CHAPTER 5

Guyer on the Value of Freedom

Patricia Kitcher

I  Guyer’s View

Over a number of years, Paul Guyer has argued that the ultimate value for Kant is not mere adherence to the moral law, but freedom and the morally deserved happiness that only a free creature can enjoy. By highlighting what he takes to be the essential roles of freedom and (warranted) happiness, Guyer aims to correct the common picture of Kant’s moral thinking as excessively focused on duty and laws. On his account, the parade case of a deontologist turns out to have his gaze firmly directed to the final end of the creation (and of morality) and the notorious happiness-denier fully to understand that humans need happiness. I agree with the broad outlines of Guyer’s kinder, gentler Kant, but I have some questions about the arguments he uses to support his interpretation; I also offer some friendly amendments to his account to make it (I hope) more distinctively Kantian.

Although I am sympathetic to Guyer’s broad aims I argue against two claims that he makes in *Kant on Freedom, Law and Happiness*. The first concerns the relation between freedom and the moral law:

Kant sees that moral reasoning must begin with a morally necessary end, his candidate for which is the fundamental value of freedom itself, and then argues that conformity to law is the means to the preservation and promotion of freedom. (Guyer 2000c:10)

I think it takes us too far from Kant’s basic moral position to suggest that following the moral law should be understood as a means to *anything* else, except perhaps a world where everyone follows the moral law. The second claim lays out the relations among freedom, happiness, and end-setting:

[Kant] can and must allow the idea of a rational and systematic conception of happiness, as the inevitable object of the free and systematic choices of human beings, to return into morality as the ultimate object of human
action that is defined for it by the fundamental value of freedom – as the capacity, after all, to set ends – itself. (Guyer 2000c:11)

Despite the distinguished adherents of this view, I doubt that Kant equated “freedom” with the “capacity to set ends.” Thus, I question whether he would welcome the “return” of the idea of a system of happiness as the goal or object of free action.

Guyer makes his case for the ultimate value of freedom and (warranted) happiness by appealing to two issues: the implications of the role of the formula of humanity in the system of categorical imperatives, and the long-standing importance of the idea of the *summum bonum* (in which happiness is proportional to desert) to Kant’s philosophical project. I have concerns about these two lines of argument, singly and collectively. Although the formula of humanity (FH) is based on the fundamental metaphysical “fact” of human freedom or autonomy, the fact on which all values depend, and although FH enjoins humans to further the ends of others, I doubt that any of this implies that Kant thinks that the ultimate locus of value seems to be universal happiness itself, as the maximal realization of a natural human end, and the law of reason, together with the freedom from which it arises, seems to be valuable primarily as the necessary condition of this universal happiness. (Guyer 2000c:107)

Despite the fact that the *summum bonum* appears in all three critiques and in other important writings, including *Religion within the Bounds of Mere Reason* and the essay on Theory and Practice, it seems to have a particular – and conflicted – role to play in Kant’s system. It is part of what seems to be a circular argument for the necessity in believing in a moral Creator of the universe – an argument that he tries to keep in the practical sphere, while all the time noting that the belief is theoretical. In the present context, I will argue that the reasoning behind the *summum bonum* argument for maximizing happiness is so different from the argument for FH that it is probably a mistake to think of them as having a common aim. After raising concerns about some of the specifics of Guyer’s proposals, I present what I hope is a defensible version of his basic interpretive strategy of rejecting the “deontological versus teleological” and the “he fails to realize the importance of human happiness” dogmas that have shaped generations of Kant interpretation.

### 2 Humanity as the Source of Value

FH has been enjoying great popularity among Kantian ethicists in recent years, but Guyer argues persuasively that a common view of the move from
the formula of universal law (FUL) to FH in section 2 of the *Groundwork* is mistaken. There are at least two versions of this view. On one, the starting assumption is that according to ordinary moral understanding, some things have at least conditional value. The argument is that if anything has conditional value, then something must have unconditional or intrinsic value, and of the possible candidates: objects of inclination, inclinations themselves, non-human animals, and rational animals, only the latter is plausible (see, e.g., Timmermann 2007:94–5). A second avenue to the same conclusion starts with the assumption that since all human action must have an end, and since all humans are bound by the moral law, then the end of acting on that law must be a common, universal, or objective end, and, again, the only plausible candidate for the end is humanity itself (Guyer 2000c:144). Guyer rejects the latter reasoning as both fallacious – why must there be just one objective or intrinsic end? – and as missing what is at issue at this point in Kant’s argument. Kant pivots to FH not (just) to find an end for moral action, but also to shore up the argument that humans are bound by FUL; he does not assume that the validity of FUL has already been shown.

