsgaard: both figures are of interest to all the critics I have named above and to many others still. I develop and defend readings of both of these figures because I believe that their explicit comparison is both textually and philosophically illuminating. Sustained treatment of either one, to the exclusion of the other, would naturally allow for greater detail in the treatment of this or that constituent subtopic. But the appropriate level of detail is always relative to the task at hand. My task is, in part, their comparison. I aim for a level of detail appropriate to that task.

## 2 | KANT'S FIRST LEVEL: HAPPINESS

In a way, it is easy to say what makes for the normative bearing, or, as I shall put it in what follows, *deliberative import* of inclinations, according to Kant. It is the necessary end of happiness:

> There is, however, *one* end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely *could* have but that we can safely suppose they all actually *do have* by a natural necessity, and that purpose is *happiness*. . . .a purpose that can be presupposed surely and *a priori* in the case of every human being, because it belongs to his essence. (Grundlegung 4:415-16)

The import of Kant's claim in this passage is a relatively neglected feature of his moral psychology.[4] That human beings wish for happiness is not something we know merely empirically: it is something we know *a priori* about all dependent, or finite, rational beings, and thereby about human beings as a species of dependent rational beings (where this means that imperatives apply to us).[5] Moreover, we human beings have this end because all dependent rational beings have it and because we are dependent rational beings. Nor is this a one-off claim: for in the second *Critique*, in a passage to which I return later, Kant he says that "[t]o be happy is necessarily the demand of every rational but finite [*endlich*] being" (KpV 5:25).

Indeed, this end is as deeply entrenched in finite, or dependent, rational moral psychology as is another, rather better known and appreciated, feature of the finite rational will: its governance by the moral law in the form of duty. Even Kant's formulations in the case of happiness and in the case of duty are similar, down to the parenthetical expression qualifying "rational" with "dependent":

[3] For the purposes of this essay, I assume that Kant's published critical texts—and in particular the *Groundwork*, second *Critique*, and *Religion*—present a single doctrine concerning the relation of happiness to morality. (This leaves room for differences in detail and emphasis on this point.)

[4] A significant exception to this is Engstrom (2009), whose very enlightening discussion of what he calls the "wish for happiness" has informed my understanding decisively.

[5] The notion of a finite will contrasts with that of the kind of will—an infinite will—to which imperatives do not apply because error is impossible for it (Grundlegung 4:414). For contemporary discussion, see Lavin (2004) and Fix (2020).

"duty…must hold for all rational beings (to which alone an imperative can apply at all) and *only because of this* be also a law for all human wills" (Grundlegung 4:425). Again, something is true of human beings because it is true of dependent rational beings and because we are dependent rational beings. Being a finite rational being, then, is sufficient both for actually having your own happiness as an end and for being subject to the moral law. One and the same thing—our being finite rational beings—is sufficient for both of these features of our rational lives.

But what exactly is happiness? Kant defines happiness in various ways, but the satisfaction of inclinations is involved in most of these definitions. Happiness is hardly ever simply identified with the satisfaction of inclinations: rather, happiness is the satisfaction of something like a system of ends identical to the objects of inclinations, subject to various constraints, such as joint realizability and attention to the temporal extendedness, the "whole existence" of the agent (KpV 5:22-25), so that her various purposes are "unite[d]" to her "own enduring advantage" (Grundlegung 4:416n.), and so that this can be glossed in terms of her "well-being in [her] present condition and in every future condition" (Grundlegung 4:418). In light of these constraints, with their focus on the future, it is unsurprising that Kant calls imperatives that tell you to take the means to your happiness imperatives of *prudence* (Grundlegung 4:416). I shall call these requirements *standards of systematicity* and shall say that your happiness is the *systematic satisfaction* of your inclinations.[6] Because happiness is a system of ends, it is an end consisting of ends in systematic connection: as I shall abbreviate this idea in what follows, it is big end with little ends in it.

It is an interesting question exactly what these standards amount to, according to Kant, and indeed whether his views about their content are consistent across his texts. But for present purposes, we need not answer this question. For whatever exactly happiness is, it is consistently defined in terms of the satisfaction of inclinations. Thus, if your own happiness is your end, this means that your inclinations have deliberative import. For ends are modifications of your will. Insofar as something is your end, you are volitionally oriented toward its realization. According to Kant, the systematic satisfaction of your inclinations is your end as long as you are a finite rational being. So the systematic satisfaction of your inclinations, and to this extent your inclinations themselves, are of deliberative import to you. Not every inclination is guaranteed to satisfy the standards of systematicity. And even some of those which do will not make it into your conception of happiness, because you will almost certainly have to make choices. But every inclination is eligible for assessment according to those standards, because it is your end to pursue the objects of your inclinations subject to their meeting those standards.[7]

[6] Kant also associates the satisfaction of inclinations, and happiness, with pleasure (KpV 5:22), so that it is unsurprising that Lewis White Beck, for example, writes that "happiness is our concept of a systematic whole of pleasures" (1960, p. 98). But this does not commit Kant to thinking that what we are inclined to do is secure pleasure. See Timmermann (2022, pp. 10–15) for discussion.

[7] That you have the end of happiness by nature does not entail that you know what means to take to realize it—that we characteristically don't know this is a favorite theme of Kant's (e.g., Grundlegung 4:417-19)—and it is controversial whether it entails that you in fact take the means when you know what they are. While Korsgaard, for example, argues that true instrumental irrationality is possible according to Kant (Korsgaard 2008, 50f., 2009, pp. 68–72)—why else would he have called the hypothetical imperative an *imperative*?—other authors, like Timmermann (2022, pp. 56–59), argue that it is not. I cannot entitle myself to either position in this essay, but I agree with Korsgaard: even knowing what it would take to pull off an end does not entail that you will (even try to) take the means, and this applies to happiness as much as to any other end. That means that the end of your own happiness is normative for you, though it is not an end assigned to you by any principle of end-selection. In claiming that your ends do not have normative status except insofar as they are assigned to you by such a principle, Korsgaard departs from Kant (Korsgaard 2006, p. 63).

That is Kant's explanation. And it is not a transformativist explanation. Your inclinations need not themselves be borne of reason to be eligible for consideration according to the standards of systematicity. For inclination-satisfaction, subject to those standards, is the end of your will, and this all by itself explains the deliberative import of inclinations.[8]

## 3 | HAPPINESS AND DUTY

### 3.1 | The basic picture

But we do not really understand this explanation except insofar as we connect it to other features of Kant's moral psychology, and in particular to the more often emphasized one I have already mentioned. The moral law is a principle of end-selection. It tells you which ends you may adopt. Obviously, there are questions about the relation of this principle to the end of happiness. For example, is it—the end of happiness—a permissible end? Since it is naturally necessary, you cannot give it up; if *ought* implies *can*, it would seem that it cannot be impermissible. On the other hand, for the same reason, it seems at the very least silly, and perhaps incoherent, to call it permissible.

But these are not the only options. For recall that happiness is a big end with little ends in it. These ends are chosen subject to the standards of systematicity. We can understand the moral law as applying to these little ends, as a constraint additional to the standards of systematicity. Since Kant says that you invariably pursue your happiness but do not invariably observe the moral law, we have to distinguish between two cases. The *good agent* fills in her conception of happiness subject to the standards of systematicity and the moral law. The *bad agent* fills in her conception of happiness subject to the standards of systematicity alone, without heeding the law.[9]

Here is an example of a passage which bears this out.[10] In this passage, Kant is telling us what it is for observance of the law, rather than any matter of the desiderative capacity—that is, rather than any inclination—to be the determining ground [*Bestimmungsgrund*] of the will (KpV 5:34):

> [T]he mere form of a law, which limits the matter, must at the same time be a ground for adding this matter to the will, but not for presupposing it. Let the matter be, for example, my own happiness. This, if I attribute it to each (as, in the case of finite beings, I may in fact do), can become an *objective* practical law only if I include in it the happiness of others.…[A]nd so the object (the happiness of others) was not the

[8] Thus Kantian inclinations are akin to what Marcus has called "orectic" desires: they neither originate in, nor are identical to, practical judgments or representations of reasons or the good (2020, pp. 890–893). But with the end of happiness, which belongs to the will, Kant has an explanation of why inclinations mean anything to the will; whereas Marcus, though he says that orectic desires have influence on and "prod" the will, does not address the question of their deliberative import.

