Courtney D. Fugate*

Problems with the Highest Good

Abstract: In this paper, I want to focus not on the problems that I believe may threaten Kant's account of the highest good, but instead on those that I believe threaten the majority of the interpretive reconstructions attempted by commentators and thus prevent the emergence of a consensus in the near future. My goal is to set forth exactly four problems to which I believe any successful interpretation or reconstruction of Kant's account of the highest good will have to provide substantive solutions and, in doing so, either break new theoretical ground or at least provide a more comprehensive methodology for sorting through the texts.

Keywords: Kant, highest good, virtue, moral philosophy

Kant's account of the highest good has been the subject of many and diverse interpretations ever since a version of it first appeared in the *Critique of Pure Reason*. Most have concluded that the argument is just plainly unsound, while a few, like myself, have taken up the challenge of discovering the suppressed premise or perhaps unnoticed distinction that would make the whole argument work out just right.[1] From this heroic – or perhaps quixotic – attempt by commentators, there has sprung up just as many different ways of framing the concept of the highest good and the function it is supposed to play in Kant's moral thought. The chief aim of all these attempts has been to successfully overcome the *prima facie* contradiction within a theory that demands we undertake virtue for its own sake, but then also argues it is necessary for us to believe this will be connected with an eventual reward, namely happiness.

As a result of these various ways of framing the highest good, we are currently further than ever from a consensus view on the answers to the following, basic questions: Does Kant have one or many conceptions of the highest good? If the

---

[1] I am aware of only two notable exceptions: Bader 2015 and Watkins 2010. Both find the argument straight-forward and both reconstruct it in close parallel to the Antinomy of Pure Reason as articulated in the *Critique of Pure Reason*. I will suggest below that the seeming ease of the argument on such readings directly results from neglect of the essentially practical nature of the concept of the highest good. See my previous treatments of this topic (Fugate 2014a, 2014b, esp. ch. 7).

*Corresponding author: Courtney D. Fugate, Florida State University, 600 W College Ave, Tallahassee, FL 32306, USA, E-mail: Courtney.fugate@gmail.com

latter, then which one of these is his central conception? What is Kant's philosophical basis for introducing the concept of the highest good? Does this concept violate the principle of autonomy and his doctrine of the moral incentive? How is it even possible for Kant to validly hold that we have a duty to promote, pursue or realize the highest good? Does this add a duty beyond the duty to act for the sake of law? Assuming we have such a duty, what exactly does it mean to promote the highest good, i.e. what is the practical content of such a supposed duty?

In this paper, I want to focus not on the problems that I believe may threaten Kant's account of the highest good, but instead on those that I believe threaten the majority of the interpretive reconstructions attempted by commentators and thus prevent the emergence of a consensus in the near future. As I see the matter, current interpretations that claim to overcome the *prima facie* incompatibility at the heart of Kant's theory of the highest good only appear to work because they frame its concept in a way that either ignores, or at least does not fully take into account, one or more of the four problems I will outline below. There is nothing unique or unifying about these four problems except that they have proven to be particularly difficult to resolve within a theory acceptable to modern commentators and yet, in my view, they must be resolved in any successful interpretation or reconstruction of Kant's account of the highest good that is faithful to the textual sources. To account for them properly, any future interpretation will therefore have to break new theoretical ground or provide a more comprehensive methodology for sorting through and excluding some of the relevant texts.[2] In general, and with only a couple of exceptions, I have avoided targeting specific authors, and have instead taken aim at the arguments themselves.

However, before explaining what these problems are, I must say something more about what I mean when I refer to the "highest good" or Kant's doctrine of the highest good. The reason for this is that many of the ways of framing the highest good found in the literature involve the contention that there are in fact two or more conceptions of the highest good in Kant's philosophy, one or some of which should be discarded or at least placed in the background in most contexts, so that the central one can receive its proper attention. Interpreters have distinguished between an immanent and a transcendent conception, a secular and a religious one, an ectypal and an archetypal one, the conception of the highest good as demand and as promise, as a duty and as an ideal, and, finally, between the highest good in a person and in a world.

---

2 To this point, this latter has been limited largely to unsubstantiated claims that Kant changes his mind at some point or did not understand his own views in one or another passage.

This last, however, is the only such distinction that has any direct basis in Kant's terminology and is the one I will be using throughout this paper. The rest arise from reconstructions aimed at solving some perceived problem in the texts. Staring with the first *Critique* and continuing right through to the *Religion,* Kant remains consistent in working with the concept of the highest good *in a world,* which he also sometimes distinguishes from that of the highest good *in a person.* Two quotations suffice to indicate the basic idea behind the intended distinction between these:

> Thus *happiness in exact proportion with the morality* of rational beings, through which they are worthy of it, alone *constitutes the highest good of a world* into which we must without exception transpose ourselves in accordance with the precepts of pure but practical reason … (A814/B842)
>
> Now, inasmuch as *virtue and happiness together constitute possession of the highest good in a person*, and *happiness distributed in exact proportion to morality* … *constitutes the highest good of a possible world*, the latter means the whole, the complete good … (AA 5:110-111, italics added)

By the highest good in a person, Kant thus means nothing else than a combination of the dual aims which he believes are shared by all human beings: we all feel the inner demand to become virtuous and we all naturally aim at our own happiness. This is each person's volitional outlook and it places no limitations on either of its components.