Kant’s explanation for why he is taking up the issue of humanity (and the formula connected to it) supports Guyer’s reading. Kant writes:

> But we have not yet advanced so far as to prove *a priori* that there really is such an imperative, that there is a practical law, that commands absolutely of itself and without any incentives and that the observance of this law is a duty. (*GMS* 4:425)

A further passage in the transition from the discussion of FUL to that of FH seems to seal the case for Guyer’s reading. To answer the question of whether FUL is a necessary law for all rational beings, then

> [that law] must already be connected (completely *a priori*) with the concept of the will of a rational being as subject. But in order to discover this connection we must, however reluctantly, step forth, namely into ... metaphysics of morals. (*GMS* 4:426–7)

When we take the step, we find something that has absolute or intrinsic value, *viz.*, rational nature. As Guyer notes, this answer is not the last word, since it raises the question of what it is about rational beings that gives them intrinsic worth. He thinks that Kant’s answer to the question is that it is their capacity to set ends (Guyer 2000c:148–9).

---

1 Translations of the *Groundwork* are from Gregor and Timmermann (2011). Translations from the *Critique of Practical Reason* are from Pluhar (1996).
This answer has also been offered by Christine Korsgaard (Korsgaard 1996b:124–32) and Allen Wood (Wood 1999:124–7), but Guyer disagrees with the reasoning that leads them to their shared interpretation. As he notes, even if the ability to set ends stops the regress of ends as means to further ends, it must be modified by something like “an ability to set ends by a creature who adheres to the categorical imperative.” Otherwise, Korsgaard and Wood would be implausibly suggesting that Kant believed that the capacity to set arbitrary or bad ends was intrinsically valuable (Guyer 2000c:151).

Although Guyer’s objection that Korsgaard’s argument rests on an ambiguity between the capacity to set ends as value-conferring for any end, and value-conferring in the case of moral ends, his proposal seems to suffer from a similar ambiguity. He suggests that Kant’s argument for the intrinsic value of end-setting is direct:

[Kant] sees our capacity to set and pursue ends of our own choice as a fundamental manifestation of our freedom and freedom itself as possessing absolute value. (Guyer 2000c:151)

The ambiguity lies in the phrase “ends of our own choice.” Does that mean simply ends that are chosen “as one pleases” rather than being determined by immediate inclination (KrV A802/B830), or does it refer to ends that are chosen independently of any inclination? Alternatively, are we dealing with merely psychological freedom or with transcendental freedom (KrV A534/B562)? Since there is little reason to suppose that Kant thought that psychological freedom was intrinsically valuable, the answer must be that what is intrinsically valuable is the ability to set and pursue ends independently of any inclination. But that capacity is not a manifestation of freedom but freedom itself (see, e.g., GMS 4:446).

Guyer’s central thesis is that the fundamental, intrinsic value of freedom is the centerpiece of Kant’s ethical theory. He defends the interpretation by invoking the capacity to set ends, which is what permits him to argue that the goal of ethical action is the maximal realization of human ends. I have argued that this defense introduces unwarranted elements into Kant’s position and I will now try to show that these elements are also unnecessary for establishing Guyer’s larger interpretive claim about the essential role of FH in the proof of the validity of FUL.

The thesis that freedom is the key to the intrinsic value of rational agency is strongly supported by the result of Kant’s quest for an a priori connection between the concept of the will of a rational being and that creature’s
being subject to the moral law, a result presented in *Groundwork* III. In that section, he argues that freedom of the will can be nothing other than autonomy, that is, the will’s property of being a law to itself and [that] proposition indicates only the principle, to act on no other maxim than that which can also have as object itself as a universal law, [i.e., FUL].

\[GMS\ 4:446–7\]

That is, the *a priori* connection between the concept of the will of a rational being and the concept of a being who is bound by FUL is found in the concept of a free will. “Freedom” is thus the unique metaphysical answer to questions about both the bindingness of the moral law and the source of non-instrumental value.

