Julian Wuerth

Guyer, the Grounding of Kant's Categorical Imperative, and the Elimination of Sensibility Procedure

Abstract: Paul Guyer interprets Kant's ethics as teleological, with humanity as the ultimate end. On this interpretation, Kant's Formula of Universal Law (FUL) on its own amounts to blind obedience to law. To make sense and to motivate, FUL must instead be interpreted as mandatory solely as a means to human freedom as spelled out in the Formula of Humanity (FH), so that FUL *rests on* FH. The current essay, however, identifies a *common* argument for *independent* FH and FUL formulations of the categorical imperative in what is termed Kant's Elimination of Sensibility Procedure (ESP). Kant's critical philosophy distinguishes in kind between sensibility and understanding, and Kant now uses this distinction to avoid the mistake that in his view has doomed every previous ethics: conflating the pragmatic and the moral. By systematically isolating the moral through ESP, Kant can finally clarify the form (FUL) and matter (FH) of the categorical imperative.

Keywords: categorical imperative, universal law, humanity

At the conference held at Brown University in honor of Paul Guyer on the occasion of his retirement from teaching, Reed Winegar asked Paul which single question he would ask Kant were Kant to return from the dead. Ignoring Fred Rauscher's sensible suggestion, "How did you return from the dead?," Paul, after some reflection, answered "What is the foundation of morality?" Three hundred years after Kant's birth and two hundred and twenty years after his death, Kant's exact answer remains enigmatic, even for arguably the greatest Kant scholar ever, and even though Kant arguably took the question of morality, "What should I do?," to be the most important question in philosophy. A year ago, in an email to me, Paul summarized his interpretation of Kant's answer to this question as I worked on my own for a forthcoming book, and he added that he looked forward to seeing my own interpretation — "even, or especially, if you differ from me!" This invitation to disagree with him captures much of the generous spirit with which Paul approaches philosophy and approached his role, decades ago, as my dissertation advisor. Much to his credit,

https://doi.org/10.1515/9783111544731-004

Paul was never interested in producing replicas of himself, much less disciples, but instead partners in philosophical inquiry and scholarship. In what follows,
I accept Paul's invitation, offering an interpretation that differs from his, even if there is much on which we agree. It is a pleasure and an honor to do so on this special occasion, made possible by *Kant-Studien*,[1] of honoring Paul for a lifetime (that is not over yet!) of exceptional achievement, on this three hundredth anniversary of the birth of Immanuel Kant.

I start with a review of Guyer's interpretation of Kant's criticism of the rationalist perfectionist tradition in ethics. As Guyer characterizes it, rationalist perfectionism falls short for Kant because it is purely formal, a problem Kant fixes by introducing into his own ethics a matter, or a good, on which to ground moral obligation. As Guyer understands it, this ultimate matter or good is for Kant humanity, which for Guyer boils down to freedom, and this is presented by Kant in his second formulation of the categorical imperative, the Formula of Humanity as an End-in-Itself (FH). On Guyer's interpretation, Kant's first formulation of the categorical imperative, his Formula of Universal Law (FUL), is by contrast with FH, and like the rationalist perfectionists' moral law, formal and empty. Obedience to FUL, without reference to FH, would for Guyer accordingly be "blind obedience" to law. It is only if and when we hitch FUL to FH, showing that the universalizability of maxims is a necessary precondition for respecting and promoting the matter, or content, of humanity, or freedom, that FUL gets off the ground with any justification: in Guyer's words in *Kant on Freedom, Law, and Happiness*[2], "obedience to universal law [is] mandatory solely as the necessary condition for the realization of human freedom" (KFLH, 1).

Guyer, as usual, musters strong textual evidence for his interpretation of Kant's views on both the rationalist perfectionist position and the relationship of the FUL to the FH, and I do not hope to refute his interpretation but instead, at best, provide a plausible alternative. I should also add that, while I provide a partial defense of the *foundations* of FUL, I believe that Kant's own formulation of FUL is highly problematic, especially its "contradiction in conception" test. My argument draws on an interpretation I have developed elsewhere[3] of a single basic underlying procedure across Kant's various arguments for the categorical imperative, grounded in my

1 Many thanks to Konstantin Pollok and Margit Ruffing for their fine editorial work on this volume.
2 Paul Guyer, *Kant on Freedom, Law, and Happiness*. Cambridge: Cambridge University Press, 2000. Hereafter "KFLH."
3 Julian Wuerth, "Categorical Imperative," in *The Cambridge Kant Lexicon*, ed. by Julian Wuerth (Cambridge, Cambridge University Press, 2021), 67–94; and Julian Wuerth, *Kant on Mind, Action, and Ethics* (Oxford: Oxford University Press, 2014).