In contrast to this, the highest good of a world or of a possible world is "*happiness distributed in exact proportion to morality.*"[3] Two things should be noticed about this definition. First, in the literature, this is sometimes paraphrased as happiness distributed *according to* morality, or happiness *corresponding to* morality. Such paraphrases, however, would seem to possibly misrepresent what Kant says in such quotations by suggesting this is only about providing happiness to the morally good. But Kant's phrasing in both the passages above lays special emphasis on the distribution being *"ganz genau in Proportion"* to morality. The demand for an *exact* proportioning, as I will argue further below, is however otiose if there is to be no denial of happiness to the vicious. This is important because it points towards two further claims we find in Kant's discussion of the highest good: that this distribution of happiness presupposes an absolutely accurate assessment of the degree of virtue attained by each individual, and that the distribution that is part of the highest good includes both the granting of happiness *and* the denial of happiness, thus both reward *and* punishment.[4]

---

3 Kleingeld understands this distinction differently (2016, p. 34).

4 The centrality of this view in Kant is supported by numerous texts, e.g.: "Hence everyone also regards the moral laws as *commands,* which, however, they could not be if they did not connect

This last fact is so frequently forgotten in the literature that it is worth providing further context for this claim. In this connection, it is important to recall the general manner in which the highest good arises in Kant's practical philosophy. Kant roughly begins the argument by noting that the moral law demands perfect virtue above all else, or as what he calls the "supreme good." This good is evidently something demanded of us, something that is therefore an object of pure practical reason, and hence also something that depends upon the exercise of the free will. Consequently, this aspect of the highest good is anything but guaranteed by God's existence or by any other condition. As Kant often stresses, belief in the inevitability of virtue would indeed have a deleterious effect on our moral striving and so must be rejected in principle. Now, with the concept of the supreme good in place, Kant then goes on to ask if it can be thought to contain all possible good and hence to be not only the supreme, but also the complete good or the maximal object of pure practical reason. The answer is clearly no; happiness is also a good, albeit a conditioned one. Moreover, the moral law itself – for admittedly unclear reasons – prescribes an appropriate distribution of happiness such that violation of the law is connected to the idea of punishment (AA 5:37f.), while observance of the law is connected to the idea of a worthiness to receive happiness in proportion to virtue. Yet these two do not stand on the same footing. The virtuous have no right to demand happiness and so cannot expect it with confidence, since fulfilling one's duty is not meritorious.[5] Happiness as reward is to be deemed appropriate, and on the assumption of God's existence, also possible. The vicious, on the other hand, can very well expect punishment, since the latter is regarded as something deserved and required by justice itself. If God exists, and is just, then the vicious must expect punishment.[6] Now, this is the overall proportioning of happiness to

---

appropriate consequences with their rule *a priori,* and thus carry with them *promises* and *threats.* This, however, they could not do if they did not lie in a necessary being, as the highest good, which alone can make possible such a purposive unity" (A811f./B839f.). The discussions of the matter in the *Religion* and second *Critique* show that Kant's theory of the relationship between human beings and God, and hence of the precise constitution of the highest derived good, is complex, since it cannot be dogmatic. Consequently, one can say nothing with certainty about what the actual highest derived good will entail or how it can be realized, but only how we must think of it so as to make it best harmonize with the duty imposed upon us by the moral law. See, e.g., AA 6:69 and AA 6:488ff. The wider significance of Kant's "vindictive" conception of punishment is discussed in O'Connell (2014).

5 "[F]or one who only does what he owes [*seine Schuldigkeit*] can have no rightful claim on God's benevolence" (8:257n.).

6 Kant expresses the above views in detail in his lectures and private notes, e.g., AA 28:1085ff. This points up the central problem in the argument of Bader 2015, which relies on an interpretation of Kant's theses that there must necessarily be a connection between virtue and happiness as meaning that virtue must necessarily entail happiness. This would turn the connection between

virtue, which, according to Kant, constitutes the second aspect of the highest good, filling out the latter and completing it as the object of pure practical reason.

Three things should be noted about this general set of views. First, the concept of the highest good has virtue as its fundamental condition and virtue is something that must be regarded as dependent upon each individual's exercise of their own freedom. Hence, the highest good must not be interpreted, as it often is, as something that must be populated by virtuous individuals by its very definition. The highest good is not a future state of affairs that is guaranteed to obtain, but a normative model towards which we must strive, and which for the same reason must include within its very concept a dependence upon the exercise of our own freedom and so also the possibility of failure.[7] Second, the belief in the existence of God, which Kant introduces as a necessary consequence of our need to believe in the possibility of the highest good, does not guarantee that if we pursue virtue we will eventually reach happiness. This seems to be a common misunderstanding, due in part to Kant's failure to make his view clear in his central published writings. But the texts where he goes into the matter more deeply are unambiguous. Those who strive towards virtue can possess no right to happiness, but at most a justification to hope for it. God's benevolence must be understood as just that, namely, an unconstrained and free granting of good to those who do not at least make themselves unworthy of it. Third, God, who Kant calls the "original good" and who

---

virtue and happiness into a certainty and would render happiness a kind of payment (*merces*) for good conduct, which Kant forthrightly denies.