Most Kant interpreters would agree with that claim, but few are content to rest with it. The problems are obvious. The alleged analytic connection between a free will and acting on the moral law seems dubious (although, see Schönecker 2013 for an influential defense of the analyticity thesis). Further, even if that point is granted, by the time of the *Critique of Practical Reason*, Kant had abandoned the argument that humans are bound by the moral law because they are, in fact, free. So there is an interpretive imperative to try to find a more solid argument for his conclusion. I, too, find this imperative compelling, but, for reasons already mentioned, I doubt that it helps to fasten on the capacity to set ends.

Kant’s sketches his argument for FH in section II. After noting that material ends, inclinations, and things are merely subjective ends – have worth only because someone values them – he highlights a special feature of human beings. Each human being necessarily represents his own existence as intrinsically valuable (\[GMS\ 4:429\]). He continues that “every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me” (\[GMS\ 4:429\]). He appends a note explaining that he is only postulating the principle here, but will supply adequate supporting grounds for it in section 3. Thus, what the final section is supposed to show is that each person regards himself as having intrinsic worth on the same ground that every other person regards himself as having intrinsic worth. To justify FH, however, he needs to show more, namely that each person must regard every other person as having intrinsic worth on the same ground that he regards himself as having intrinsic worth.

If Guyer is right that the fundamental value for Kant is freedom, then the proposition to be established in *Groundwork* III would be: Each person
must regard himself as free on the same ground that he does or must regard every person as free. When we turn to what I take to be the key text in that section, that is just the proposition to be proved. Interestingly, the proof is presented as something that will establish the validity of FUL, not FH. That textual fact is strong support for Guyer’s claim that FH plays a critical role in proving the validity of FUL – because the argument that completes the proof for FH is also the argument that establishes the bindingness of FUL for all humans.

Groundwork III is to eliminate the possibility that morality is chimerical by showing that the moral law is objectively valid for all rational beings, because they are free. Since the argument is based on freedom, freedom must be shown to be a property of all rational beings (GMS 4:447). Kant lays out how the proof works and its first steps, as follows: \(^2\)

1. One must prove it [freedom] as belonging universally to the activity of rational beings endowed with a will as such. 2. Now I say: every being that cannot act otherwise than under the idea of freedom is actually free, in a practical respect. . . 3. Now I assert: that we must necessarily lend to every rational being that has a will also the idea of freedom, under which alone it acts. For in such a being we think [denken] a reason that is practical, i.e., has causality with regard to its objects. 4. Now one cannot possibly think of a reason that would self-consciously receive guidance from any other quarter with regard to its judgments, since the subject would not then attribute the determination of judgment to his reason, but to an impulse. (GMS 4:448, my numbering and underscoring)

The lynchpin of the argument for the universality of freedom and so for the objective validity of the moral law is premise (3), the claim that we must “lend” the idea of freedom to any being we think as having a reason that is practical, and so as having a will, as the only idea under which she can act. Why the peculiar expression “lend” in place of the more straightforward “attribute”?

I think the answer to this puzzle can be found in a little-noticed discussion at the beginning of the Paralogisms chapter of the first Critique. Kant is going to use the results of his exploration of the requirements of thinking from the Transcendental Deduction to illuminate the errors of Rational Psychology, and he pauses to consider a possible objection:

---

\(^2\) The next seven pages draw on material that first appeared in my 2017a. See pages 224, 225, 228, 232, 234, 243.
It must, however, seem strange at the very outset that the condition under which I think at all, and which is therefore merely a characteristic of myself as subject, is to be valid also for everything that thinks; and that upon a proposition that seems empirical we can presume to base an apodictic and universal judgment, *viz.* that everything that thinks is of such a character as the pronouncement of self-consciousness asserts of me. The cause of this, however, lies in the fact that we must necessarily ascribe to things a priori all of the properties that make up the conditions under which alone we think them. (KrV A 346/B 404–5, my underscoring; see also KrV A 353–4)

That is, it is legitimate for Kant to use himself (or for any rational thinker to use herself) as a model for all rational thinkers, because that is the only way they have of understanding others as thinkers at all. Humans have no access to thinking through outer or inner sense. Rather, they come to understand thinking by engaging in it.