interpretation of Kant's account of our mental faculties and theory of action, which I call Kant's Elimination of Sensibility Procedure, or ESP. I have argued that Kant views the history of ethics before him as united in its conflation of intellectual and sensible sources of the moral law and as thus united in its mixing together "in proportions unknown" of moral and pragmatic considerations, spoiling all efforts to identify a properly *moral* law. The rationalists, in particular, intellectualize sensibility, while the empiricists sensualize the intellect. Operating against the backdrop of his distinction in kind between sensibility and understanding, Kant's strategy with ESP is to systematically eliminate as inadequate all offerings of sensibility: our sensible feelings, our sensible desires, the objects of our sensible inclinations, and also our sensible cognitions of various moral codes in practice. Kant rejects these offerings of sensibility for a reason: they are not up to the demands of a categorical imperative or a good will determined by it. In this essay, I add to my ESP argument a new historical piece. I look first at some of Kant's views in the early 1760s and then at how Kant's ethics transitioned away from moral sense theory in 1769 with the advent of his distinction in kind between sensibility and understanding. At this point Kant gave up on his attempts to accommodate the categorical nature of moral imperatives within a moral sense theory, and he turned instead to an ethics of reason. Next, I argue that Kant was rejecting rationalist perfectionism in good part because of its failure to draw the distinction in kind between sensibility and understanding, which meant that its concept of perfection did not even limit itself to the intellectual but instead gestured vaguely toward an undifferentiated combination of moral and pragmatic goods. Kant, by contrast, starting from our shared understanding of either the value of a good will or the nature of a moral law as a categorical imperative for us, uses ESP to systematically reject sensibility and isolate pure practical reason, and this argument yields FUL. On my interpretation, then, one does not need to rest FUL on FH.

1 Guyer on Kant's Rejection of Rationalist Perfectionism and the Dependence of FUL on FH

Paul Guyer has written extensively in defense of his view that Kant's ethics is a form of teleology, whose telos, goal, end, or matter, is humanity, freedom, or autonomy — Guyer tends to use these terms interchangeably. This matter is what Kant presents in the second formulation of his categorical imperative, his famous Formula of Humanity as an End in Itself (FH). As Guyer sees it, Kant's first formulation of the categorical imperative, his equally famous Formula of Universal Law, which

specifies the *form* of morally permissible willing, is not an equal partner alongside the second formulation of the categorical imperative; if taken on its own, the first formulation would be left to insist upon "obedience to law for its own sake," or "blind obedience to law," and it avoids this fate only to the extent it is hitched to the second formulation as a means to the end set out in this second formulation, demanding universalizability of maxims merely as the necessary condition for satisfying the demands of the second formulation that we respect and promote humanity/freedom in ourselves and others. In the Introduction to his *Kant on Freedom, Law, and Happiness*, Guyer lays out his impressively bold thesis:

> Yet Kant has also seemed to many to advocate an insistence upon obedience to law for its own sake, an insistence upon blind obedience to law that in the twentieth century just ended has been associated with the destruction of everything for which the Enlightenment stood. A profound paradox can be avoided only if it can be shown that Kant intended obedience to universal law to be mandatory solely as the necessary condition for the realization of human freedom and through that freedom a systematic and unselfish distribution of happiness among all persons seeking a systematic union of their purposes in a world both natural and moral in which each and every person is treated as an end and never merely as a means. On this account, while all human beings must be treated as ends in themselves, the sheer fact of adherence to universal law is not an end in itself but is rather the means to the realization of the human potential for autonomy or freedom in both choice and action. (KFLH, 1)

Guyer's interpretation of Kant's ethics sees important seeds for Kant's mature ethics in Kant's 1764 *Inquiry Concerning the Distinctness of the Principles of Natural Theology and Morality*, otherwise known as his "Prize Essay" — even though Moses Mendelssohn won the actual monetary prize while Kant merely had his essay published along with Mendelssohn's as a consolation prize of sorts. In that essay, Kant, who at the time was still a moral sense theorist, faulted rationalist perfectionism, as practiced by Wolff and Baumgarten, for offering a purely formal ethics that was empty, pointing out that when regarded on their own, "no specifically determinate obligation flows from [the rationalist perfectionists'] two rules of the good" (UD, AA 02: 299). These two rules of good tell us to "perform the most perfect action in your power" and "abstain from doing that which will hinder the realization of the greatest possible perfection" (UD, AA 02: 299). But while useless on their own, these formal rules can be rendered useful — if, and only if, Kant tells us, they are "combined with indemonstrable material principles of practical cognition" (ibid.). At the time, as a moral sense theorist, Kant saw this material being supplied by our "unanalyzable feeling of the good" (ibid.). Surveying this material, Guyer identifies in it a two-part lesson that Kant takes to heart and never surrenders, namely, that a formal law is useless on its own and that a formal law must ultimately rely on a material law: "the key insight for Kant's eventual practical philosophy contained in this work is

then the idea that [i] moral philosophy too must begin with a fundamental though indemonstrable material or substantive principle, a fundamental source of value, and [ii] cannot rely on any purely formal law alone (KFLH, 8; see also 41f). I will return to the *Inquiry* later to suggest other lessons Kant takes from it for use in his mature ethics.

Of course, if Guyer wants to defend an interpretation of Kant's moral philosophy as recognizing a moral matter or moral end toward which moral choice needs to be responsive, he will need to defend himself against the inevitable objection that Kant's ethics is deontological, not teleological. Guyer does exactly this, examining Kant's well-known discussion in his 1788 *Critique of Practical Reason* in which he asserts what is often taken for the textbook definition of deontology, that the right must come before the good. As Guyer makes clear, however, Kant's point is *not* that the moral right has to come before a *moral good* but only that the moral right must come before, or be the condition under which we pursue, the good understood as the *empirical, pragmatic good, of our happiness*. As long as the sought-after unconditionally-valuable good is found, which would require that we go outside of the conditioned empirical realm, Kant's concern about the good coming before the right would be addressed and the good would *not* need to come after the right. Guyer (KFLH, 132f) quotes Kant on this point:

> The concept of the good as of an object does not determine and make possible the moral law, but on the contrary the moral law first determines and makes possible the concept of the good insofar as it deserves this name.
>
> This remark, which concerns merely the method of the highest moral inquiries, is of importance. It explains at once the ground responsible for all confusion of philosophers in regard to the highest principles of morals. For they sought for an object of the will in order to make it into the matter and ground of a law (which would then be the determining ground of the will not immediately but only mediately through the object that is brought to the feeling of pleasure or displeasure), when they should have first sought for a law that would determine the will immediately and an object only in accordance with this law. Now whether they placed this object of pleasure, which was to yield the highest concept of the good, in happiness, in perfection, in moral feeling, or in the will of God, their principle was always heteronomy, and they must unavoidably have stumbled into empirical conditions for a moral law (KpV 5:64).

I agree with Guyer fully on this point: Kant is by no means opposed to a "good" understood as an "end," or a "matter," *per se*, serving as the goal of a moral imperative; he is instead opposed only to having an *empirical* good, end, or matter, rooted in incentives and the goal of happiness, serve as the goal of a moral imperative. We see the same point made by Kant in the *Groundwork*, in so many words, when he also addresses what he later cites in the *Critique of Practical Reason* as "the ground responsible for all confusion of philosophers in regard to the highest principles of morals":

> When we look back upon all previous attempts that have been made to discover the principle of morality, there is no reason now to wonder why they one and all had to fail [...]. For when man is thought as being merely subject to a law (whatever it might be), then the law had to carry with it some interest functioning as an attracting stimulus or as a constraining force for obedience [...] duty was never discovered, but only the necessity of acting from a certain interest [...] [so that] the imperative had to be always conditional and could never possibly serve as a moral command. (GMS 04: 432f)

Here Kant makes the same point, namely, that all previous moral philosophies have failed, and he explains that they have failed because they have not identified a law of *pure practical reason*, which *we* ourselves recognize as valid; instead, previous moral philosophies have understood us to be subject to laws imposed from *without*. This relates to the point in the cited passage from the *Critique of Practical Reason* because, as Kant here explains, if we do not ourselves recognize the validity of a moral imperative, we will only subject ourselves to it from a "certain interest," i.e., *on the basis of our subjective empirical incentives, and so with an eye toward our own happiness*. Not surprisingly, then, and consistent with Guyer's point, when Kant searches, in the *Groundwork*, for an end with absolute value that could serve as the end set out by a moral law, he accordingly specifically rejects "*subjective* ends, which rest on incentives" (GMS, AA 04: 427, emphasis added), while embracing "*objective* ends [or "matter" GMA, AA 04: 436], which depend on motives valid for every rational being" (GMS, AA 04: 427, emphasis added). Humanity is this objective end, or matter, that depends on motives valid for every rational being. Because "practical principles are formal when they abstract from all subjective ends," but "material [...] when they are founded upon subjective ends" (GMS, AA 04: 427), and because the Formula of Humanity abstracts from all *subjective* ends, it follows that, even though the Formula of Humanity supplies the *objective* end, or matter, for the categorical imperative, it is nonetheless in the sense here described, a "formal" practical principle consistent with Kant's rejection of material ethics.

As Guyer sees it, Kant's presentation of the Formula of Humanity, with its objective end, or matter, is therefore the moment where Kant comes full circle from his discussion in the *Inquiry*. There Kant argued that the rationalist perfectionists' formal principles were empty without combination with "material principles" (UD, AA 02: 299), and now, Guyer argues, Kant has found an end, or matter, in which to ground his mature moral principle and give it content, in effect supplying the Formula of Universal Law with the content it needs.

A few more details need to be reviewed regarding Guyer's account, however. First, when Guyer speaks of the end, matter, or source of value in Kant's ethics, he tends to cluster a number of closely related concepts, namely, those of humanity, freedom, autonomy, and good will. I do not have space here to properly defend a position on the rich question of the relation of these four concepts, but I will say a few

things. First, humanity, as the combination of the higher faculties of cognition, feeling, and desire that together constitute rational nature in humans and allow us to set ends for ourselves, is for Kant free, and so humanity also entails the *capacity* for a good will and autonomy, even if we can never know for certain whether someone (even ourselves) has ever had a good will (e.g., GMS, AA 04: 407; V-Anth/Collins, AA 25: 24; MS, AA 06: 447). It is this capacity, specifically, for a good will that comprises the worth of humanity, so that without this capacity we would lose our worth and descend to the rank of mere animality. In the *Metaphysics of Morals*, Kant explains that if we ceased to have moral feeling, which is what "make[s] us aware of the constraint present in the thought of duty," we would become "morally dead," and "then humanity would dissolve (by chemical laws, as it were) into mere animality" (MS, AA 06: 399–400). Here we would not strictly be reduced to animality, because animality, as defined by Kant, lacks self-consciousness, which we would here retain along with any number of our higher faculties. But we would be *like* animality in one key regard, in that we would now lack the capacity to act out of respect for the moral law, and so we would now, like animals (defined by Kant not biologically, but simply as beings lacking self-consciousness and so lacking the capacity for morality) lose all *worth in ourselves*. Kant explains this point in the *Critique of Practical Reason*, making clear that even if we retain the capacity for setting ends for ourselves and the capacity for using reason in the process of setting ends for ourselves, we would still lose all value if we could use reason only *instrumentally*, as a "slave of the passions," to use Hume's words, to pursue happiness alone and not the higher purpose of morality: "That he has reason does not in the least raise [a human] in worth above mere animality if reason only serves the purposes which, among animals, are taken care of by instinct; if this were so, reason would be only a specific way nature had made use of to equip man for the same purpose for which animals are qualified, without fitting him for any higher purpose" (KpV, AA 05: 61). So the good will is at the center of Kant's account of value, being something he and everyone else, in his view, recognizes as self-evidently valuable without qualification, while we likewise all recognize the value in itself of humanity insofar as we recognize it to be a capacity to set ends and have a good will. Here are some supporting quotations, first from his lectures on ethics from the year of the *Groundwork's* publication, in which he repeats his famous claims in the *Groundwork* about a good will: "a good will is simply good without restriction, for itself alone, in every respect and under all circumstances. It is the only thing good without other conditions" (V-Mo/Mron II, AA 29: 599). As Kant explains in the *Groundwork*, it is because humanity is capable of morality that it can have value in itself, or "dignity": "morality is the condition under which alone a rational being can be an end in himself [...]. Hence morality and humanity, insofar as it is capable of morality, alone have dignity"