[9] The terms "good agent" and "bad agent" are labels. There is more to success in action than filling in your conception of happiness subject to the standards of systematicity and the moral law. If you are instrumentally irrational, unskilled, or cursed by unfavorable circumstance, you may fail to secure your ends, however perfectly they have been selected. But Kant singles out the pursuit of happiness heedless of the law as a structurally distinct kind of practical failure, one to which we are tempted in a way in which we are not tempted to instrumental irrationality, lack of skill, or unfavorable circumstance. It is this kind of practical failure with which I am concerned here.

[10] Like any reading of any passage in Kant's texts, this one will be controversial. Nonetheless, I submit that it will be very obviously grounded in this and other texts in the second *Critique* and the *Religion*, though there is always more than one way to read almost any passage.

> determining ground of the pure will; this was, instead, the mere lawful form alone, by which I limited my maxim based on inclination [*meine auf Neigung gegründete Maxime*] in order to afford [*verschaffen*] it the universality of a law and in this way to make it suitable for pure practical reason… (KpV 5:34-5)

To understand the passage, we have to understand the terms of art used in it. First, a *maxim based in inclination* is a *material practical principle*. For "practical *principles* are propositions that contain a general determination of the will" which are "subjective, or *maxims*, when the condition is regarded by the subject as holding only for his will" (KpV 5:19). Such principles are material, or empirical, when they "presuppose an *object* (matter) of the desiderative capacity as the determining ground of the will," an "object whose actuality is desired" (KpV 5:21). This is to say that they presuppose an object of an inclination.

Kant has already said at this point that "all material practical principles as such are, without exception, of one and the same kind and come under the general principle of self-love or one's own happiness" (KpV 5:22). The principle of self-love he also describes as the principle of making happiness "the supreme determining ground of choice" (KpV 5:22). This principle is a principle of the will: for "[p]ractical *principles* are propositions that contain a general determination of the will, having under it several practical rules" (KpV 5:19). It is, however, a merely "subjective" principle, one "regarded by the subject as holding only for his will" (KpV 5:19). That this is a principle of my will is the same as that of my own happiness is my end.[11] Thus, material practical principles—maxims based on inclination—"come under" the end of happiness. This coheres with Kant's definitions of happiness in terms of inclination-satisfaction. And it corroborates my claim that inclinations have deliberative import by virtue of the end of happiness. A maxim—a determination of the will, by virtue of which something is my end—can be based in inclination only insofar as it comes under the end of happiness.

Now for Kant's distinction between "adding matter to the will" and "pressupposing it." To add matter to the will is to adopt an end. The law, Kant says, is the ground of adding matter to the will: you may adopt ends only insofar as they conform with the law. But the law is not the ground of "presupposing" this matter. By "presupposed" matter I take Kant to mean matter available for addition to the will by virtue of some inclination, in keeping with KpV 5:21, quoted above. Thus, the ground for presupposing matter is that it is the object of an inclination which comes under the end of happiness. In these terms, we can redescribe the good and bad agents. The good agent treats the law, something additional to the standards of systematicity, as the ground of adding matter to the will: she treats the law as the ground which it indeed is. The bad agent treats the standards of systematicity as on their own sufficient grounds for such addition. She is heedless of the law.

An important implication of this is that while Kant says that every finite rational being has as her end the systematic satisfaction of her inclinations, it must be that his conception of what it is to have this end allows that there are different ways of having it, both the way in which the good agent has it and the way in which the bad agent has it: as qualified or unqualified by the moral law. The same thought is expressed, with a further specification, in Kant's statement of what has been termed by Allison (1990) the "incorporation thesis" (p. 40) in the *Religion Within the Boundaries of*

[11] Because I hold these to be equivalent notions, I find it misleading to call the principle of self-love a "principle of choice," as, for example, Reath does (2006, p. 41). I agree with Reath that because I have the end of happiness, I engage in associated characteristic modes of deliberation (2006, p. 42). But "principle of choice" suggests a principle for selecting an end I might not have, rather than a principle which guides me in filling out the content of an end I already have.

*Mere Reason*. Here, Kant speaks of an agent's "supreme maxim" (Religion 6:31), which constitutes her practical orientation as a whole (rather than this or that policy or an intention to perform this or that action), so that Gregor (1963) calls it the "fundamental attitude of will" (p. 155). If what I have said in this section thus far is correct, we should expect that there are two possible fundamental attitudes: having the morally qualified end of happiness, as the good agent does, and having the morally unqualified end of happiness, as the bad agent does. And this is, indeed, exactly what we find:

> [The human being] is…dependent on the incentives of his sensuous nature…and he incorporates them too into his maxim (according to the subjective principle of self-love). If he took them into his maxim *as of themselves sufficient* for the determination of his power of choice, without minding the moral law (which he nonetheless has within himself), he would become morally evil. […] [T]he difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim (nor in the material of the maxim) but in their *subordination* (in the form of the maxim): *which of the two he makes the condition of the other*. (Religion 6:36)[12]

In the first two sentences, Kant distinguishes between the good agent and the bad agent. Both agents "incorporate" the "incentives" of their "sensuous nature" into their fundamental maxim "according to the subjective principle of self-love": that is, the principle of having happiness as your end. But the bad agent does so "without minding the moral law," whereas the (here unmentioned) good agent does it in observance of the moral law. (The third sentence specifies the characterization of the bad agent. Not only does she not qualify happiness with the law; she qualifies the law with happiness.)

## 3.2 | Intellectualism?

Before I proceed to Kant's explanatory level two, let me bring together two things I have said. First, being a finite rational being is sufficient both for having happiness as an end and for being governed by the moral law. As I read Kant, to be a finite rational being is the same as to possess a finite will, since "the will is nothing other than practical reason" (Grundlegung 4:412). And the end of happiness could only belong to a will, not to what he calls a "lower" desiderative capacity (KpV 5:22), because the latter's modifications are not ends, but rather inclinations, which supply the matter for happiness but are not identical to it. Second, either you are the good agent, in which case you exercise your will—*alias* practical reason—in such a way as to pursue happiness qualified by the moral law; or you are the bad agent, in which case you exercise your will—*alias* practical reason—in such a way as to pursue happiness unqualified by the moral law.

This second point brings me into agreement with Allison (1990) in thinking that every action we undertake is either undertaken in observance of the law or in subordination of the law to self-love, that no action we undertake is an exercise of a capacity per se indifferent to morality (p. 5). Recently, Timmermann (2022) has argued against Allison (and those who have agreed with him) that Kant's incorporation thesis means merely that I cannot act on an inclination without

[12] It does not matter here whether we speak of having an end or of having a maxim. Doing the latter always involves having an end.

*choosing* to do so, and that this act of choice, an exercise of the merely elective power [*Willkür*], itself involves no assessment of my action against the standard of the law, an assessment performed by the will [*Wille*] or practical reason, which itself has nothing to do with happiness (contrary to my first point in the previous paragraph). Timmermann's (2022) argument is multipronged, and I cannot respond to all of it. I shall focus here on his main objection of principle to the Allisonian reading.[13] And that is that Allison—and Korsgaard, Andrews Reath, and Stephen Engstrom following him—are "intellectualists" who cannot account for the "possibility of a conscious, deliberate departure from one's own better judgment" (p. 94). This is because "intellectualism is committed to the thesis that, at the time of a wrongful act, we are deluded about our obligations" (2022, p. 103).