Consider making the judgment “nine” on the basis of counting or inferring that Socrates is mortal from the usual premises. Kant argues that these acts of rational cognition are possible only because the thinker is conscious in combining some representations in others. The reasoner must be conscious in concluding that Socrates is mortal from the premises, or her conclusion would not be an act of rational cognition, because she would not know the reasons for it (A 303/B 359–60); The counter must be conscious, in judging “nine,” that she does so on the basis of counting “one,” “two,” etc. Otherwise, she would have no idea why she is saying “nine” (KrV A 103–4). She would not be using a concept but parroting a phoneme. The act-consciousness that is essential to rational cognition is not sensory (KrV B 153). Cognizers do not know the reasons for their judgments or inferences because they have observed themselves thinking through inner sense, but because they have consciously performed acts of judging and inferring – as they must for these to be the rational acts that they are. If Kant is right about the requirements of thinking – and I believe he is – then although it is inaccessible to the senses, it is known through the doing of it (see my 2017b: 165–70 for further defense).

Under these circumstances, I can understand others as thinkers only through the one case I know, *viz.* my own. Kant is not committing the infamous single-case induction to solve the problem of other minds. He is explicit that the self-as-model for others move is not an empirical generalization, but merely *seems* to be one. It is, rather, a recognition from your own case that a special type of activity – thinking – exists, and that what it is, is actively, consciously combining some mental states in others. Alternatively, the point at issue in these preliminaries to the arguments of the
Paralogisms chapter is not how humans know that other humans have minds, but how they must understand what it is for another to have a mind.

Given this background, we see why Kant says (in 4) that no rational agent could think of “judging” as receiving direction from the outside. Any such agent must understand that that would not be judging, and that any “actions” or bodily movements that flowed from such impulses would not be rational actions. We can also see why he might use the odd locution “lend.” Kant cannot mean that we must think of others as temporarily or provisionally free, since the conclusion to be established is that they are free simpliciter. Rather he must mean that any rational agent will project on the action-producing judgments of others the same conscious activity or spontaneity that she enjoys when making action-producing judgments. On this reading, Kant would be providing just the argument he promised in section 2. Insofar as we take ourselves to have absolute or intrinsic worth because we recognize that we are free judgers – we would take any others we understand to be rational agents to have absolute worth on literally the same ground, viz., our recognition of ourselves as free judgers. In this way, he would be demonstrating that rational agents really are bound by FUL (via the argument that a “free will” and “a will subject to FUL” mutually imply each other [GMS 4:447]) by completing his argument for FH, as the texts that transition from FUL to FH claim he will, and as Guyer argues, he must.

Unfortunately, Kant has not shown that the will is free from all incentives. He assumed that practical reasoning was syllogistic: The major premise contained the law or principle of action; the minor, subsumption under the law; and the conclusion laid down what to do in the present circumstances (RL 6:313). But the argument about lending the idea of freedom to any rational actor concerns only the reasoning from premises to conclusion and not the source of the principle. He extends the argument as follows:

5. Reason must view herself as the author of her principles, independently of alien influences, and must consequently, as practical reason, or as the will of a rational being, by herself be viewed as free; i.e., its will can be a will of his own only under the idea of freedom, and must thus for practical purposes be ascribed to all rational beings. (GMS 4:448, my numbering)

Then he raises a devastating objection to the argument:

It seems as if in the idea of freedom we actually just presupposed the moral law, namely the principle of the autonomy of the will itself, and could not prove its reality and objective necessity. (GMS 4:449)
Given the alleged analytic connection between freedom and FUL, it would be inappropriate to try to prove one simply by assuming the other. In making the additional assumption (in step 5) of freedom or autonomy in the generation of the principle or maxim, he has thus begged the question (see Berger 2015 for a discussion of the apparently circular reasoning in *Groundwork* III). He was not, however, supposed to be assuming freedom, but arguing for it from the nature of rational willing. Since considerations about the necessary conditions for judging in general do not, however, imply (5), he has to simply assume it and thereby beg the question.

It is widely agreed that the failed argument of *Groundwork* III is recast in the fact of reason texts in the second *Critique* (i.e., in sections 6 and 7 and the Corollary to 7). As we have just seen, the argument failed, because Kant had no way to establish that reason must regard itself as the author of its principles of action (maxims) independently of alien influences. He returns to the issue of maxims in section 6 of the second *Critique*, after noting that knowledge of freedom and the moral law must begin with the latter:

> therefore, it is the moral law of which we become conscious directly (as soon as we frame or draft or pose maxims of the will for ourselves) which first offers itself to us, and which – inasmuch as reason displays it as a determining basis not to be outweighed by any sensible conditions and indeed entirely independently of them – leads straight to the idea of freedom. (*KpV* 5:29–30, my underscoring)