(GMS, AA 04: 435). So, as I have very briefly argued, while the concepts of humanity and a good will are closely related, a good will is the only thing with unqualified value in itself, while humanity also has value (but not unqualified value, given its capacity for moral failure) in itself, *because* it has the *capacity* for morally worthy choice. At the end of the day, I am not sure whether this interpretation squares with Guyer's.

Next, and importantly, Guyer argues not only that Kant recognizes a matter, end, or telos of sorts — again, he refers mostly to freedom, though also to humanity, autonomy, and good will — but also that the Formula of Universal Law *relies on* the Formula of Humanity and its specified end to "give any rational being a reason to adhere to the principle of morality" (KFLH, 145). In other words, when it comes to the Formula of Universal Law, with its demand that we act on universalizable maxims, there is nothing to "*possibly motivate us to act on the basis of such a law*," Guyer suggests, unless we rest FUL on FH and the value of humanity, or freedom. Accordingly, we ought to act on universalizable maxims (per FUL) *because* this is how we respect and promote humanity (per FH), so that FUL acquires normative authority only to the extent it is now recognized to specify the means (universalizable maxims) to the end laid out in FH, of respecting and promoting humanity.

2 An Alternative Approach: Kant's Rejection of Rationalist Perfectionism and the Grounding of FUL through the Elimination of Sensibility Procedure

I've now briefly sketched Guyer's interpretation of Kant's views on how form and matter work together in his ethics. Guyer sees Kant rejecting rationalist perfectionism in his 1764 *Inquiry* as formal and empty and proposing a solution to this problem with the introduction of matter into this rationalist perfectionist account — at this early point in Kant's development, the matter contributed by moral feeling. Guyer sees echoes of this strategy in Kant's *Groundwork*, arguing that Kant again views his formal Formula of Universal Law as incomplete, providing no reason or motivation for acting morally, and that it is only the second formulation of the categorical imperative that offers this basis and motivation for the moral law with its account of the matter, or end, of humanity. What I now propose is an alternative narrative spanning the same period considered by Guyer, from the *Inquiry* to the *Groundwork*. While this alternative opposes Guyer's view that there is no reason or

motivation to act on FUL without appeal to FH, it may well be compatible with Guyer's interpretation in most other regards.

As Guyer says, Kant's 1764 *Inquiry* criticizes the rationalist perfectionists for presenting a purely formal account that offers no results on its own, and Kant later repeats this criticism of the rationalist perfectionists in many other places, adding that matter needs to be introduced for this perfectionist account to yield results. But there is another criticism that Kant offers of rationalist perfectionism that he will continue to offer throughout his critical philosophy as part of his most sweeping criticism of all previous ethics. This is the criticism that we saw Kant making, above, in the *Groundwork* (GMS, AA 04: 432f) and the *Critique of Practical Reason* (KpV, AA 05: 64): that all systems of ethics prior to Kant failed because they did not properly distinguish, in kind, between the moral and the pragmatic, because they did not properly distinguish, in kind, between sensibility and understanding.

We can see the crude beginnings of this criticism of the rationalists' conflation of the moral and pragmatic in a source from the same year as Kant's *Inquiry*, namely, his 1763–64 *Lectures on Ethics Herder*, and in the *Inquiry* we find additional seeds for Kant's eventual, deeper criticism of the rationalists' conflation of the moral and the pragmatic rooted in Kant's later, mature distinction in kind between sensibility and understanding. First, after Kant introduces a distinction between moral feeling and practical feeling in his 1763–64 *Lectures on Ethics Herder*, he attacks Baumgarten and Wolff for failing to recognize the distinction between moral and practical perfection within their perfectionism: "this distinction is bungled by Baumgarten […] though everything he says may make for great practical perfection, it does not constitute moral perfection. The latter he omits to define, according to the taste of Wolff, which continually based perfection on the relation between cause and effect, and thus treated it as a means to ends grounded in desire and aversion" (V-PP/Herder, AA 27: 16). While Kant here frames the distinction between the moral and the practical within his merely pre-critical moral sense theory — following the moral sense theorists in their distinction between "self-interested feeling" (V-PP/Herder, AA 27: 03) and (against the "selfish-theorists" such as Hobbes) an additional, "disinterested," or "noble" moral feeling (ibid.) — Kant nonetheless has managed to introduce an important theme that will remain a fixture throughout his later, critical thought: the view that the perfectionists and all other philosophers have failed to properly distinguish the moral from the pragmatic. Moving to Kant's 1764 *Inquiry*, we see Kant for the first time drawing his famous distinction between hypothetical and categorical imperatives, though the terminology differs: "Now, every *ought* expresses a necessity of the action and is capable of two meanings. To be specific: either I ought to do something (as a *means*) if I want something else (as an *end*), or I ought *immediately* to do something else (as an *end*) and make