7 But if some, perhaps all, are not virtuous, and less happiness will be distributed than if this were not the case, does this not contradict the very concept of the highest good as also being the complete or maximal good? Kant's answer is no; for the punishment of the vicious, as demanded by justice, is itself good. Attempts to separate out descriptive and normative features of the concept of the highest good so as to create two distinct concepts are, I think, misguided. Kant seems well-aware that normative concepts cannot be separated from theoretical ones. This is precisely how the postulates of the second *Critique* are related to the moral law. His main contention is that these do not describe known things, but instead must be reduced to and understood in terms of the practical significance of our belief in them. The concept of the highest derived good is no different: If we regard it as an actual future state of affairs, then it either must be thought as populated by a number of virtuous individuals, perhaps all, or not. But this is a transcendent question, and as such has no meaningful resolution. This is the upshot of Kant's discussion of the possibility of a genuine theodicy (see esp. AA 8:257ff.). And yet, as Kant argues, there are practical grounds (rooted in the question: *What should I believe in order to best devote myself to pursuing the highest good?*) for believing, in various respects, both that all people will eventually become virtuous and that those vicious in this life will only get worse in the next (e.g., AA 28:1085). There is no question, in this regard, that it would undermine moral striving if we believed that in the highest good, which God guarantees, everyone will be virtuous by definition. It seems to me that we are yet to fully understand this complex feature of Kant's view of normativity generally, and of his account of the highest good in particular.

makes the highest (derived) good possible, remains such even if we should fail; for the punishment of this failure is itself good.[8]

I here contend that Kant always bases his arguments on this concept of the highest good in a world, and never on that of the highest good in a person or on any of the other conceptions of the highest good found in the literature. Hence, in the following, when I speak of the highest good, I always mean by this the conception of the highest good of a world.

1 The Problem of Heterogeneity

Some of the most influential interpretations of Kant's views on the highest good target the objection that the inclusion of happiness in the highest good is inconsistent with a central pillar of his moral theory, namely, his view that virtue consists in acting "from duty," or what is the same, in acting from the incentive of respect.[9] According to this rather common-sensical objection, if Kant's intention is to say that the highest good contains the idea that virtue will be rewarded with happiness, and vice punished with its opposite, then it can only serve to motivate us to be virtuous as a means to happiness; but this would spoil the moral incentive and violate autonomy.[10]

Interpretations targeting this objection generally argue that the happiness Kant is speaking of in connection with the highest good is not a kind of happiness that is contrary to the moral incentive or to autonomy. Instead, this is the happiness of others, or perhaps my own happiness, but only after I've been transformed such that what I want, and would make me happy, is just what is morally necessary. Indeed, beginning with Wood 1970, it has been repeatedly argued that the necessity of including happiness within the highest good is basically a direct consequence of Kant's understanding of the structure of rational willing.[11] Wood's basic idea is that since all maxims necessarily have both a form and a matter, and Kant – so it is argued – identifies happiness in the context of the highest good with the end or matter of our maxims, the maxims of a virtuous person must also aim at some object, even if this object is not the motive, and the total object of all such maxims is just what Kant means by "happiness" in the context of the highest good.

---

8 As Kant writes, "punishment in the exercise of justice is founded in the legislating wisdom not at all as mere means but as an end: trespass is associated with ills not that some other good may result from it, but because this connection is good in itself, i.e. morally and necessarily good" (AA 8:257n.).
9 This issue is thoroughly discussed in Denis (2005).
10 This view is already found expressed in Hegel, but its modern proponent is Beck (1960).
11 This view has been developed by many others, not always with reference to Wood, but most fully in the writings of Engstrom. See, e.g. Engstrom (1992).

A simple way to put this idea is that the happiness Kant has in mind is just the sum-total of whatever the virtuous happen to have as the object of their willing. I imagine the further, underlying suggestion here is that although virtue and happiness are at loggerheads for the morally imperfect, as one becomes more and more virtuous one will also increasingly come to desire the realization of a world in which virtue and happiness go hand in hand. In such a world, for example, we will not only tell the truth because it is our duty but will also desire to tell the truth and be pleased when it is told, because we have come to love honesty for its own sake.

A different version of this same basic idea is found in Silber (2012), esp. ch. V., and again more recently in Kleingeld (2016). Silber's idea is that the happiness we will in pursuing the highest good is the same happiness of others that is said to be a duty in the *Metaphysics of Morals*. But this cannot be right for two reasons. First, there is strictly speaking no such duty to be found in that work or elsewhere. The closest thing is what Kant calls the duty of beneficence, which consists only in the duty to help others when they are in need and to not tempt them into vice (AA 6:393f.). Kant's argument for the first part of this duty is roughly: We necessarily will that we are the end of others, i.e. that they promote our happiness when we need them to. But this cannot stand as a law unless we ourselves also include them as an end in our maxims, which means taking their ends as our own. Hence, we have in this sense a duty to promote the happiness of others. However, as Kant then goes on to stress, this really amounts to nothing more than a duty to assist others when they are in what he calls "true need," i.e. it amounts not to making them positively happy, or in providing them with happiness in proportion to their virtue, but in relieving them of clear distress with respect to their most basic needs. Moreover, according to Kant, this duty to help others in need can never, on pain of contradiction, require the sacrifice of one's own true happiness. Such a duty is clearly nowhere near equivalent to a duty to promote an exact proportioning of happiness to virtue as described in the highest good. Indeed, never in his explanation of beneficence does Kant even mention worthiness as a *condition* of providing such help, but instead seems to think that we should provide assistance to those in need simply for that reason, and not contingent upon an assessment of their worthiness of receiving our aid.[12]