That is, when humans form a maxim or principle of action, they are immediately conscious of the moral law: “would they will a world where their intended action was universal law?” And they are not just conscious of the moral law, but conscious of it as determining their willing. That is the thesis to be established in the second *Critique* – not that pure reason has a hand in some action, which we know from the *Groundwork* could never be established with certainty (*GMS* 4:407) – but that pure reason is sufficient by itself alone to determine the willing (*das Wollen*) (*KpV* 5:15), i.e., to alter the subject’s motivational structure. Kant’s dramatic example of an individual who is threatened with immediate hanging if he refuses to give false testimony against an honest man provides a *Gedanken* experiment for his reader. In imagining being in that situation, the reader recognizes that, even when threatened with the extinction of all future pleasures, he still feels the pull of duty (*KpV* 5:30). Although not given a name until §7, that is the fact of reason doctrine – the fact that in deliberating about what to do, and so about what maxim to adopt, the
subject is conscious of the moral law as producing a determinate willing, a rational desire to follow or abstain from a particular course of action.

Despite the very bad reputation of the “fact of reason” doctrine, it is not an alien presence in Kant’s theory, something that arrived from Mars. It is implied by his account of moral motivation and by his theory of rational judging. By the argument of *Groundwork* I, ordinary moral agents take the moral character of an action to depend on its motive and they take an action to have moral worth only if it is motivated by duty (GMS 4:397–401). Ordinary rational agents also understand that a rational agent must be conscious of the grounds of her judgments and actions. It follows that, on the ordinary understanding, a subject can perform a morally worthy action only if she is conscious of its ground, and only if that ground includes her recognition that she can will a world that everyone does what she intends to do. The existence of the fact of reason is thus a consequence of his analysis of the ordinary human understanding of morality and of judging.

In section 7 the fact of reason is offered as a proof of the proposition that “you ought to act so that the maxim of your will could be a universal law.” The proof invokes no premises and makes no inferences. It appeals only to the reader’s consciousness in practical deliberation of the moral law as moving her willing. The Corollary to this section generalizes the efficacy of the moral law to all humans:

> Pure reason is practical by itself alone and gives (to the human being) a universal law, which we call the moral law. (KpV 5:31)

If this can be shown, then the promise of a proof for FH would finally be redeemed. Through its representation of the moral law, reason alone (i.e., without any incentives) would move the willing of rational beings, thereby showing them to be capable of free action and so have ultimate worth. Proving the Corollary would also finally establish the validity of FUL.

Kant opens the argument in support of the Corollary with the claim that the fact of reason is undeniable. Unfortunately, he then presents a step that we should reject and that he later rejects:

> Now, this principle of morality precisely on account of the universality of the lawgiving that makes it the formal supreme determining basis of the will regardless of all subjective differences of the will, is declared by reason at the same time to be a law for all rational beings insofar as they have a will at all, i.e., a power to determine their causality by the representation of rules. (KpV 5:32, my underscoring)
If reason declares that all rational creatures must have the moral law within, then it must be possible to infer the presence of an efficacious moral law from rationality. That view is, however, inconsistent with the claim he has just made, viz., that the fact that reason moves the willing through the moral law cannot be derived from anything else (KpV §31).

Kant could, however, defend the Corollary without this false step. All he needs to do is redeploys the line of reasoning that he uses in the prefatory material of the Paralogisms. Through the Gedanken experiment, the reader is reminded of what she already implicitly knows. When she deliberates about what to do, she is aware of considerations about whether she could will that everyone does what she proposes to do as moving her willing. That is how she knows that she has an efficacious moral law within/is free and so has intrinsic moral worth. It is also how she understands what practical reasoning is. Since practical deliberation is a species of thinking, Kant must hold that it is inaccessible by the senses. Hence, insofar as one human can understand another as engaging in practical deliberation at all, she must use herself as a model for that person.

Although the Corollary has the form of a generalization, it is plainly not empirically based. Normal humans do not conduct experiments on other humans to determine whether they are rational or moral agents. They simply presume that their conspecifics are like themselves in their capacities for thinking and practical deliberating. This presumption can be defeated, but when it is not, then they must model other humans’ practical deliberating on their own – because that is the only way that they can understand it at all. In that case, however, they would regard others as intrinsically valuable on exactly the same grounds that they regard themselves as valuable.