it actual. The former may be called the necessity of the means (*necessitas problematica*), and the latter the necessity of the ends (*necessitas legalis*)" (UD, AA 02: 298). Kant then tells us that "the first kind does not indicate any obligation at all" and then adds — to support Guyer's point about the need for an end in itself — that "They cannot be called obligations as long as they are not subordinated to an end which is necessary in itself" (UD, AA 02: 298f). Kant has now juxtaposed, within his moral sense theory, the distinction, on one hand, between "self-interested feeling" and the "noble" (moral) feeling, and, on the other hand, the distinction between hypothetical and categorical imperatives. The problem is that talk of categorical imperatives, or imperatives that rest on "an end which is necessary in itself" (UD, AA 02: 298), is incompatible with Kant's view expressed in the same source, in keeping with the views of Hume, that our "feeling of the good" never provides insight into "a thing absolutely but only relatively to a being endowed with sensibility" (UD, AA 02: 299). In other words, *moral* imperatives are supposed to be unique in speaking of an end that is necessary *in itself*, while our moral feeling can register value only *relatively*, subject to the nature of our sensibility. In other sources from the 1760s, Kant does make attempts to square the subjective nature of moral feeling with the need to identify an end that is "necessary in itself," following Rousseau in arguing for the special, universal, and even "unerring" nature of *some* of our feelings, namely those that are *natural* moral feelings rather than those feelings corrupted by *custom*: "My reason can err; my moral feeling, only when I uphold custom before natural feeling" (V-PP/Herder, 27: 06). But whether Kant was really convinced even at this stage that feelings could identify ends necessary in themselves is unclear, especially when one considers that, a mere page after writing about moral feeling and ends that are necessary in themselves in the *Inquiry*, Kant concludes the work with a *question* of whether feeling or cognition determines the first principles of practical philosophy (UD, AA 02: 300). In sum, at this point in the early 1760s, Kant is already faulting the rational perfectionists for not adequately distinguishing between moral and pragmatic imperatives, and his first talk about a categorical moral imperative and ends that are recognized as necessary in themselves hints at an even deeper divide to come between the moral and the pragmatic, where reason, and not moral feeling, achieves insight into the moral imperative.

To make a long story short, Kant's switch away from a moral sense theory grounded in sensibility to an ethics of reason happens in 1769–70, with his critical turn to a distinction-in-kind between sensibility and understanding in his 1770 *Inaugural Dissertation* and various *Reflexionen*. In a *Reflexion* from 1772, Kant explains the modal challenge: "The whole difficulty in the dispute over the *principium* of morality is: how is it possible to have a categorical *imperativus*, which is not conditional, under neither problematic nor apodictic conditions (of skill, prudence)"

(R 6725, Refl, AA 19: 141 [1772]). In another *Reflexion* likely from the same year, Kant argues that the moral law's modal status as necessary rules out moral sense as its possible source: "If such a moral sense were possible, then necessary, categorical, and universal laws could not be grounded on it" (R 6754, Refl, AA 19: 149 [1772?]). In a *Reflexion* from 1769–70, Kant explicitly rejects Hutcheson's moral sense theory: "Hutcheson's principle is unphilosophical, because it introduces a new feeling as a ground of explanation, and second *because it sees objective grounds in the laws of sensibility*" (R 6634, Refl, AA 19: 120 [1769–70]). And in the same 1769–70 period, Kant makes clear that he now views *reason* as the source of moral imperatives: "morality is an objective subordination of the will under the motivating grounds of reason" (R 6610, Refl, AA 19: 107 [1769–70]).

And now we return to the rationalist perfectionists. In addition to rejecting moral sense theory once and for all, Kant now, in the *Inaugural Dissertation*, also rejects the rationalist perfectionists. And while he will continue to fault them for offering an empty formalism, his main focus here is on their failure to appropriately distinguish between the moral and the pragmatic. In his earlier *Lectures on Ethics Herder*, we saw Kant level the same basic criticism, but now this failure is due also to the rationalists' failure to draw a distinction in kind between understanding and sensibility. Rather than view the distinction between sensibility and understanding as a distinction in kind, they view it as a distinction merely in degree of logical distinctness: "the sensitive is poorly defined as that which is *more confusedly* cognized, and that which belongs to the understanding as that of which there is a *distinct* cognition. For these are only logical distinctions which *do not touch* at all the things *given*, which underlie every logical comparison. Thus, sensitive representations can be very distinct and representations which belong to the understanding can be extremely confused" (UD, AA 02: 394), Kant explains. Indeed, in Kant's view, were logical distinctness the criterion for determining what is sensible rather than intellectual, the line might even be drawn in the wrong place altogether, given that some sensible cognitions are more distinct than intellectual ones. In order to draw this line in the right place, one needs to instead consider the point of *origin* of these objects of cognition. Here Kant makes this point, faulting Wolff in particular. What's more, he now draws the connection to his ethics, underscoring the crucial role that this distinction plays there:

> Each and every one of these cognitions preserves the sign of its ancestry, so that those belonging to the first group [sensibility], however distinct they be, are called sensitive because of their origin, while those belonging to the second group continue to belong to the understanding, even though they are confused. **Such, for example, is the case with *moral* concepts, which are cognized not by experiencing them but by the pure understanding itself**. But I am afraid it may be that the illustrious WOLFF has, by this distinction between what is sensitive and what belongs to the understanding, a distinction which for him is only logical,

> completely abolished, to the great detriment of philosophy, the noblest of the enterprises of antiquity, the discussion of the *character of phenomena and noumena*, and has turned men's minds away from that enquiry to things which are often only logical minutiae (UD, AA 02: 395, boldface added).

In the later (1781/87) *Critique of Pure Reason*, Kant famously criticizes both the rationalists and the empiricists for failing to draw a distinction in kind between sensibility and understanding, with the result that "Leibniz **intellectualized** the appearances, just as Locke **sensitivized** the concepts of the understanding" (KrV, A 271/B 327), and what we have seen here in the moral realm with Kant's assessment of the moral sense theorists and Wolff is the same point: on the empiricist end, the moral sense theorists, including the early Kant himself, sensualized the intellect, understanding reason's moral law to be a form of moral sense; on the rationalist end, Wolff intellectualizes sensibility, not properly isolating understanding and its unique moral law from sensibility because he construes the distinction as one in logical distinctness only. Neither the rationalists nor the empiricists will accordingly be able to properly isolate the intellectual from the sensible and in turn the moral from the pragmatic. The result will be that neither will isolate what Kant will understand to be the unique voice of pure practical reason, and as a result both will fail.

Here Kant's challenge is accordingly not the skeptic, we will see. Kant instead thinks that we all know the moral law, even if this moral law is not always clear. And Kant also thinks we all accept its authority, even if we do try to free ourselves from its specific demands through a process of rationalizing that Kant refers to as the "natural dialectic" (GMS, AA 04: 405). In defense of Kant's optimism about our recognition of a moral law, it can be noted that even Hume kicks off his *Enquiry Concerning the Principles of Morals* by asserting on the first page that it is only "disingenuous disputants" who have ever "denied the reality of moral distinctions," and that it is not conceivable "that any human creature could ever seriously believe that all characters and actions were alike entitled to the affection and regard of every one." Instead, Kant's challenge will be to confront the ubiquitous failure amongst his predecessors and peers to draw a distinction in kind between sensibility and understanding, a failure manifesting itself in the conflation of the moral and the pragmatic. Kant's resulting strategy? With an eye to this distinction in kind, as well as to other sub-faculties of sensibility and understanding, Kant will focus his readers on the unique unconditioned, categorical nature of moral imperatives and the unique unconditioned and unqualified good of a will determined out of respect for the moral law. With his readers focused on these unique, unconditioned features of the moral law and action on it, Kant will then be able to guide them through a systematic, not haphazard, rejection of sensibility as the possible source of the moral law. Having thus isolated pure practical reason, Kant will finally be positioned to

clarify the nature of its moral law. This is the argumentative strategy that Kant uses across his mature ethics, which I refer to as Kant's Elimination of Sensibility Procedure. It is a process that Kant also likened to chemistry in his *Critique of Practical Reason*, where he tells us that we should, in approaching moral judgments, adopt "a process similar to that of chemistry, i.e., we may, in repeated experiments on common sense, separate the empirical from the rational, exhibit each of them in a pure state, and show what each by itself can accomplish" (KpV, AA 05: 163; cf. GMS, AA 04: 388f). In taking his readers through this process whereby they first reject sensibility, Kant will in effect have his readers recapitulate his own evolution in ethics.

If we jump ahead from 1769 to the mid-1770s, to Kant's *Lectures on Ethics Collins*, we can see this new strategy in action, in the form of an argument for an early version of the categorical imperative that provides us with an overview of the strategy in *Groundwork I* and *Groundwork II*. Kant takes it for granted that we to some extent know the moral law and accept its authority. The task is instead one of clarifying the *source* of the moral law — and specifically whether morality is grounded in sensibility or understanding — and of clarifying what the *nature* of this moral law is. Kant thus asserts that "morality rests either on empirical or intellectual grounds," before setting out to determine which of these it is, and he also asks "what, then, is the supreme principle of morality whereby we judge everything, and in what way does moral goodness differ from every other sort of goodness?" (V-Mo/Collins, AA 27: 252). He turns first to empirical grounds, to systematically reject all of them, telling us first that "If a system of ethics is based on empirical grounds, it rests either on inner or outer grounds, drawn from the objects of inner and outer sense" (V-Mo/Collins, AA 27: 253). The inner grounds are physical feeling, or self-love, and moral feeling, with Kant listing Epicurus among the proponents of the former and Shaftesbury and Hutcheson among the proponents of the latter. Kant rejects both of these inner empirical grounds because of their *contingency*: regarding the possibility of the moral law resting on self-love, Kant explains by way of *reductio* that if the imperative to not lie rested on self-love, then the imperative "would run: You are not to lie only if it brings harm your way, but if it profits you, then it is permitted"; likewise, if the imperative rested on moral feeling "then anyone not possessed of a moral feeling so fine as to produce in him an aversion to lying would be permitted to lie" (V-Mo/Collins, AA 27: 254). These are the inner grounds, which concern the sensible faculties of feeling and desire and on which Kant will focus in *Groundwork I* before he there turns to practical reason stripped of such sensible feelings and desires, i.e., before he turns to what is effectively *pure* practical reason. Kant next moves on, in the *Lectures on Ethics Collins*, to the *outer* empirical grounds, of education and government, with which he will begin *Groundwork II*. Here again, the contingency of the imperatives disqualifies them on intuitive