If "intellectualism" means Socratic intellectualism, this last claim is true. And I agree with Timmermann that Kant is not a Socratic intellectualist. But the part of the Allisonian reading of the incorporation thesis with which I concur involves no commitment to Socratic intellectualism. (Nor, in my view, do the readings of Korsgaard, Reath, and Engstrom, but that is another matter.) Timmermann supposes that unless there is a distinct act of election—the exercise of the power of choice—in which I do not assess myself by the moral law, it is impossible for me to act knowingly contrary to the law. This is because he supposes that if the elective act is the same as the evaluative act, I cannot but elect in accordance with my evaluation. But that supposition is false. For here is a perfectly coherent view, which I shall develop in Section 5: the finite rational will is a capacity to pursue happiness in observance of the moral law. Capacities are, in general, such as to admit of failure in their exercise. Given the description of this capacity, we can identify as among its failures instrumental irrationality (a failure in pursuit), prudential irrationality (a failure in applying the standards of systematicity), and evil (a failure in observation of the law, perhaps by consciously rejecting it). When you choose to satisfy your inclination in spite of the law, you exercise the capacity to pursue your happiness in observance of the law, but badly: in at least one of the several determinate modes of failed exercise. The possibility of failure does not require that we possess a distinct elective capacity indifferent to morality. And indeed, Kant explicitly states that "freedom of choice cannot be defined…as the ability to make a choice for or against the law.…Only freedom in relation to the internal lawgiving of reason is really a capacity [*Vermögen*]; the possibility of deviating it is an incapacity [*Unvermögen*]" (MS 6:226-7).[14]

Of course, evil is an especially interesting mode of failure, because it involves the rejection of the principle of the very capacity that you exercise.[15] To that extent, it is not exactly like tripping and falling on your way to the supermarket, though just as that would be a failure in walking, evil is a failure in exercising the finite rational will. For this reason, there is something deeply mysterious about it. According to Kant, we accommodate its possibility in philosophy because "according to the cognition we have of the human being through experience, he cannot be judged otherwise, in other words, we may presuppose evil as subjectively necessary in every human being" (Religion

[13] I respond to those arguments of Timmermann's which would target my first point more directly in footnotes 19 and 20.

[14] Here, I read Kant as a capacities-constitutivist. For a discussion of capacities-constitutivism in the contemporary metaphysics of normativity, see Fix (2021b). For discussion of the possibility of failure in particular, see Fix (2021a, pp. 1081–1084). For a presentation of the application of capacities-constitutivism to the interpretation of Kant, see Schafer (2023). For discussion of constructivism in ethics, including of contemporary Kantian varieties, see Bagnoli (2022).

[15] How exactly this rejection works is beyond my scope. Sometimes Kant suggests that it has to involve some kind of self-deception involving self-conceit (KpV 5:73n.). The point here is that the case of rejection is distinct from the case in which you commit an innocent misapplication of the law or are blamelessly ignorant of morally relevant features of your circumstances.

6:32). Nonetheless, its possibility is incomprehensible: "[t]he rational origin" of the "propensity to evil…remains inexplicable to us," because "the original predisposition…is a predisposition to the good; there is no conceivable ground for us, therefore, from which moral evil could first have come in us" (Religion 6:43).

This mystery about evil is an important topic in its own right. Here, I wish to note only that that evil is incomprehensible, whereas good is not, at least not in the same way. For this asymmetry is something that readings like Timmermann's have difficulty explaining. For Timmermann, the capacity to elect is in its nature indifferent to moral evaluation (2022, p. 108). This means that election according to moral evaluation is no more comprehensible, as an act of election, than election against it. Timmermann (2022) insists that good and evil are not "normatively on par" (p. 107), but what he means by this is that *Wille* instructs us to choose the law and not self-love. As an exercise of *Willkür* as such, the choices *are* on a par, so that Timmermann has not, as he claims, accounted for the *Vermögen/Unvermögen* passage, which explicitly concerns freedom of choice, nor Kant's claim that the propensity to evil (which is a propensity of choice) is unintelligible because the original propensity is the propensity to good.

Timmermann acknowledges the consequences of this position for intelligibility, saying that "[b]y dissociating choice from practical reason, the elective model of *Willkür* introduces an element of arbitrariness and inexplicability," so that Kant's view is "philosophically unsatisfactory"; but, he says, "[w]e may still have to attribute the elective model to Kant if key passages support if and alternative readings are equally unsatisfactory" (2022, p. 108). But not only does Timmermann's objection of principle to the alternative model presuppose something false—that is, that unless the elective act is different from the evaluative act, there is no way to explain evil—his positive alternative fails to account for the asymmetry between the intelligibility of good action and the unintelligibility of evil as exercises of the power of choice. The *Vermögen/Unvermögen* passage from the *Metaphysics of Morals* and the passage I have quoted from the *Religion* are not, of course, the only texts to be consulted on this matter. But they are enough to ground a suspicion that the "elective" model of *Willkür* may be, not merely philosophically unsatisfactory, as Timmermann admits, but textually unsatisfactory too—and so enough to ground the interest of developing the very different kind of reading which I favor, on which Kant's motivational dualism of happiness and morality appeals to no morally indifferent volitional powers but rather to complexity of structure in the one and only volitional power we have, the rational will, *alias* finite practical reason, to whose exercise choice belongs.[16]

## 4 | KANT'S SECOND LEVEL: NEEDINESS

Kant says that you have to have the systematic satisfaction of your inclinations as your end, either with or without qualification. That is so as a matter of the essence of your will. This, if true, is enough to secure the deliberative import of inclinations. But just to make this claim is not to explain why the will has this end, whereas this is something Kant professes to explain. It is only

[16] A comprehensive treatment of the discussion of the *Wille/Willkür* distinction in the secondary literature is beyond my remit. Timmermann sides with those who believe that the point of the distinction is to accommodate the possibility of bad action by the introduction of a morally indifferent elective capacity. I believe that the point of the distinction (in Kant's critical-period ethical works) is to distinguish between two conditions of the will, one in which it stands ready to realize its end, because its possessor has the relevant skills and is in favorable circumstances, and that in which it is not, in which case its act is called mere "wish" [*Wunsch*]. See, for example, KU 5:177-8 and Engstrom (2009, 28ff.) for discussion.

in light of the explanation given at this second level that the explanation at the first level achieves its full potential. Thus:

> To be happy is necessarily the demand of every rational but finite being, and therefore an unavoidable determining ground of its desiderative capacity. For, satisfaction with one's whole existence is not, as it were, an original possession and a beatitude, which would presuppose a consciousness of one's independent self-sufficiency, but is instead a problem imposed upon him by his finite nature itself, because he is needy and this need is directed to the matter of his desiderative capacity. . . . (KpV 5:25)

What ultimately explains the attributes of the finite will is its finitude, and what this means in this context is something that human beings share with nonrational animals: neediness. The human will is a desiderative capacity (KpV 5:15), and as such it has the function for us that animal desiderative capacities have for animals: it is in its exercise that an individual maintains herself for the complete course of her life. For it is in its exercise that such an individual lives: as Kant is recorded as saying, "[l]iving beings have a desiderative capacity; one can make this into a definition of living beings" (L2 28:587; see also KpV 5:9n., MS 6:211); as he says in the *Groundwork*, "[*w*]*ill* is a kind of causality of living beings insofar as they are rational" (Grundlegung 4:446).[17] Her inclinations have the function of pointing her towards the things she needs to do that.

This is also true for nonrational animals. But nonrational animals do not have their own happiness as an end. In the first place, they have no ends at all: they act from their inclinations in accordance with instinct, not in the exercise of a will (e.g., KpV 5:61). And this difference is not merely terminological. To have the end of happiness is to have a desiderative consciousness of the complete course of one's life, with a focus on the future, all the way forward, however long it will last. Kant thinks that only a rational being could have such a consciousness. In his *Conjectural Beginning of Human History*—his philosophical reading of *Genesis* 2–6—he writes that

> [t]he third step of reason, after it had mixed itself into the first immediately sensed needs, was the deliberative *expectation of the future*. This faculty of not merely enjoying the present moment of life but of making present to oneself the coming, often very distant time, is the most decisive mark of the human advantage of preparing himself to pursue distant ends in accordance with his vocation—but also simultaneously it is the most inexhaustible source of cares and worries which the uncertain future incites and from which all animals are exempt. (Anfang 8:113)

As this passage shows, Kant thinks of desiderative future-directedness as part of what you get when you add apperceptive self-consciousness to animal desire. Nonrational animals have a desiderative past, present, and future, to be sure, all of which eventually adds up to the complete course of their life, but at no point in their lives do they live under the idea of the course of their lives. But rational animals do, and that is the same as their having the end of happiness. The end of happiness is the self-consciousness of this aspect of animal desire.[18]

[17] See also Volckmann 28:499 and Dohna 28:679-80.

[18] Kant often characterizes happiness in terms of the notion of a rational being: "a rational being's consciousness of the agreeableness of life uninterruptedly accompanying his whole existence is *happiness*" [5:22]; "*happiness* is the state of a rational being in the world in the whole of whose existence *everything goes according to his wish and will*" [5:124]. He never suggests that non-rational animals can be happy, and his word for happiness—*Glückseligkeit*—where "*Seligkeit*" means

In Section 2, I said that the import of Kant's claim about the necessary end of happiness in the *Groundwork* is underappreciated. He claims, at what I have called level one of his explanation, that this feature of our finite rational moral psychology is as deeply entrenched as our subjection to the moral law in the form of duty. Here, at level two, we have an explanation of why he thinks that. The end of happiness is not some arbitrary or gratuitous end which comes preinstalled in your will for reasons unknown. It is explained by what you are: a self-conscious animal, which is to say a self-conscious temporally extended desiderative being, and in particular one with a future. The standards of systematicity are thereby grounded in the temporal extendedness of life.