Kant did not extend the reasoning of Groundwork III in this way, but he could have, thereby finally filling in the argument for FH sketched in Groundwork II. Further, like the failed argument of Groundwork III, this argument for the validity of the moral law rests on a metaphysics of morals, on the non-empirical fact that humans have a will that can be moved by practical reason, and hence have absolute worth. Critics will protest that the fact of reason argument proves nothing more than that people are not aware of the forces directing their wills. A deliberator may believe that her consciousness of the moral law moves her willing against a proposed course of action, but how could she eliminate associative conditioning as the true source of the repulsion? Although this line of objection has problems – it seems to threaten the very possibility of rational action – I am not going to
pursue them. Kant thought that the fact of reason demonstration worked, and I want to consider what FH would enjoin if it were proven along the lines that he originally sketched. In particular, would it require moral agents to promote maximal human happiness?

According to the fact of reason, humans automatically respect the moral law at the level of thinking. That is the fact of reason: When a person deliberates, she is conscious of the moral law moving her willing. Applying the considerations raised at the beginning of the Paralogisms, it would follow that when humans encounter conspecifics, they automatically regard them as moral deliberators, as rational beings whose reflections on what to do are affected by their consciousness of the moral law (because they model them on themselves). It would thus also follow that humans always and automatically respect other humans in thought. The idea that Kant takes humans automatically to respect the moral law and other moral agents at some level is consonant with his many remarks that, however depraved someone is, he can never turn off or even turn down the voice of reason (e.g., \(KpV\) 5:35, 5:69). This inevitable, inescapable respect in thought is the basis of the duty for respect in action.

Kant traces a complex psychological route from thought to action and employs surprisingly purple prose to explain the effect on people of realizing that they have the moral law within, that they and all (non-defective) humans are moral lawgivers (\(GMS\) 4:435–6, \(KpV\) 5:87). I am going to avoid most of the complications of his psychology of respect and focus on just two, seemingly incontestable, elements of his account. The first has already been discussed: Humans have exactly the same reasons for taking others to be (in part) pure, holy, etc., as they have for taking themselves to have these qualities. The second is also implied by the preceding discussion: What gives humans their exalted status is their moral law producing practical reason. In the second \(Critique\), Kant argues from the lawgiver status of humans to FH. Precisely because every human is autonomous, no one can

\[
\text{use [such a] subject merely as a means ... [by] subjecting him to any aim that is not possible in accordance with a law that could arise from the subject himself who undergoes [the action], thus never to use this subject merely as a means, but always at the same time as himself a purpose.} \quad (KpV \ 5:87)
\]

To treat a person in a way that he could not will as a universal way of acting would be inconsistent with understanding him to be a moral lawgiver – as an arbiter of what is to be done.
I think that Kant is making the same point in his discussion of how FH rules out lying promises in *Groundwork* II:

For, he whom I want to use for my purposes cannot possibly agree to my way of behaving toward him, and so himself contain the end of this action. (*GMS* 4:429–30)

This discussion may seem to support the idea that FH requires containing others’ ends, and thus pursuing them and thereby enhancing their happiness. Telling the truth is, however, a strict duty, so the issue cannot be increasing happiness or even minimizing misery. A note clarifies the situation through a contrast with the negative version of the Golden Rule. The point is not about what people would hate having done to them. These matters can be subjective, whereas what is needed to ground duty is universal law. It is not that the victim would disagree on the subjective ground that he hates being lied to. It is that he cannot agree as a moral lawgiver and so could not contain the end of lying – in his own case or any other.

Kant discusses containing others’ ends in the application of FH to the fourth example, but he presents it as a somewhat indirect duty. There would not be 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 also be my ends, if that representation is to have its full effect on me. (*GMS* 4:430, my underscoring)

The last phrase suggests that what is good about promoting the ends of others is not the happiness thereby achieved, but the full recognition of others as ends in themselves.

Kant’s explanation of the duty of beneficence in the *Metaphysics of Morals* takes a different approach:

The reason that it is a duty to be beneficent is this: Since our self-love cannot be separated from our need to be loved (helped in the case of need) by others as well, we therefore make ourselves an end for others; and the only way this maxim can be binding is through its qualification as a universal law, hence through our will to make others our ends as well. The happiness of others is therefore an end that is a duty. (*TL* 6:393)

Here, the duty to promote the ends of others is not connected to their status as ends in themselves, and so to FH, but is defended via FUL. Further, the duty is based on a universal, but completely contingent fact
about humans, namely, their desire for happiness. Further still, the only way to make one’s contingent desire for happiness morally acceptable is to strive for the happiness of others. Hence, the duty to promote the happiness of others does not rest on the intrinsic value of happiness. Promoting happiness in general is an instrument for making one’s contingent desire for happiness morally acceptable.