---

12 Silber argues that one's own happiness, under the restriction of universality, gives rise to the demand to further the happiness of others and that this is the second element of the highest good. Note, however, that Kant instead *equates* one's own happiness with this second element. Yet Silber's solution still has the virtue that it attempts to derive this demand to pursue the happiness of others from the demand for one's own happiness, thereby recognizing that the root of the Antinomy is precisely this second concept of happiness. Kleingeld, by contrast, disregards this feature of Kant's argument, a feature which is not restricted to the treatment in the second *Critique* but indeed is found in nearly all the relevant texts. Based on this argument, I believe the promotion

Second, the duty of beneficence is just that, a duty of virtue: it is an end that one has a duty to promote. But this is not at all the way Kant frames happiness in the highest good. *That* happiness is the desire that all go according to *our* natural "wish and will." *This* happiness of *others* demanded in the duty of beneficence, by contrast, is pursued as an essential part of *becoming virtuous*, in other words, of becoming – as Kant so often states – worthy of *that personal happiness* contained in the highest good. The primary use Kant makes of rephrasing virtue as "worthiness to be happy" is precisely to indicate the conditionally of my own happiness on my own conduct, not the conditionality of others' happiness on theirs or on mine. Furthermore, I ask: Does this duty to provide aid to others really make sense as a possible stand-in for the happiness in the highest good? Are we to presume that, e.g., Kant's argument there is that we cannot be certain of our capacity to fully provide the aid that others need and so we must assume the existence of a God that would provide this aid in our stead? That seems highly implausible. Another problem with this view is that it does not at all account for the negative side of Kant's notion of exact proportioning, that is, there exists in Kant's writings no complimentary duty to deny help or even to punish (though, to be sure, there is a duty for a judge to punish). However, the greatest impediment to accepting any such interpretation must surely be Kant's repeated equating of the second element of the highest good with one's own happiness ("the second element of the highest good, namely one's own happiness" (AA 5:127); see also AA 5:112).[13]

After this look at Silber's suggestion, let us return to the more general idea that the happiness contained in the highest good is simply the object of a virtuous will as such. According to this idea, the virtuous must only aim at virtue, and happiness will be whatever they turn out to will in that case, and so not something that would adulterate their pure moral respect for the law. Unfortunately, there is a serious problem with this interpretation, which I believe cannot be resolved without doing an unacceptable amount of violence to Kant's texts. This problem stems basically from a fuller understanding of why, on Kant's view, the two components of the highest good, namely morality and happiness, are radically heterogeneous.

Let me take a moment to explain what I mean by such radical heterogeneity. In the Dialectic of the second *Critique*, Kant claims that in the concept of the

---

of others' happiness, which is surely a component of the highest good, is not equivalent to or based upon this duty of beneficence, but instead extends beneficence, based on other grounds, beyond what is strictly required.

13 By this I do not mean to deny that promoting the highest good also entails promoting universal happiness (see, e.g., AA 8:279). But this we are supposed to do because of duty, not because it would please us, and thus not according to the principle of self-love. Hence this promotion of others' happiness cannot be substituted for the supreme principle of one's own happiness, which Kant employs to generate the Antinomy and so to provide a deduction for the highest derived good.

highest good both virtue and happiness proportioned thereto are thought of as necessarily combined. Now the Greeks attempted to combine them, Kant explains, by choosing one sort of maxim, namely either one based on principles of happiness or one based on principles of virtue, and regarding the other as consisting only in the consciousness of them: "The Epicurean said: to be conscious of one's maxim leading to happiness is virtue; the Stoic said: to be conscious of one's virtue is happiness" (AA 5:111). Notice what Kant is arguing here. The Greek attempt to combine the two elements of the highest good was not by envisioning some ideal world or something in which the two elements are combined, but rather by seeing those two elements as resulting from one and the same supreme principle of their conduct.

Kant then explains that any such attempt to see virtue and happiness as aspects of one and the same maxim of action is bound to fail, because the two concepts are so "extremely heterogeneous." In the space of one page, Kant repeats the importance of this "heterogeneity" three separate times. In the last instance, he writes:

> Now, it is clear from the Analytic that the maxims of virtue and those of one's own happiness are quite heterogeneous with respect to their supreme practical principle; and, even though they belong to one highest good, so as to make it possible, yet they are so far from coinciding that they greatly restrict and infringe upon each other in the same subject. Thus the question, *how is the highest good practically possible?* still remains an unsolved problem … The Analytic has, however, shown what it is that makes the problem difficult to solve, namely that happiness and morality are two specifically quite *different elements* of the highest good … (AA 5:112)

Here, the problem of combining virtue and happiness in the concept of the highest good is framed in terms of a fundamentally irresolvable conflict of maxims. The maxims of virtue and the maxims of happiness not only do not coincide, they fundamentally "infringe upon" one another because they rest on incompatible supreme practical principles, namely, the moral law and the principle of self-love; to act on maxims of happiness is *eo ipso* not to act on maxims of virtue, and to act on maxims of virtue is *eo ipso* not to act on maxims of happiness. To have happiness as one's object here just means acting from the supreme practical principle of self-love, whereas to have virtue as one's object is to have the moral law for one's principle. This is extremely important. Many commentators understand Kant's argument to be: *We have a duty to further the highest good, of which happiness is a component. As part of this, we must promote happiness in proportion to morality, but we have no guarantee that this will be fully achieved. Hence, we need to believe in the existence of a God who will supplement our lack of power.* But the point Kant here makes about these two maxims, which he develops further into the

Antinomy of Pure Practical Reason itself, shows that his idea is altogether different. The real problem is not that our happiness is not fully within our power, but instead that following maxims of virtue and following maxims of happiness, and thus aiming at the virtue and aiming at the happiness envisioned as objects within in the complete highest good, are totally incompatible on the level of principles of conduct.[14] Put differently, our human capacity is insufficient to bring about the highest good in the world, not because we aim at but cannot fully achieve the happiness contained within it, but because virtuous human activity does not – in terms of the supreme principle of its maxims – even really appear to be aiming at our own happiness, at least not according to the empirical principles of the latter. In short, the two objects cannot be combined into one object, because, to have them both as object of one will would require that will to act from two different and fundamentally incompatible supreme principles.