This level-two explanation arms Kant against attacks by some of his sympathetic readers. Korsgaard (2008) has argued that, in claiming the end of happiness to be inevitable, Kant contradicts his own doctrine of freedom, which, as she understands it, says that all ends are freely chosen (pp. 80–81). But if the end of happiness is just the self-consciousness of animal desire insofar as it is temporally extended, with especial focus on the future, it is not some particular end which you are arbitrarily forced to pursue. Nonrational animals pursue particular ends with no thought as to their fitting into a course of practical life. Rational animals pursue particular ends under the idea of such a course. That is part of what it is for a rational animal to exercise her will at all. There is no such thing as your choosing not to pursue happiness, according to this conception of pursuing happiness: it would be choosing to have no desiderative orientation to your future, like a nonrational animal.

You are free not in the selection of the big end of happiness (because you do not decide to be a rational, rather than a nonrational animal), but in your selection of the little ends that you pursue in your distinctively rational way. According to Kant, however, this freedom cannot be explained except in terms of your subjection to the moral law. You have a free will, in the sense with which he is concerned, when your choice of these little ends is subject to a principle of end-selection, the moral law, over and above the standards of systematicity. Nonetheless, because your choice of these little ends is indeed subject to such a principle, there is a straightforward sense in which your pursuit of happiness is free. And that is that your selection and pursuit of these little ends, which together make up your conception of happiness, are acts of freedom, being subject to the law.

Here is a passage in which this feature of the interplay between the end of happiness and the moral law comes out. Kant speaks here of our reason as a particular "arrangement of our nature" [*Naturanstalt*]:

> For that he has reason does not raise him in worth above mere animality if reason is to serve him only for the sake of what instinct accomplishes for animals; reason would in that case be only a particular mode nature had used to equip the human being for the same end to which it has destined animals, without destining him to a higher end. No doubt once this arrangement of nature has been made for him he needs reason in order to take into consideration at all times his well-being and woe; but besides this he has it for a higher purpose: namely, not only to reflect upon what is good or evil in itself as well—about which only pure reason, not sensibly interested at all, can judge—but also to distinguish the latter appraisal altogether from the former and to make it the supreme condition of the former. (KpV 5:61-2)

"blessedness" or "beatitude"—strongly discourages this: "[a] state whose comfort has its source merely from things of nature… and of which I am not the author through my freedom, would not be called '*selig*'" (Vigilantius 27:644). Nonetheless, my point here is not that non-rational animals cannot be happy but merely that they do not have their happiness as an end.

If human practical reason were limited to what you get when you add apperceptive self-consciousness to the temporal extendedness of animal desire—that is, if it were limited to the end of happiness and thereby to the application of the standards of systematicity and instrumental reasoning—human beings would be "destined" to nothing higher than are nonrational animals. We would just do what they do in a distinctive way.

Elsewhere, Kant doubts whether practical reason could be thus limited. In the passage concerning the "favored creature" in the *Groundwork*, for example, he suggests that a being with practical reason limited to the end of happiness (and instrumental reason) would be contrapurposively constituted, because instinct would serve an even theoretically rational being better than the end of happiness alone would (Grundlegung 4:395). There is some reason to suppose that this is not always his position: witness his distinction between humanity and personality (Religion 6:26). But in the latter half of this passage from the second *Critique*, he says something about human practical reason—which is not, in fact, thus limited, according to him—which does not depend on his considered view on the question whether there could be a kind of practically rational being not governed by the moral law. He says that in our case, because we are, by virtue of our reason, destined to a higher purpose, we *also* need reason to consider our well-being and woe. In the *Conjectures* we saw Kant imagine the addition of apperceptive self-consciousness to animal desire to yield the end of happiness. Here, we see Kant imagine the addition of *moral* apperceptive self-consciousness in particular to animal desire. And he says that when you add practical reason which comes with the moral law as a principle of end-selection to animal desire, you get the end of happiness too, automatically.

This corroborates my claim in Section 2 that in saying that being a finite rational being is sufficient for having happiness as your end, and in saying that this end belongs to the essence of a finite rational being, Kant is saying that the end of happiness belongs to its rational will, which is the same as its practical reason. For when you add moral practical reason to animal desire, you get not just government by the moral law, but the end of happiness. And moreover, just a few lines before this passage, Kant says something which is difficult to read any other way: that

> [t]he human being is a being with needs, insofar as he belongs to the sensible world, and to this extent his reason certainly has a commission from the side of sensibility which it cannot refuse, to attend to its interest and to form practical maxims with a view to happiness in this life and, where possible, in a future life as well. (KpV 5:61-2)

If reason cannot refuse sensibility's commission, there must be something about reason which makes this the case: the necessary end of happiness, explained as Kant explains it at level two.

Now, of course, there is no such thing, literally speaking, as adding reason to a nonrational animal. The cash value of the metaphor is that happiness as an end and government by the moral law in the form of duty are explained neither by your being a *rational* being in general (since God is also a rational being and has neither of these features), nor your being an *animal* in general (since nonrational animals are also animals and have neither of these features), but by your being a rational animal.

It is striking that in passages like KpV 5:61-2, Kant connects these two features of finite practical reason by saying that any animal governed by the moral law also exercises reason in the pursuit of happiness. It is striking because it seems not to be the only philosophical possibility. Here, for example, is a coherent position which is not Kant's. We are theoretically conscious of our future, but not desideratively; like those of nonrational animals, our inclinations motivate us by virtue of instinct, not in a way that is mediated by the end of happiness. We are to that extent like the

favored creature. But we are also rational beings governed by the moral law, and sometimes we act from the law: we intervene in our own animal action, prevent our inclinations from realizing themselves, and act from reason instead. On this view, our practical reason has no commission from sensibility. It is utterly indifferent to our securing what we are inclined to pursue. To be sure, it does not interfere with that pursuit unless it violates the moral law, but otherwise it falls outside the scope of reason's concern. This is what it would be like if our practical reason were simply plopped on top of our animality. Its only relation to the latter would be to inhibit it when it happened to proceed contrary to reason's principle.

If our practical reason were like that, Kant could not say that "our well-being and woe count for a *very great deal* in the appraisal of our practical reason" or that "happiness is not *the only thing* that counts" (KpV 5:61), since our well-being and woe would count, according to practical reason, for *nothing at all*.[19] We would have two distinct capacities to act, an animal capacity and a rational capacity, the first untouched by reason, the second untouched by animality.[20]

But that is not Kant's view. His conception of happiness makes sense only given a conception of finite practical reason as itself internally connected to our animality, as having a commission from sensibility. Literally speaking, this commission is the end of happiness, which belongs to finite reason, or the finite rational will, neither merely to our animality (insofar as we share it with nonrational animals) nor merely to our rationality (insofar as we share it with God).[21]

Nonetheless, Kant's view is not transformativism either, because while finite practical reason is, in the end of happiness, oriented towards inclinations, inclinations themselves are blind to finite practical reason. The difference between nonrational animals and rational animals lies not in the representational purport of their inclinations, as the transformativist believes, but merely in their functional profile in the animal's desiderative economy as a whole. Nonrational animal inclinations realize themselves in action mediated solely by instinct: the "*animal* power of choice (*arbitrium brutum*)" is "pathologically necessitated" (A534/B562). Rational animal inclinations depend for their realization on mediation by practical thought: "[t]he human power of choice is indeed an *arbitrium sensitivum*, yet not *brutum* but *liberum*, because sensibility does not render

[19] Timmermann (2022, p. 46) claims that, in Kant's mature (post-1784) view, "practical reason is restricted to *pure*, i.e. moral, practical reason." This is because Kant comes to think that hypothetical imperatives are really theoretical (e.g., KpV 5:25 n.). But this view about hypothetical imperatives does not entail that happiness, as an end, does not belong to the finite rational will and thereby to finite practical reason. Nor is Timmermann's position compatible with Kant's claim that "our well-being and woe count for a *very great deal* in the appraisal of our practical reason" in a book published in 1788. Timmermann himself says that "[w]hat is practical…pertains…to final ends" and that "[t]here are…only two such ends, one's overall happiness…and morally good activity" (2022, p. 49), which equally contradicts his claim that practical reason is restricted to moral reason. To be sure, in his sketch of the "emergence" of practical reason, he suggests a way out of this contradiction: he says that the "commission" that sensibility cannot refuse is merely practical reason's first or "original" assignment, one which it abandons with the emergence of morality (2022, pp. 76-84). But this is not supported by the "commission" passage, which he cites. To all appearances, this passage is about what is the case with us now, constituted as we are.