The *Metaphysics of Morals* is both late and meant to elaborate the fundamental principles uncovered in the *Groundwork* in relation to the human condition. But the *Groundwork* itself presents helping the needy as an exemplary case of a moral duty in the context of illustrating FUL, as well as FH, and this discussion should not rest on contingent facts about humans. The example has been widely discussed and I am mainly going to follow Barbara Herman’s analysis (Herman 1984). If the needs in question are understood as “true needs” – that are virtually essential to the continuing capacity for autonomy – then the duty to help the needy stems not from the value of maximal human happiness, but from the value of preserving their status as moral beings. I say “mainly,” because the discussion of this case in relation to FH seems to add a further dimension. There is also merit in promoting inessential wants of people, because it is a recognition of their special status. In sum, although Kant clearly holds that it is a duty to promote the happiness of others, it is not at all clear that he takes the basis of that duty to be the intrinsic value of happiness.

3 God’s Justice

I will treat the issues surrounding happiness and the *summum bonum* much more briefly. Kant is inconsistent about whether the religious postulate arises from practical or theoretical reason. In the Canon of the first *Critique*, he famously asks three questions: What can I know? What ought I to do? What can I hope? (*KrV* A804–5/B832–3). The answer to the second question is that you should do that which makes you worthy of happiness, which is to try to bring into being a moral world – a world in complete accord with all moral laws. He then claims that the answer to the third question about hope is, like that to the second, practical. But he continues by explaining that

just as moral principles are necessary according to reason in its practical use, so it is equally necessary also according to reason in its theoretical use to assume that everyone has cause to hope for happiness insofar as he has made himself worthy of it in his conduct. (*KrV* A809/B837)
Since this result cannot be guaranteed through the operation of nature, the connection between worthiness to be happy and happiness can only be secured through a supreme intelligence, i.e., God. Is this argument for God supposed to be practical or theoretical, or both?

Kant suggests the third answer in a passage that lends credence to Christian Garve’s criticism that he maintains that the moral law cannot motivate without the promise of commensurate happiness (TP 8:279):

Reason finds itself compelled either to assume such a being . . . or to regard the moral laws as idle chimeras, because without this presupposition the necessary result that reason connects with these laws would have to vanish. (KrV A811/B839)

Here, he seems to say that without the theoretical presumption of a God, and so the distribution of proportional happiness, then reason would find it impossible to act on moral laws.

In the second Critique, Kant’s considered argument for the religious postulate brings the issues into sharper focus. The argument for God is practical, and not at all theoretical (see KpV 5:125), and there is a clear difference between the driver of morality (the Triebfeder [KpV 5:72ff.]), which can only be the moral law itself and the respect that is inseparable from the consciousness of it, and the ultimate object of moral action, viz., the highest good. Why must humans aim at the highest good and thus presume a moral Creator of the universe? Kant’s answer is clear:

Virtue is not yet . . . the whole and complete good as the object of the power of desire of rational finite beings. For, in order to be that, happiness too is required in addition, and this not merely in the partial eyes of a person . . . but even in the judgment of an impartial reason, which regards a person as such . . . as a purpose in itself. For, to be in need of happiness, and also worthy of it, but nonetheless not to partake of it is not at all consistent with the perfect volition of a rational being that also had all power, even if we only think such a being by way of experiment. (KpV 5:110, my underscoring)

But why should an ordinary, rational agent engage in this thought experiment, since, as far as he knows, there are no rational beings who are all-powerful?

In a Reflection, Kant observes that if the moral law obligates humans only because they believe in God and an afterlife, then it would be absurd to argue that they must believe in God and an afterlife in order to meet their moral obligations (R6432, AA 18:714). The belief in God and an
afterlife is what is creating the need for the belief in God and an afterlife! Why is it any less circular to argue that, since a world lacking in the highest good would be inconsistent with an all-powerful (and all-benevolent) Creator, then to avoid this inconsistency rational beings must believe in an all-powerful and all-benevolent Creator who would bring about a world in which the highest good is realized? The supposition of an all-powerful and all-benevolent Creator is the source of the inconsistency with a world where the morally worthy are unhappy, and so of the rational need to suppose the existence of an all-powerful Creator who ensures that they are justly rewarded. Without this supposition, unrewarded virtue is a matter of bad luck, not inconsistency.