The Antinomy explains that the problem comes down to this: "either the desire for happiness must be the motive to maxims of virtue or the maxim of virtue must be the efficient cause of happiness" (AA 5:113). Again, we must be careful about what Kant is and is not saying in this sentence. He is not saying that there is an antinomy because virtuous action does not lead to happiness in the empirical world, which could then be resolved by postulating that God will somehow make this so. Rather, what he is saying is that either the principle of self-love must also be the principle of virtue or the principle of virtue itself must be the supreme principle of happiness or self-love. This has nothing to do with empirical virtuous actions failing to reach their ideal goals. The conflict is rather one of maxims and hence of courses of action envisioned from the first-person perspective. Resolving the antinomy therefore requires a way of removing this conflict. But how? That maxims of happiness would result in virtue is impossible, as it runs counter to the chief result of the Analytic. But that the maxim of virtue would result in happiness *appears* impossible, because there is no known connection between the two. Indeed, in the common course of life, following the maxim of virtue very often infringes upon our pursuit of what we want and, when it is a case of a *moral* decision, ought to always supplant any concern for happiness whatever. Indeed, in the world as we know it, following maxims of virtue seems to be a pretty good way of forfeiting one's happiness. (Note that this claim, which Kant explicitly makes, is sufficient to show

---

14 It may sound as if this means Kant is here instead working with the concept of the highest good in a person and not that of a world. However, the latter concept, namely, happiness in exact proportion to virtue, the essential component of which is the limiting relation between the two elements, remains the same. What this analysis shows, however, is that Kant's conception of the highest good of a world is also conceived of as the ideal object of a certain kind of conduct in which the maxims aiming at its two components infringe upon one another. Thanks to an anonymous reviewer for asking me to clarify this point.

that by happiness in the highest good he does not mean the matter or end of virtuous conduct!) The resolution to this problem, finally, is that this last point is only true of the empirical world and will be false if – in the intelligible world – there exists a God who will make it such that happiness is exactly proportioned to virtue. In this case, we are not guaranteed that maxims of virtue are also maxims of genuine happiness, but only that these alone have the *possibility* of being such.[15] Kant thus claims that this resolution to the antinomy is "synthetic" precisely because it combines these two heterogeneous concepts – *concepts that we cannot conceive of as coinciding naturally at all in respect to principles of our conduct* – through an intelligible synthetic and causal connection first made *thinkable* by the existence of God.

What all of this serves to highlight is that the fundamental heterogeneity of virtue and happiness, the dual components of the highest good, is rooted in the fundamental heterogeneity of the two supreme principles of action, namely the principle of virtue and the principle of self-love. This is why, in the quotation above, Kant directs the reader back to the results of the Analytic. There it was shown that all principles of conduct ultimately reduce to one of these two supreme principles. The principle of happiness, which Kant also refers to as the principle of self-love, places the determining ground in the matter or object of the maxim, while the principle of virtue places it immediately in the formal suitability of the maxim to serve as universal law. This is a distinction not between objects of willing, but between principles of willing. Hence, a maxim aiming at any object of whatever kind, even if it should also be the same object that would be pursued by the most virtuous possible conduct, would fall under the principle of self-love if that object were taken as the determining ground of our maxims (cf. AA 5:34). Now, it is happiness in *this* sense – namely, as the principle of placing one's determining ground in the matter of one's maxims – that Kant employs to generate the Antinomy of Pure Practical Reason. Consequently, the view that Kant's inclusion of happiness in the highest good does not spoil the moral incentive simply because this happiness is nothing but the object of virtuous conduct itself is ruled out. For there are only two possibilities: A person is virtuous and so acts on the supreme formal principle of virtue, which means not acting on the supreme principle of self-love and so not aiming at happiness in this sense. Or, a person aims at realizing their own ends, i.e. happiness, no matter whether these are selfish ends or rather even the same kinds of ends that the virtuous would pursue, and so acts on the supreme principle of self-love. In the former case, which is the one envisioned by those who take the happiness in the highest good to be simply the end of virtuous conduct, the end of conduct is not the determining ground of

---

15 This I take to be the point of Kant's qualified statement that, understood in this way, virtue can be thought of as a doctrine of happiness, namely, as a doctrine of the hope for such (AA 5:130).

one's maxim and hence is not the happiness Kant uses to generate the Antinomy in the first place. And, in any case, if this interpretation were right, then it would at once show that there is no incompatibility between virtue and happiness and hence no heterogeneity and so also no legitimate ground for an Antinomy of Pure Practical Reason. Of course, I have not explained how Kant himself can entirely resolve this conflict when it is understood as a conflict of maxims; but that must be reserved for another time.

2 The Problem of Proportionality

In every substantive discussion of the highest good of a world, Kant stresses that it must include the concept of an exact proportioning of happiness to virtue.[16] This already appeared above in quotations from the first and second *Critiques*, and many more examples from the latter could be provided at will. In the *Gemeinspruch* essay, Kant writes similarly: "For it [i.e. the highest good] contains no prospect of happiness absolutely, but only of *a proportion* between it and the worthiness of the subject, *whatever that may be*" (AA 8:279n.).