[20] Timmermann (2022, p. 48) suggests that our practical reason "needs to take care of the empirically conditioned side of action" *only* because if "prudential matters" were "left to blind, mechanical instinct in creatures endowed with pure practical reason", "reason would constantly clash with instinct as it puts checks on the pursuit of happiness." But he cites no passages in support of this claim. And since by "reason" he means a capacity whose sole concern is morality, I do not see why an instinctual organization of inclinations would be significantly more likely to result in clashes with morality than a prudentially reasoned-out, but morally indifferent, conception of happiness.

[21] As Herman (2007) puts it, "[h]appiness is a problem imposed on us, in that each of us has to decide how to live *a life*.…Given our natures as rational and sensible beings with the concept of a life, we have the end of happiness" (p. 182). See also, for related discussion, Hills (2006), Engstrom (2009 p. 85).

its action necessary, but in the human being there is a capacity to determine oneself from oneself, independently of necessitation by sensible impulses" (A534/B562). The inclination of a rational animal is never sufficient, even assuming the requisite skills and favorable circumstances, for the animal's acting accordingly. For the rational animal to act, she must choose to act on her inclination—a choice *per se* governed by the standards of systematicity and the moral law. So in a rational animal, though not in a nonrational animal, the functional profile of inclinations is limited to their role as constituting proposals for the addition of their objects to the will. This difference in functional profile makes a difference to what inclinations are in a rational animal, if (as I believe) functional profile is part of what makes each element of Kant's moral psychology what it is. But it does not assimilate them to practical judgments, to representations of reasons or of the good. On the contrary: their function is utterly heterogeneous from that of exercises of the will as such. Kant might say about inclinations and the will what he says about sensibility and the understanding in the theoretical case: "these two capacities cannot exchange their functions" (A51/B75).[22]

## 5 | HAPPINESS AND DUTY, *redux*

### 5.1 | Universal rational life

Now that I have introduced Kant's explanatory level two, another question arises concerning the relation of happiness and morality. It may seem to be a mystery why we are subject to the moral law, if this is a constraint on end-adoption over and above the standards of systematicity. Those standards are grounded in the self-consciousness of the temporal extendedness of a living being's existence. Why would, why *should*, such a living being subject itself to any constraint on end-adoption over and above those expressive of life itself? What could explain Kant's claim that "happiness is not *the only thing* that counts" (KpV 5:61)?

The answer to this comes in two parts, one at each level of Kant's explanation of the deliberative import of inclinations. Here is the first, at level one. Two things belong to the finite rational will: the end of happiness and its government by the moral law in the form of duty. But the finite rational will is a capacity. And every capacity is a capacity to do something, something called its *exercise*. So we have to say what the finite rational will is a capacity to do, such that these two things feature in its exercise. The most straightforward way of putting the answer is to say that the finite rational will is, as I said in Subsection 3.2, the capacity to *pursue your happiness in observance of the moral law*: that whole thing. Pursuing your happiness heedless of the law, like the bad agent, is not doing something complete and self-standingly intelligible as an exercise of your will. It is exercising this capacity in a privative, because incomplete, way. It is leaving off the moral qualification that completes the exercise. That is why you count as the bad agent when you do that. Goodness and badness are defined in terms of completeness and incompleteness in the exercise of the capacity.

But that answer on its own is unlikely to seem satisfactory. After all, what *is* a capacity to pursue happiness in observance of the moral law? I am insisting that it is that whole thing, rather than the "pursue happiness" bit, which is a self-standingly intelligible unity, but it looks rather like two things welded together. This is an interpretive problem, because as long as this appearance persists, we have not made it intelligible how reason, whose principle is the moral law, could have

[22] For discussion of this passage, see Pendlebury (2022).

a commission from sensibility. That we are animals with a sensibility is, again, not enough to explain this commission, since it is coherent (though not true, according to Kant), as I said in Section 4, to suppose our reason utterly indifferent to sensibility.

This is a question any answer to which requires undertaking further interpretive commitments. I limit myself to the consideration of a candidate answer, to illustrate what the moral psychological structure I have attributed to Kant allows us to say. Thus far, I have ignored the question of what morality requires. But now I want to consider what Kant says about his fourth example of a duty in the *Groundwork*, described under the second formulation of the categorical imperative:

> [T]he natural end that all human beings have is their own happiness. Now, humanity might indeed subsist if no one contributed to the happiness of others but yet did not intentionally withdraw anything from it; but there is still only a negative and not a positive agreement with *humanity as an end in itself* unless everyone also tries, as far as he can, to further the ends of others. For, the ends of a subject who is an end in itself must as far as possible be also *my* ends, if that representation is to have its *full* effect in me. (Grundlegung 4:430)

This suggests the following picture. Each of us has, as her end, her own happiness, the systematic satisfaction of her inclinations. That is for self-love to belong to her volitional orientation. And that is to say that she loves herself. But insofar as each of us is a good agent, she qualifies this end by the moral law. What that means, according to the fourth example under the second formulation of the categorical imperative, is that she qualifies her own pursuit of happiness in light of the morally qualified pursuit of happiness by others. She loves her neighbor as herself. Her life consists not merely in the systematic satisfaction of *her* inclinations, but such satisfaction of her inclinations as is limited by the practical recognition of the other morally qualified systematic inclination-satisfiers. This practical recognition involves noninterference and furtherance. For me to further the ends of others is not the same as for them to pursue their own ends. If it is your end to varnish a bookcase this afternoon, I do not further your end by seizing the brush from your hand, expelling you from the room, locking the door, and finishing the job myself. That would be interference with your end. I further your end by picking up some more polyurethane from the hardware store, by grabbing another brush and joining you, or indeed in finishing the job—if that is what you ask me to do and I agree to do it. So, strictly speaking, if it is your end to varnish a bookcase, when I make that end my own, I do not adopt the end of *varnishing a bookcase*, but the end of, roughly speaking, *not needlessly interfering with your varnishing a bookcase and perhaps helping you varnish it if that's what you and I agree on*.

That means that the finite rational will is aptly described as the capacity *to pursue happiness in cooperation with other happiness-pursuers*.[23] The end of happiness is indeed grounded in the character of life itself. But the moral law is not simply *additional* to what is grounded in the character of life itself. To govern yourself by the law is not to evince concern with something other than life, though it is to evince concern with something other than *your* life: it is, namely, to live your life in

[23] Here, I am singling out the fourth example under the Formula of Humanity in the *Groundwork* as expressive, not merely of one duty among others, but of what morality is all about. This is suggested in the idiom of the passage itself, when Kant speaks of what it is for the representation of humanity as an end in itself to have its "full effect" on me. To the extent that I fulfil merely duties to myself, or those plus duties of respect to others, and do not love my neighbor, my moral practice is incomplete; it comes to full fruition, as it were, in practical love. For another articulation of this idea, see Pendlebury (2024, pp. 435–436).

practical observance of the lives of others. You retain a particular concern for your particular life, of a kind which no one else could have. But this concern is inflected by a universally commanded concern for universal rational life.

This is not to say that every species of happiness-pursuer of which we can frame a logically coherent concept is like this. It is not to say that no kind of rational being could not be concerned with her life without being universally concerned with rational life. But I do not take the intelligibility of the kind of rational will which Kant thinks *we* have to depend on claims like that—claims about which, as I have already noted, he seems himself ambivalent. The point is that morality is not an unintelligible external constraint on the way in which a rational life is lived. It is a way in which a rational life—and, according to Kant, *our* rational lives—can be lived. But it is a way that involves a different conception of what it is to live a life than applies to the nonrational animals, not just because we are self-conscious in our desiderative temporal extendedness, but because while life for the individual nonrational animal (and for the amoral happiness-pursuer, if there is such thing as such a being) consists wholly in its securing what it needs, our lives consist in more than this: each of us secures what she needs in cooperation with other individuals like us (and not because it conduces to her own individual need-satisfaction). This is the higher purpose of which Kant speaks.