If this argument were successful, then it would show that the goal of morality is maximal human happiness. Any moral agent must aim to produce a world in which everyone acts morally (KrV A808/B836; TP 8:280), and if happiness must be proportional to desert, then in a perfect world, everyone would receive his or her full measure of happiness, compatible with all others receiving their full measures. Further, even if unsuccessful, the argument can support Guyer’s interpretive thesis that Kant took maximal happiness to be the ultimate aim of morality – since he presumably accepted considerations based on the *summum bonum*.

Still, I am not convinced, because the pieces of the argument from FH and from the *summum bonum* fit uneasily together. In the latter argument, maximal happiness enters the moral calculus very differently from the way in which the “happiness” of others becomes a duty in relation to FH. According to FH, I must help those in need, either to prevent the loss of their autonomy or to show my respect for them. By contrast, the maximal happiness that would be part of the complete good is a matter of distributational justice. The difference in the bases for promoting the happiness of others parallels Stephen Darwall’s well-known distinction between two kinds of respect (Darwall 1977). The happiness I must promote on the basis of FH applies equally to all humans; the happiness deserved in a perfect world is a matter of the kind of person someone is.

Further, at least in the 1790’s, Kant offers theses that are incompatible with the interpretive claims that he believed that what is valuable about humans is their rationality (their ability to set ends), and that he took FH to be the basis of the duty of promoting the ends of others. As noted, in his final attempt to explain the duty to help others, in the *Metaphysics of Morals*, he does not rely on FH, but on FUL. I have some obligation to give you a dollar so that you have enough to buy the hot fudge sundae you
really want and would much enjoy, because I could not will a world in which no one helped those in want. This particular obligation has nothing to do with your rational capacity to set ends, however, and everything to do with your physical capacities to yearn for things and to enjoy them. The ability to set ends means that people have many projects for which they need the assistance of others, but for Kant, although the satisfaction of bringing such projects to fruition may be more intense or more permanent than that of eating ice cream, the pleasures are not different in kind \((\text{KpV} 5:23)\). More importantly, at least late in his life, he was clear that the ability to sustain rational projects is not what gives humans their ultimate worth. In \textit{Perpetual Peace} (1795), he explains that even a nation of devils could set up a state based on Hobbesian principles of rational self-interest \((\text{ZeF} 8:366)\). Presumably such folks are seriously lacking in ultimate worth.

For these reasons, I do not think that Kant took maximal happiness to be the ultimate value on the basis of his advocacy of FH. Still, it seems to me that the prominent place of the \textit{summum bonum} in his practical thinking does support Guyer’s more limited claim that he took human happiness to be valuable. Happiness must be supplied to meet the demands of cosmic justice (under the supposition of a moral Creator), but that reward would be appropriate only if happiness were valuable, or at least valued. Kant clearly thinks that all humans value happiness, but he may also think that happiness is valuable, under the obvious restriction, \textit{viz.}, that the desire for pleasure or success does not lead the person to contravene her duty. That condition is met when happiness is proportional to virtue, so in discussing the \textit{summum bonum}, Kant can endorse the value of happiness without in any way suggesting that it is good without qualification, and so a suitable basis for morality. In this way, Kant can find a place for the constrained but real value of happiness alongside the ultimate value of humanity.

Although I think the argument works differently than he does, I have also offered reasons in favor of Guyer’s view that the ultimate proof of FUL rests on the proof of the value of humanity. This result supports his broad interpretive claim that the deontological in Kant is inseparable from the teleological. I argued further that having the moral law within not only gives humans intrinsic worth, it also sets them the goal of creating a world where human behavior is governed by a system of common, objective laws. So the validity of FUL rests on a metaphysical fact about value – that humans have FUL within – and it creates the end of a morally harmonious world. Kant’s theory is not about value-free procedures that make sense
separately from metaphysical assumptions. Since being free is the same as having an efficacious moral law within, Guyer is also right that freedom is the ultimate value for Kant.

Exploring the exact roles of value, teleology, and happiness in Kant’s moral theory is crucial to adequately understanding it. Or this project certainly seems crucial now that Guyer has shown that, contrary to popular belief, these elements are not outcasts of his ethical thinking.