In the *Religion* as well, Kant touches on this idea first in stating that "in following this idea, he [i.e. a person pursuing the highest good] might see himself in danger of forfeiting much in the way of personal happiness, for it is possible that he might not be adequate to what reason makes the condition for it" (AA 6:6). Then, later, he remarks that "if the strictest observance of moral laws is to be thought of as the cause of the ushering in of the highest good (as end), then, since human capacity does not suffice to effect happiness in the world proportionate to the

---

16 Marwede attempts to solve some problems raised by proportionality by denying it is a quantitative relationship (2018, p. 174ff.). However, this claim flatly contradicts both the textual sources and reason. In the tradition Kant is a part of, quantities are any properties that cannot be known except by comparison with something else (see, e.g., Baumgarten 2013, p. 112–13), and hence includes anything of which there can be more or less, a feature that Marwede admits applies to happiness. The very language Kant repeatedly employs and stresses, namely "exact proportioning," is also itself of mathematical origin. Marwede's main argument in this regard seems to be that Kant does not believe *we* possess an objective measure for happiness or virtue, and hence these things can have no objective quantity. This, however, is beside the point, since Kant maintains that God does have such a measure and ascribes omnipotence to the divine being precisely so that we can ascribe it complete knowledge of our exact degree of virtue and happiness. As for Marwede's textual claim that Kant certainly "nowhere" discusses degrees ("*Grade*") of happiness: "Glückseligkeit ist die Befriedigung aller unserer Neigungen (sowohl *extensive* der Mannigfaltigkeit derselben, als *intensive* dem Grade und auch *protensive* der Dauer nach" (B834). Here Kant not only does so, but also applies to it all the essential mathematical concepts from the principles of pure understanding elucidated earlier in the *Critique*.

worthiness to be happy, an omnipotent moral being must be assumed as ruler of the world …" (AA 6:6n).[17]

All of these quotations contain or imply two key ideas: First, as we saw earlier, the highest good – however we otherwise frame it – does not necessarily entail that all, or even anyone, will become worthy of happiness. In other words, the idea of the highest good contains virtue as a condition, not as a final state that we populate, as it were, by definition. Second, the happiness contained in the highest good is to be thought of as exactly proportioned to each person's worthiness – to use the phrase of the *Gemeinspruch* – *"whatever this may be."* What is more, these two ideas seem to be inseparable: If the concept of the highest good allows for the possibility that it is not populated by fully virtuous individuals, then such beings are not worthy of all happiness but only of that of which they have made themselves worthy. Likewise, talking about "exact proportionality" would make no sense at all if we were to assume that in the highest good everyone is necessarily perfectly virtuous. Such beings would obviously deserve all possible happiness. The suggestion that this is not true, because happiness is only a conditioned good and so more happiness than is warranted by one's degree of virtue would be bad, is incorrect for two reasons: first, Kant expressly states that there is no such limit on how much happiness perfect virtue would deserve[18] and, second, there is no reason to think that the conditionality of happiness as a good works like this, i.e. that happiness that does not result necessarily from virtue is also necessarily bad. All in all, there is no doctrine more prevalent or more straightforward in Kant's discussions of the highest good than that it includes an exact proportioning, and hence both reward *and* punishment.

Now, most recent interpretations attempt to make Kant's argument work either by equating the highest good with an ideal state in which everyone is virtuous, in which case all that is required from God is a supplement of happiness, or, as discussed above, by equating the happiness contained in the highest good with the matter or end of the virtuous. But neither of these, so far as I can see, is consistent with the doctrine of exact proportionality. In the first case, the demand for exact proportionality would be otiose. And in the second, it would make little sense; for if all happiness meant were the achievement of a more just world or a world in which *other people* are happier, then what sense would there be in God's preventing the realization of such a world as punishment for *my* vice? Put another way, if the happiness envisioned in the highest good is what the virtuous would strive for,

---

17 The ubiquity of this doctrine is also stressed in O'Connell (2012).

18 "The *worth* of a disposition *completely* conformed with the moral law is infinite, since all possible happiness in the judgment of a wise and all-powerful distributor of it has no restriction other than rational beings' lack of conformity with their duty" (AA 5:128).

then how can denying that be an appropriate punishment for vice? The demand for proportionality manifestly only makes sense when the happiness envisioned in the highest good is the happiness corresponding to the principle of self-love. And as we saw in considering the issue of heterogeneity, this is precisely how Kant frames it in the second *Critique*.

3 The Problem of the Determinations of the Attributes of the Divine Being

This problem is an extension of the preceding, but also shows how the problem of exact proportionality ties into the structural connection between Kant's moral and religious thought. Kant's moral proof of God's existence turns specifically on the idea that only a being such as God could fulfill the function, or serve as the cause, of the highest derived good. That is to say, the highest good is itself so constituted such that its possibility cannot be guaranteed by *any other object* than one answering to a very specific concept, namely, that of a being possessing omnipotence, omniscience, supreme justice, supreme benevolence, and so on.