## 5.2 | The highest good

Before I turn to Korsgaard, let me very briefly sketch the relation of what I have said to Kant's doctrine of the highest good—in which happiness is important—as I understand it. The highest good is defined in terms of the good. The good is defined as the "necessary object of the desiderative capacity…in accordance with a principle of reason" (KpV 5:57-8). This object, as I understand it, is not a particular action but the form of every permissible action: it is "moral possibility" (KpV 5:57-8). In in accordance with what I have said thus far, we can say that insofar as you pursue happiness in observance of the law, your action is good. Otherwise, it is evil. Happiness is part of the good.[24]

Kant discusses this in the Analytic of the second *Critique*. He does not discuss the highest good until the Dialectic. The Dialectic is an apposite place for this discussion for several reasons, among which is that Kantian Dialectics treat of unconditioned totalities, and the highest good is meant to be such a totality: "pure practical reason [reason]…seeks the…unconditioned totality of the object of pure practical reason, under the name of the *highest good*" (KpV 5:108). The difference between the notion of the good and that of the highest good is, in my view, solely that the latter notion is that of an unconditioned totality. The notion of an unconditioned totality, when applied to moral possibility, does not yield the notion of happiness. Happiness cannot make its first appearance in the good when we frame the idea of the unconditioned totality of the good. It must be there already.

Thus Kant writes in the Dialectic that it "has been proved in the Analytic" that virtue is the "*supreme* good," but that

> it is not yet, on that account, the whole and complete good as the object of the desiderative capacity of rational finite beings; for this, *happiness* is also required, and that not merely in the partial eyes of a person who makes himself an end but even in the

[24] That does not entail that it *is* good. What is good is securing morally qualified happiness.

> judgment of an impartial reason, which regards a person in the world generally as an end in itself. (KpV 5:110).

As I read this passage, by the "whole and complete good," Kant does not mean an unconditioned totality. He means the good as discussed in the Analytic. For notice that none of the reasoning in this passage makes any reference to unconditioned totalities. That a person makes herself an end, and that she is as an end in itself, are pre-Dialectical claims. The complete good is the object of the will of the good agent: happiness qualified by the moral law.

What, then, is the highest good? For the answer, we have to consult Kant's resolution of the antinomy of pure practical reason, in which he says that

> it is not impossible that morality of disposition should have a connection, and indeed a necessary connection, as cause with happiness as effect in the sensible world, if not immediately yet mediately (by means of an intelligible author of nature), a connection which, in a nature that is merely an object of the senses, can never occur except contingently and cannot suffice for the highest good. (KpV 5:114-15)

The highest good requires at least that there be a necessary connection between virtue—morality of disposition—and happiness. It is a condition in which, to the extent that you qualify your happiness by the moral law, to that extent you are guaranteed, by a necessary connection, to succeed in action and thereby be happy. Your happiness is not conditioned by anything outside your moral disposition: no external hindrance can possibly stand between the internal perfection of your will and your achievement of happiness.[25] That is what "highest" adds to "good": not happiness, but an unconditioned connection between virtue and happiness.

It is an interesting question how Kant purports to show, as he claims, that we are rationally required to pursue the realization of such a condition, and how exactly his argument for the necessity of faith in God as the principle of the necessary connection in question works. But my point here is merely that his initial description of the good agent—as someone whose pursuit of happiness is qualified by the law—is pre-Dialectical and does not invoke a necessary connection between such qualification and the achievement of happiness. Kant's doctrine of the highest good *presupposes* this pre-Dialectical unity of morality and happiness in the good, as the object of the good agent's will.[26]

## 6 | KORSGAARD'S MORAL PSYCHOLOGY

### 6.1 | Principles of end-selection

That, I believe, is Kant's basic story about the deliberative import of inclinations. It is a two-level explanatory structure, at the first of which is the attribution to the finite will of something which confers deliberative import upon inclinations—the end of happiness—and at the second of which

[25] Kant says earlier that "virtue and happiness together constitute possession of a highest good in a person" (KpV 5:110), but given his later remarks, he must mean virtue, happiness, *and a necessary connection between them* constitute possession of a highest good in a person. So in a person the highest good involves not merely a necessary connection between the degree of virtue and that of happiness but also, perhaps, maximal virtue, and thereby maximal happiness.

[26] Here I am informed by Engstrom (1992).

is the explanation of why the finite will has this end. The first thing to note in connection with Korsgaard is that if she thinks that something about the human will confers deliberative import upon inclinations, if she endorses anything structurally analogous to Kant's first level, it must be something different in content from Kant's. This is because Kant's view is that happiness belongs to the will as a necessary end, and Korsgaard rejects this doctrine, for two reasons.

First, she claims that it is not *formal*: Kant's imperative of prudence, she thinks, "fits so badly with the rest of his theory, since his other principles are formal ones" (Korsgaard 2009, pp. 57–58). What she means by a formal principle is one which springs from the essence of the human will. And evidently, she supposes that nothing having to do with happiness can belong to the essence of the human will. This is brought out in her opposition to the egoistic principle:

> [I]t is not even remotely plausible to suppose that *the egoistic principle* captures the very essence of self-determination. That is, it is not plausible to think that you only succeed in exercising self-determination if you aim to maximize the satisfaction of your desires; or that you are not really willing at all unless *what you will* is the maximum satisfaction. If we accept the stronger version of the Kantian conception, then the egoistic principle simply seems to be *the wrong sort of thing* to be a principle of practical reason. To put the point in somewhat more old-fashioned Kantian terms, the egoistic principle is concerned with the *content* of the will, not with the very *form* of willing. (Korsgaard, 2008, p. 86)

This passage shows that for Korsgaard, the human will is a capacity for self-determination, and *that's all there is to it*. But Kant would not accept this characterization of the essence of the human will, since, as we have seen, he includes in that essence that it is a will which has happiness as an end. Since "formal" means "springing from the essence of the human will," this objection to Kant's position amounts to a disagreement about that essence.

Second, Korsgaard (2008) claims that to attribute any end to the will as naturally necessity is to claim that the will is unfree—because freedom is the freedom to select ends—and is thus incompatible with Kant's doctrines concerning human freedom (pp. 80–81). As I have said, this criticism depends upon an understanding of Kant's doctrine of freedom which I do not share. But it is necessary to mention it here because it brings out an important feature of Korsgaard's own position. Because she rejects the attribution of any end to the will as naturally necessary on the ground that this would constitute it as unfree, she believes that the only features of the will's essence which have normative bearing are its principles of end-selection.[27] This allows us to conclude that if anything about the Korsgaardian will confers deliberative import on inclinations, it must be its principles of end-selection.[28]

## 6.2 | The constitution model

You might think this conclusion is hasty. When Schapiro, for example, considers the question of how Korsgaard explains the deliberative import of inclinations, she examines Korsgaard's

[27] See also Korsgaard's argument that principles of end-selection are required to "give normative status to our ends" (2008, p. 63).

[28] This is not to say that Korsgaard does not think of the finite rational will—the capacity—as having a purpose and a function. But the way in which it fulfills its function towards that purpose, on her account, is that its possessor selects ends of action—of the *exercise* of that capacity—according to a principle of end-selection.

"constitution model," according to which the roles of inclination and the will (or practical reason) in action are analogous to those of different kinds of office-holder in a polity.[29] "Inclination presents the proposal; reason decides whether to act on it or not" (Korsgaard, 2009, p. 154). It may seem that even if no relation to inclination is encoded in the will's principle of end-selection, because the will is assigned the role, in the constitution of action, of reviewing proposals constituted by inclinations.

But I do not think it makes sense, for Kosgaard, to think that while the will is the source of its own principle of end-selection, its application of that principle to proposals constituted by inclinations is explained by the assignment to it of a role. That would mean that from the point of view of the will as such, inclinations have no deliberative import. And that would mean that the assignment directs the will to something to which its own nature does not. But if the will's having a necessary end constitutes unfreedom, on Korsgaard's view, this would too. An assignment of a role cannot be imposed on the will if it is to be free. It must be, then, that if talk of assignment is acceptable from Korsgaard's point of view, it must be merely a way to express a point about the will itself: that is, it must be that by virtue of its own principles of end-selection, it has the role of reviewing proposals constituted by inclination.