grounds, with Kant again arguing by *reductio*: "were the prohibition of lying to rest upon education or government, then anyone educated or living under a regime where lying is permitted, would be at liberty to lie" (V-Mo/Collins, AA 27: 254). These "outer empirical grounds" concern the sensible faculty, in particular, of *cognition*, on which Kant focuses at the start of *Groundwork II* before turning to practical reason, at which point he will there reject *instrumental* practical reason because its hypothetical imperatives are, again, contingent and incompatible with our idea of a moral imperative. What we see in Kant's *Lectures on Ethics Collins*, then, is that a lot of work *precedes* the point where — in the later *Groundwork* — we might try to extract the FUL from pure practical reason. Here Kant takes it for granted that we all agree that there is a moral law, and that we all agree that the authority of a moral law should not hinge on arbitrary empirical factors such as whether our government, our system of education, our feelings of morality, or our feelings of self-love just so happen to endorse this moral law. This rejection of such inner and outer empirical grounds leaves our practical reason free of all such arbitrary and possibly distorting influences, which means that our practical reason is now guided by the concern — anticipating FUL's test of impartiality — to have "all acts of my choice concur with universal validity" (V-Mo/Collins, AA 27: 257).

This brings us to the *Groundwork*, Kant's criticisms of the rationalist perfectionist Wolff, and Kant's argument leading up to FUL. In the *Groundwork*, Kant once again criticizes the manner in which the rationalist perfectionists' account is *empty*, consistent with Guyer's interpretation, telling us, for example, that the rationalists' concept of perfection "is empty, indeterminate, and hence of no use for finding in the immeasurable field of possible reality the maximum sum suitable for us" (GMS, AA 04: 443). No doubt this example of a purely formal ethical code would benefit from the addition of some matter, and now Kant would be able to offer something other than the moral feeling of the *Inquiry*, namely, the matter of humanity presented by Kant in the FH. But our question is whether Kant's own FUL has a similar fate, of requiring the matter of humanity as a reason or motivation for FUL. And on this score we find that the *Groundwork* once again points to the way in which the problem of the emptiness of rationalist perfectionism could be addressed, instead, by recourse to the distinction in kind between understanding and sensibility — without relying on the matter of humanity. Turning to the Preface to the *Groundwork*, we find Kant there arguing that the rationalist perfectionists have failed precisely because they have not properly cordoned off a priori reason from merely empirical motives. In other words, Kant here criticizes the rationalist perfectionists *not* for having *too little content*, as he of course also does elsewhere, but once again for having *too much empirical content*. In other words, he is criticizing the rationalists for being *bad rationalists*, i.e., for *not being rationalist enough*, allowing empirical

content to mix in indiscriminately under the heading of "indistinct" understanding for lack of a distinction in kind between understanding and sensibility:

> they do not distinguish the motives which, as such, are presented completely a priori by reason alone and are properly moral from the empirical motives which the understanding raises to general concepts merely by the comparison of experiences. Rather, they consider motives irrespective of any difference in their source; and inasmuch as they regard all motives as being homogeneous, they consider nothing but their relative strength or weakness. In this way they frame their concept of obligation, which is certainly not moral, but is all that can be expected from a philosophy which never decides regarding the origin of all possible practical concepts whether they are a priori or merely a posteriori. (GMS, AA 04: 391)

Kant also draws an analogy to general logic versus transcendental logic, telling us that

> Wolff has not taken into consideration any special kind of will, such as one determined solely by a priori principles without any empirical motives and which could be called a pure will, but has considered volition in general [...]. And thereby does his propaedeutic differ from a metaphysics of morals in the same way that general logic, which expounds the acts and rules of thinking in general, differs from a transcendental philosophy, which treats merely of the particular acts and rules of pure thinking, i.e., of that thinking whereby objects are cognized completely a priori" (GMS, AA 04: 390).

A look at Kant's *Critique of Pure Reason* helps us to appreciate Kant's point here. There Kant tells us that

> [...] a distinction between pure and empirical thinking of objects could also well be found. In this case there would be a logic in *which one did not abstract from all content of cognition*; for that logic that contained merely the rules of the pure thinking of an object would exclude all those cognitions that were of *empirical content* [...] while general logic, on the contrary [...] deals only with the form of the understanding" (KrV, A 55/B 79–A 56/B 80).