This basic idea is the subject of a line of thought we find developed extensively in both the second and third *Critiques*. As Kant famously contends, all concepts employed by the Critical philosophy must be validated in some manner and their limits understood through their specific origin in reason itself. So any concept of God that might, for example, later play a role in his religious thought, must be properly examined and determined prior to its use. Now, speculative reason, Kant explains in the passages I have in mind, is unable not only to prove the existence of a God; by itself, it cannot even produce a valid concept of such a being that would deserve of being called "God" (AA 5:139). Theoretical considerations may, he says, be sufficient to justify the idea of an intelligent source of nature, or of a powerful source of nature, but not an omniscient or omnipotent one, let alone one possessing morally relevant attributes like the supreme justice and benevolence required for the God of religion.

What this means is that the very possibility of rational religion depends upon reason's ability to provide a valid ground for the determination of the concept of a deity. Now, as Kant further explains, this ground is located precisely and only in our duty to pursue the highest good insofar as this is understood to consist in happiness in exact proportion to virtue. Now, to fulfill this role, the being that provides the synthetic causal link between virtuous conduct and happiness must: "be *omniscient,* in order to cognize my conduct even to my innermost attitudes in all possible cases throughout the future; *omnipotent,* in

order to assign to this conduct the appropriate consequences; likewise *omni-present, eternal,* etc." (AA 5:140). In other words, the determination of God's attributes derives from this being's role as the basis of the exact proportioning of virtue to happiness. For example, in order to judge my inner self as to my dispositions, God must be intelligent, and to judge them all, he must be omniscient, and so on. As Kant states so clearly in the second *Critique:* "Thus the moral law, by means of the concept of the highest good as the object of pure practical reason, determines the concept of the original being as the *supreme being,* something that the physical (and pursued higher, the metaphysical) and so the whole speculative course of reason could not effect" (AA 5:140).[19] According to the third *Critique* and other texts, the concept of the highest good allows us not only to determine the properties of the deity, but even to determine their necessary order (AA 5:442-45; AA 8:257n.).[20]

We can see from this that the idea of an exact proportionality between virtue and happiness is such a basic part of the way Kant seeks to ground his religious thought in his moral philosophy that without it one could not – by Kant's own lights – think of the moral proof as a proof of a *God* at all. Moreover, this same connection is what underpins, ultimately, Kant's claim that morality, through the concept of the highest good, leads to "*religion,* i.e. to the cognition of all duties as divine commands" (AA 5:130).

The problem here is this: If a reconstruction of Kant's account of the highest good cannot account for his doctrine of exact proportionality – which, I repeat, includes both reward and punishment essentially – then it must also reject this fundamental argument regarding the determination of the attributes of the divine being and so also the structural linkage between Kant's moral philosophy and his religion. Alternatively, if one wants to argue that Kant's treatment of the highest good is altered in later texts such that it no longer requires the idea of such a God, then one will have to admit that by the same token he has torn away the foundation of his religious thought.

---

19 This line of thought is examined in Chance and Pasternack (2019).

20 Marwede (2018) rejects the connection of the proportionality thesis with the concept of divine justice in accordance with his view that the highest good only concerns providing happiness to the virtuous. Aside from the evidence against this latter view presented above, Kant clearly connects precisely divine justice with the limiting principle of happiness in the divine will (see, e.g., AA 5:131n.; AA 5:444). The connection of concepts is that in order for the highest good to be possible, there must exist a being who limits happiness exactly in accordance with virtue, and hence we must ascribe to God supreme justice.

4 The Problem of Choosing and Justifying One's Textual Basis

Perhaps the most fundamental problem facing commentators – because it must be answered before any of the previous can even be approached – lies in the solution to the question: *What should be the primary textual basis for interpreting Kant's doctrine of the highest good?* On the face of it, the answer to this question would appear to be straightforward. The *Critique of Practical Reason* contains, by far, Kant's longest and most detailed treatment of the concept. Yet, despite at least one claim to the contrary, the majority of interpretations, even those that proport to be based at least in part on the text of the second *Critique*, in fact ignore either the bulk or at least large parts of that very text. Often this is explicitly accompanied by the claim that such a choice is justified by the poorness of Kant's argument or by his having become confused about what he wanted to say. Instead, current interpretations mostly prefer to focus on the brief passage in the *Critique of Pure Reason* – a passage written prior to Kant's discovery of autonomy,[21] prior to his having developed a theory of the moral incentive, prior even to his discovery, or perhaps invention, of a dialectic of pure practical reason.[22] Due to some of the problems it raises the account in the second *Critique* has even been characterized as the worst and most misleading of all Kant's discussions of the highest good, or at least as resting on a conception of the highest good that does not fit the rest of his philosophy.[23] Beck's now classic treatment, which has been presented by many a subsequent author as a thumbnail sketch of Kant's entire doctrine of the highest good, is an example that possesses all these features. After essentially dismissing Kant's framing of the highest good and the dialectic, he provides a summary of Kant's argument that consists of nothing but a paraphrase of a mere half paragraph from the second *Critique* (see Beck 1960, p. 274 and for my criticism of his argument see Fugate 2014a, 2014b, esp. ch. 7). It would seem that, aside from a sentence here and there, the other thirty-odd pages of Kant's treatment can be ignored.