## 6.3 | The content of the principles

If that is true, then we are faced with this question: is the role of the will in relation to inclination, as depicted by the constitution model, encoded in the content of the principles of end-selection which Korsgaard attributes to the will? I think the more plausible answer is "no." This is suggested by the passage from "The Myth of Egoism" which I have already quoted in which she argues that the egoistic principle cannot be a principle of end-selection for the will because it does not capture "the very essence of self-determination." The review of proposals constituted by inclinations does not belong to the very essence of self-determination. In *Self-Constitution*, Korsgaard's idiom is different, and indeed more Kantian, but the same point may be made in its term. She argues that the categorical imperative is the will's one principle of end-selection because "[t]he essential characteristics of an agent"—that is, a being with a will—"are efficacy and autonomy" (Korsgaard, 2009, p. 82), and the categorical imperative is the principle governing autonomous efficacy as such.[30] Note again the contrast with Kant: while he ascribes to the essence of the finite rational will both the end of happiness *and* the principle of end-selection which is the categorical imperative, Korsgaard ascribes to the essence of the rational will *only* the latter.[31,32]

[29] Schapiro argues that Korsgaard's view does not properly capture the sense in which inclinations are merely *provisionally* deliberative, though she believes that it captures a sense in which they are deliberative, full stop (2021, pp. 122–123).

[30] Korsgaard's distinction between the categorical imperative and the moral law (1996, p. 99) will not matter for this argument. It may be understood that by "categorical imperative" I mean "the moral law" (in her sense).

[31] It might be suggested that the notion of autonomy could do the work, inasmuch as, for Kant, at least part of what it means to be autonomous is not to be determined by your inclinations. But that does not give them deliberative import.

[32] You might think that Korsgaard could explain the deliberative role of inclinations in terms of her notion of practical identity (1996, 101ff.). For "[m]oral identity does not swamp other forms of identity: no one is simply a moral agent and nothing more" (1996, p. 125). But I do not see how. Your practical identity (or the totality of your practical identities) functions as a law more determinate than the categorical imperative (1996, 121f.), but nothing in the notion requires that this specification—for example, a law governing you as a Quaker or you as an American citizen—confer the status of proposal upon inclinations. Nor do her remarks about the relation between practical identities and desires furnish an

## 6.4 | Why are inclinations the practical given?

Above I said that Kant explains, at a second level of explanation, our having the necessary end of happiness in terms of our consciousness of our neediness and the function of our inclinations of directing us toward what we need. I think that a final possibility for Korsgaard to escape the transformativist objection admits of reconstruction from what she says in a register analogous to Kant's level two.

Her second-level idiom is that of what is *good for* me rather than, like Kant's, that of what I *need*. This is perhaps more obvious in the idiom of her more recent work on the good than in *Self-Constitution*. According to the position presented in *Fellow Creatures*, for example, our inclinations direct us toward what is good for us (2018, p. 138), and in adopting their objects as our ends we do so—subject to the constraint of the moral law—solely because they are good for us: "I have no other reason for taking my end to be good absolutely, than the fact that it is good *for me*" (2018, p. 144). So it might be suggested on her behalf that my principle of end-selection is a principle for the review of proposals constituted by inclinations because inclinations direct me towards what is good for me. If that is true, then there must be some internal connection between the nature of my will—whose principle is in question—and what is good for me. For example, Korsgaard might say that my will is, in part, my capacity to secure what is good for me. More fully characterized, it is, roughly, my capacity to secure what is good for me compatibly with others' securing what is good for them. That would allow her to rebut the transformativist. But if that is what she has in mind, she does not make it explicit. She does not say that, in addition to efficacy and autonomy, securing what is good for the agent is essential to agency. And so, according to the letter of her texts, it is unintelligible how "the fact that it is good *for me*" could count as a reason for adopting the object of my inclination as my end. For end-adoption is an act of the will, and unless the will is oriented toward what is good for me, there is no such thing as my adopting an end because it is good for me—at least where that "because" gives a reason for performing an act of the will and is not merely a causal explanation of the "brute-force" kind.

## 7 | CONCLUSION

I have considered two ways in which the nature of the will might be thought to confer deliberative import upon inclinations: Kant's way, via a naturally necessary end; and (if it is her way) Korsgaard's, via the form of principles of end-selection. I wish to conclude by reflecting on something important they have in common, as an two-level explanatory structural alternative to the transformativist conception of desire.

In Section 4, I said that while the transformativist says that inclinations share roles with exercises of reason, according to Kant they have a distinct function, one which they have by virtue of reason's constitution, and in particular by virtue of the end of happiness. Here, I want to bring out

account of the deliberative role of inclinations: "the desires and impulses associated [with our practical identities] do not just *arise* in us. When we adopt…a conception of practical identity, we also adopt a way of life and a set of projects, and the new desires which this brings in its wake" (1996, p. 239). This is to say that the principle of the will might make a difference to the desires we have, but it is not to say that that principle confers upon those desires the status of proposals. Indeed, Korsgaard insists that my inclinations' origin in my own activity makes no difference to their significance "at the moment of action," because at that moment, however much I regard my desires as "my own productions," I have no choice but to identify *not* with them but with my "principle of choice" (1996, p. 241). So if the inclinations mean anything to me at the moment of action, my principle of choice must confer deliberative import upon them.

more sharply the implications of this structural difference between the Kantian and the transformativist, toward highlighting what is both historically distinctive and philosophically promising in the Kantian conception of our mindedness.

According to a certain kind of rationalist, finite reason is independently intelligible, constituted independently of its relation to our sensibility. It might be characterized in terms of judgment, or inference, or knowledge, or thinking, or rational action. Since some of our judgments, inferences, knowledge, and thinking seem not to depend on sensibility—mathematical judgments, for example—and since some our rational action is action contrary to our inclination—as when we act as morally required, though we do not feel like it—it is not unreasonable to suppose that our reason as such is what it is independently of our sensibility, though in some restricted range of applications it involves a relation to sensibility. It is natural for the rationalist, against the background of this picture, to suppose that insofar as there is such a restricted range of applications, our sensibility must be fit to play its role in them. It must harmonize with reason; it must have the characteristics that make it suitable as a co-conspirator in reason's exploits. Its representations must itself amount to judgments so that they can be premises in inferences, or must otherwise be in the right shape to serve as inputs into rational processes.

But this is not Kant's picture. For Kant says that it belongs to finite reason, just as such, that each of us has her own happiness as an end. And happiness is defined in terms of inclination-satisfaction: in terms, that is, of desiderative sensibility. This allows that sensibility need not be explanatorily assimilated to reason to play a role in rational life. While the rationalist forces the explanatory burden onto sensibility—it must match up with an independently constituted reason; it must share reason's functions—Kant makes reason shoulder the burden all by itself. Our practical reason is a capacity to satisfy our inclinations subject to the standards of systematicity and the moral law. Once we write that down, there are no further questions to pose about whether our inclinations suit our reason.[33] Our reason is *per se* constituted to deal with them as they are. I submit that transformativists have been animated by rationalist sympathies, sympathies which Kant does not share because he is not, in this sense, a rationalist.

And here is one further gesture. This kind of rationalism may hold especial appeal to those who think that our practical reason is attuned to something external to itself: to reasons, or to the good (construed independently of the will, as Kant does not construe it), or to normative facts or properties. For if you endorse this kind of externalist rationalism, it is natural to think that our desiderative sensibility can play a role in our rational life only if it too is attuned to that to which our practical reason is attuned.[34] And sometimes transformativists can talk as though this view is animating their arguments, even when they do not endorse it.

For example, Boyle considers an alternative to his position—that desires have the significance of practical judgments—according to which there is a "linking disposition…normally present in rational subjects, which ensures that normally, when S desires E, E appears worth pursuing to S" (2016, p. 541). He argues that this disposition, if it is to do its explanatory work, must be one "whose continued operation depends on [the subject's] acceptance of it as sound" (2016, p. 541), acceptance which amounts to the "belief" that her desires normally direct her to things that she

[33] I believe that there are analogous problems with prominent arguments for the necessity of conceptual content in perception, such as those due to McDowell (1994) and discussed by Boyle (2016, pp. 533–535). Boyle's views about the case of perception have since changed in important respects—see Boyle (2022)—so that the position articulated here, about practical reason, is analogous to the position he has adopted in the case of theoretical reason. But he has not himself extended his position to the practical case, and I do not know whether he would want to.