Kant's point is that general logic does not consider the *pure thinking of objects* (to the exclusion of empirical thinking of objects), as does *transcendental* logic, but instead sticks to rules for thinking in general; and that, by parallel, rationalist perfectionism *does not consider the pure, properly moral contributions of reason* (again, to the exclusion of empirical incentives), as does Kant's moral philosophy, but instead sticks to "volitions in general." While transcendental logic, Kant's a priori ethics, general logic, and perfectionistic rationalism, are therefore all *formal* in one sense, of being stripped of all *empirical* content, transcendental logic and Kant's a priori ethics *offer pure (but not empirical) content*, providing new insight into how reason thinks a priori about objects and volition.

Turning to FUL, we find that Kant approaches this formulation of the categorical imperatives from two starting points. In *Groundwork I* Kant famously starts

from what he takes to be the universally shared conviction that a good will, or action from duty in our case (as finite rational beings), has special, unconditioned, and unqualified value. This is a *self-evident* starting point for Kant, not a conclusion reached by argumentation, placed boldly in the first sentence of the *Groundwork*, and it has pulled in readers for over two hundred years rather than alienating them. Kant never wastes a moment with talk of a skeptic about the value of a good will, and, indeed, when discussing examples of a heroic good will, he often describes the way in which we one and all take inspiration from the hero in question and wish to be more like them — even if not in their dangerous predicament. We readily recognize the unconditional value of the good will and are presumably interested to learn more about the *nature* of the moral law that directs such a will, *not whether there is a moral law at all*. Once Kant has taken us through the remainder of his usual Elimination of Sensibility Procedure, systematically (and not haphazardly) touring the various sub-faculties of the faculty of sensibility and rejecting self-love and moral feeling as possible sources of determination of such an unconditionally valuable good will — just as he had done in his *Lectures on Ethics Collins* when he toured our inner empirical grounds, rejecting self-love and moral feeling as inner empirical grounds — he has now effected the elimination of that part of our sensibility that might guide practical reason in a merely *instrumental* manner, so that he has now "deprived the will of every impulse that might arise for it from obeying any particular law" (GMS, AA 04: 402). It is at this point, after all of this argumentation, that Kant presents the Formula of Universal Law (FUL) for the first time, asserting that "there is nothing left to serve the will as principle except the universal conformity of its action to law as such, i.e., I should never act except in such a way that I can also will that my maxim should become a universal law" (GMS, AA 04: 402).

We now return to the beginning: Guyer's point about the relationship between the FUL and FH. In Guyer's words,

> Kant has also seemed to many to advocate an insistence upon obedience to law for its own sake, an insistence upon blind obedience to law that in the twentieth century just ended has been associated with the destruction of everything for which the Enlightenment stood. A profound paradox can be avoided only if it can be shown that Kant intended obedience to universal law to be mandatory solely as a necessary condition for the realization of human freedom (KFLH, 1).

What we have now seen, in addition to Guyer's strong evidence for his interpretation, is another, partly overlapping possibility. Kant's early engagement with the rationalist perfectionists revealed Kant's desire to add substance to the rationalist perfectionists' formal account, as Guyer notes, but it also revealed Kant's determination to draw a sharp line where none existed in the perfectionists' account

between pragmatic imperatives and moral imperative — moral imperatives of a *categorical* sort. With the introduction of his distinction-in-kind between sensibility and understanding around 1769, Kant could now systematically identify all sensibly-grounded imperatives per se as inadequate to the categorical demands of morality. And so when Kant discusses the moral law directly (as in *Groundwork II*) or the good will as determined by the moral law (as in *Groundwork I*), he can now confidently and safely guide us from each of these starting points, with an awareness always of the unique, unconditioned nature of the moral, through a systematic process of elimination of those faculties of sensibility whose products have in Kant's view adulterated all previous ethics. Guiding us from the self-evident starting point of the unconditioned value of a good will (or the unconditioned demands of the moral law) past these empirical shoals, Kant brings us to the high ground of pure practical reason, of practical reason now operating *without* the leading — or misleading — strings of empirical desires that compromise our impartiality. Kant chooses to couch this resulting impartiality in terms of universality, and certainly his at-times tortured account of "contradictions in conception" — and just as importantly the uncritical reception of it on the part of some of Kant's interpreters — invites questions about blind rule worship or blind deference to Kant, but what he describes is (as ought to be the case, given that Kant adamantly insists that he never dreamed of inventing a new moral law, but only to render the moral law clearer (KpV, 8n)) a new and arguably improved version (in its inclusion of duties to self and both negative and positive duties to all) of familiar, well-worn, and well-respected, if unclear and incomplete, moral concepts of impartiality and fairness as found in the following: the Golden Rule, Hume's Judicious Spectator, and Rawls's Original Position, among many others. We could move to justify the FUL as a necessary means to FH and the value of humanity, but recall that the value of humanity is at least arguably attributable to humanity's capacity for a good will, whose goodness would then, again, be presupposed within FH. So now the advantage of one strategy is to render FUL a means to respecting and promoting humanity (as something capable of a good will), and that of the other to render FUL a form of willing that itself constitutes a good will. Either way, the unconditioned value of the good will is self-evident and universalizable willing is our moral path forward, and on that harmonious note I will end this essay, on the happy occasion of honoring both Immanuel Kant and Paul Guyer.

References

Guyer, Paul (2000). *Kant on Freedom, Law, and Happiness*, Cambridge.
Wuerth, Julian (2014). *Kant on Mind, Action, and Ethics*, Oxford.
Wuerth, Julian (2021). "Categorical Imperative." In *The Cambridge Kant Lexicon*, ed. by Julian Wuerth, Cambridge, 67–94.