[34] Consider, for example, Scanlon's (1998) view that desires are perceptions of reasons (17ff.).

has some reason to pursue (2016, p. 541). The subject must believe that her desires are attuned to that to which her practical reason is attuned: reasons for action. It would be natural to expect him to say that my having the end of happiness, as in Kant's account, must similarly be undergirded by such a belief.

I do not wish to attribute externalist rationalism to Boyle; I wish only to note that if, like Kant, you reject rationalist externalism, his reasoning is not convincing. According to Kant, the nature of the will grounds the standards of its exercise, which may be expressed in its principle of end-selection and the necessary end of happiness. At the level of description at which it makes sense to speak of the rational intelligibility of an individual agent's actions or dispositions of the will, everything bottoms out in the conformity of those actions or dispositions to these standards, identified at Kant's explanatory level one. What it is for her actions to be intelligible just is for them to satisfy these standards. There is, therefore, no such thing as an agent's subjecting any of these standards, or the properties of the will in which they are grounded, to assessment for rational intelligibility, drawing on normative resources constituted externally to the will. There is no such thing as the agent's asking whether her naturally necessary end of happiness is sound and then having a belief about that. This is because this end is not an attunement of the will to reasons constituted independently of itself, but—if we want to use the idiom of "reasons"—what, alongside the law, fully determines what reasons the agent has.

Of course, questions about why our will is so constituted—questions at explanatory level two—are perfectly apt. But the answers to them do not express reasons which agents have for possessing or exercising a will with the internal standards in question; that thought makes no sense within this framework. Rather, such answers describe the place of the will and its exercise in the life of the kind of being to which it belongs. If you are tempted to say that these explanations (if they work) work only on the assumption that *we have reason* to live a life so described, you have already rejected Kantian constitutivism. You are presupposing an explanatory framework in which the nature of our capacities is not the source of normativity. Perhaps your presupposition is true, but it is not by itself an argument.

For the Kantian, the fundamental question is simply not that of what I have reason to do, but that of what my will, what my practical reason, is: what I, a finite practically rational being, am. And if the will is, in its essence, directed toward the satisfaction of inclinations, then it does not matter whether or to what extent they are borne of reason. They need do nothing, explanatorily speaking, to establish standing to play a role in our rational life. For the will's prerogative is absolute. If our desires are summoned to the court of reason, then that is where they must go.

## ACKNOWLEDGMENTS

I wish to thank, for comments, conversation, and correspondence, Carla Bagnoli, Sam Berstler, Matthew Boyle, Pirachula Chulanon, Sanford Diehl, Stephen Engstrom, Matthias Haase, Chandler Hatch, Timothy Hinton, Mathis Koschel, Maya Krishnan, Jed Lewinsohn, Michael Pendlebury, Jacob Rosen, Karl Schafer, Jensen Suther, and Mary Tjiattas; the attendees of a January 2024 meeting of the German Philosophy Workshop at the University of Chicago, of a panel on constitutivism at the World Congress of Philosophy in Rome in August 2024, and of the Chicago Talks on Agency conference in April 2025; two anonymous referees, whose suggestions led to considerable improvements in the paper; and, above all, Christine Korsgaard.

## REFERENCES

Citations of Kant's works (except of the Critique of Pure Reason, in citing which I supply A/B pagination) are to the Academy Edition. Translations are based on those of Guyer & Wood (1992-2016) unless otherwise noted.

## Works by Kant

[A/B] (1781/1787). *Kritik der reinen Vernunft*.
[*Volckmann*] (1784-5). *Metaphysik Volckmann*.
[*Grundlegung*] (1785). *Grundlegung zur Metaphysik der Sitten*.
[*Anfang*] (1786). *Mutmaßlicher Anfang der Menschengeschichte*.
[*KpV*] (1788). *Kritik der praktischen Vernunft*.
[*KU*] (1790). *Kritik der Urteilskraft*.
[$L_2$] (1790-1). *Metaphysik* $L_2$.
[*Dohna*] (1792-3). *Metaphysik Dohna*.
[*Vigilantius*] (1793). *Moral Vigilantius*.
[*Religion*] (1793). *Die Religion innerhalb der Grenzen der bloßen Vernunft*.
[*MS*] (1797). *Metaphysik der Sitten*.

## Works by Others

Allison, H. (1990). *Kant's theory of freedom*. Cambridge University Press.
Bagnoli, C. (2022). *Ethical constructivism*. Cambridge University Press.
Beck, L. W. (1960). *A commentary on Kant's critique of practical reason*. University of Chicago Press.
Boyle, M. (2016). Additive theories of rationality: A critique. *European Journal of Philosophy*, *24*, 527–555. https://doi.org/10.1111/ejop.12135.
Boyle, M. (2022). The rational role of perceptual content. In M. Boyle & E. Mylonaki (Eds.), *Reason in nature: New essays on themes from John McDowell* (pp. 83–110). Harvard University Press.
Ebels-Duggan, K. (2023). Bad debt: The Kantian inheritance of empiricst desire. In D. Heide & E. Tiffany (Eds.), *The idea of freedom* (pp. 150–175). Oxford University Press.
Engstrom, S. (1992). The concept of the highest good in Kant's moral theory. *Philosophy and Phenomenological Research*, *52*, 747–780. https://doi.org/10.2307/2107910.
Engstrom, S. (2009). *The form of practical knowledge*. Harvard University Press.
Fix, J. D. (2020). The error condition. *Canadian Journal of Philosophy*, *50*, 34–48. https://doi.org/10.1017/can.2019.14.
Fix, J. D. (2021a). Practical cognition as volition. *European Journal of Philosophy*, *30*, 1077–1091. https://doi.org/10.1111/ejop.12709.
Fix, J. D. (2021b). Two sorts of constitutivism. *Analytic Philosophy*, *62*, 1–20. https://doi.org/10.1111/phib.12166.
Fix, J. D. (2022). The unity of normative thought. *Philosophy and Phenomenological Research*, *104*, 639–598. https://doi.org/10.1111/phphr.12798.
Gregor, M. (1963). *Laws of freedom*. Basil Blackwell.
Herman, B. (2007). *Moral literacy*. Harvard University Press.
Hills, A. (2006). Kant on happiness and reason. *History of Philosophy Quarterly*, *23*, 234–262.
Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge University Press.
Korsgaard, C. M. (2008). *The constitution of agency: Essays on practical reason and moral psychology*. Oxford University Press.
Korsgaard, C. M. (2009). *Self-constitution: Agency, identity, and integrity*. Oxford University Press.
Korsgaard, C. M. (2018). *Fellow creatures: Our obligations to the other animals*. Oxford University Press.
Lavin, D. (2004). Practical reason and the possibility of error. *Ethics*, *114*, 424–457. https://doi.org/10.1086/381695.
Marcus, E. (2020). Wanting and willing. *European Journal of Philosophy*, *29*, 887–899. https://doi.org/10.1111/ejop.12610.
McDowell, J. (1994). *Mind and World*. Harvard University Press.
Moran, R. (2022). Self-consciousness and self-division in moral psychology. In T. Schapiro, K. Ebels-Duggan, & S. Street (Eds.), *Normativity & agency: Themes from the philosophy of Christine M. Korsgaard* (pp. 95–125). Oxford University Press.
Pendlebury, T. A. (2022). The shape of the Kantian mind. *Philosophy and Phenomenological Research*, *104*, 364–387. https://doi.org/10.1111/phphr.12767.
Pendlebury, T. A. (2024). The will of all in Kant's *Groundwork*. *Kantian Review*, *29*, 423–445. https://doi.org/10.1017/S1369415424000347.

Reath, A. (2006). *Agency & autonomy in Kant's moral theory*. Oxford University Press.
Schafer, K. (2023). *Kant's reason: The unity of reason and the limits of comprehension in Kant*. Oxford University Press.
Schapiro, T. (2021). *Feeling like it: A theory of inclination and will*. Oxford University Press.
Scanlon, T. M. (1998). *What We Owe to Each Other*. Harvard University Press.
Timmermann, J. (2022). *Kant's will at the crossroads: An essay on the failings of practical rationality*. Oxford University Press.

**How to cite this article:** Pendlebury, T. A. (2025). The Kantian's desire. *Philosophy and Phenomenological Research*, 1-24. https://doi.org/10.1111/phpr.70041