Now, why might passing over the finer details of the argument in the second *Critique* and relying instead on other texts be such a problem? I have already stated several reasons why relying on the first *Critique* in particular might be a one,

---

21 Bader (2015, p. 184) makes this point as well.

22 This is argued persuasively in Klemme (2010).

23 Again, Bader 2015 and Watkins 2010 are exceptions. Yet Watkins still regards what I think to be essential parts of the text as confusions on Kant's part.

namely, because it precedes the development of so many of Kant's most innovative and characteristic doctrines. But beyond this, it is notable that Kant never explicitly retracts any portion the argument of the second *Critique,* and never even attempts to provide an alternative foundation for this doctrine. The other relevant texts, such as the *Gemeinspruch* essay, the *Critique of the Power of Judgment* and the *Religion* all revisit or expand parts of Kant's earlier discussion, but none presents an entirely new argument for the place and role of the highest good in his philosophy. On the face of it, then, the second *Critique* should be regarded as containing Kant's "official" or formal basis for introducing this concept into philosophy and the foundation for its valid use. For, as is well-known, Kant himself regards transcendental deduction in general as having the essential function of determining and validating the use of a concept by laying forth its proper content, function and limits. Unlike many philosophers who are satisfied to employ commonplace concepts without scrutiny, Kant is adamant that fundamental concepts not drawn from experience must always be validated prior to use. Thus, any conception of the highest good that does not receive a deduction – at least from Kant's own point of view – should be regarded as illegitimate until one is provided. But as it happens, it is in the second *Critique* alone, that Kant claims to provide such a transcendental deduction of one particular conception of the highest good, namely that of the *highest good of a world understood as an exact proportioning of happiness to morality, or the worthiness to be happy.*[24]

Placing the text of the second *Critique* at the center of one's interpretation, based upon the above arguments, is nevertheless only one possible way in which this problem may be solved. It remains an open question whether there are other, perhaps stronger, arguments for taking other texts as central, or for segregating the texts into groups based upon changes in Kant's views, or even for adopting some entirely different method for sifting the evidence in the texts.[25] I thus conclude this section by stating this final problem in the form of a desideratum: Namely, that future reconstructions of Kant's account of the highest good either account for more than one or two brief passages from the second *Critique* – and preferably for the problems of heterogeneity, exact proportioning, and the derivation of the divine attributes – or else explain what basis not only they, but also Kant, has for employing whatever conception of the highest good they would like to propose.

---

24 AA 5:113, AA 5:126.

25 E.g., one might wish to develop a Kantian theory that omits certain doctrines, not because it is less Kantian, but because it is less philosophically justified.

5 Conclusion

Contemporary interpretations of Kant's doctrine of the highest good seek to overcome the *prima facie* incompatibility at the heart the highest good by employing one or a combination of four strategies, namely: (1) framing the concept of the highest good so that its second element, happiness, turns out to be just what the virtuous already desire; (2) minimizing those elements that most appear to offend against Kant's principle that virtue be sought for its own sake; (3) regarding the highest good as an ideal whole necessarily and fully populated by perfectly virtuous individuals enjoying complete happiness; (4) suggesting that we should disregard certain texts that would appear to be central to Kant's considered account. In discussing problem 1 above, I showed that the first strategy fails on a close reading of the Dialectic of the second *Critique*, because the *prima facie* incompatibility within the highest good has its original roots in a principled and irresolvable conflict or heterogeneity between the maxims that aim at virtue and those that aim at happiness. According to Kant, the apparent conflict that commentators attempt to overcome by reframing the concept of the highest good itself (so that there is no longer a conflict) is a true and irresolvable conflict on the level of maxims, *and precisely for this reason,* requires the introduction of something *outside* of the highest good itself, namely, a divine being, to find its resolution.[26] So far, no interpretation I am aware of has recognized, let alone incorporated into a positive account of Kant's position, this deeper root of the *prima facie* incompatibility mentioned.

Another way to attempt to overcome this apparent incompatibility is to frame the concept of the highest good in a way that minimizes those elements that most obviously offend against Kant's principle that virtue be sought for its own sake. Undoubtedly, the most offensive such doctrine is Kant's claim that within the very concept of the highest good there is contained the idea of an exact proportioning of happiness to virtue, hence the idea of rewards and punishments. This doctrine is wholly at odds with strategy 1, but is also difficult, if not impossible, to accommodate by those employing strategy 2; hence proponents of those strategies have also invariably employed 3 as well. As I argued in my discussion of problem 2 above, the issue of proportionality is not, however, so easily dismissed. Indeed, there is no feature of Kant's discussion of the highest good that is more pervasive and consistent than the view that it entails both reward and punishment and that this is what we are to think of when we think of the "exact proportioning" of

---

26 Of course, Kant refers to the highest good as happiness proportioned to virtue more strictly as the "highest derived good" and to God as the "highest original good." But that is a detail that does not touch my point here.

happiness to virtue.[27] In discussing problem 3, I went on to show that this understanding of the "exact proportioning" between the two elements of the highest good is so central to Kant's views, that removing it would undermine the very link between his moral and religious thought, leaving the latter without even a properly determined concept of a divine being.

The fourth strategy, namely, that of suggesting we must disregard what would otherwise appear to be important texts, is one employed by nearly all commentators. In discussing problem 4, my aim was not so much to argue against such a methodology generally or to suggest that it could never be justified. Rather, I wanted to draw attention to how *pervasively* and *uniquely*, and yet *casually,* this strategy has been employed with respect to the highest good. Its *pervasiveness* can be judged fact that nearly all recent papers on the topic begin by telling us they will exclude large parts of what Kant says on the topic, while those that claim to focus on the account in the second *Critique* do not in fact follow up this promissory note. The *uniqueness* of the way this strategy is employed in this case is evident in that its main function has been to avoid confrontation with texts articulating the doctrines discussed under the three other problems mentioned above. Finally, the *casualness* of this strategy – which is what I would most like to see remedied – lies in its being employed as an argument for disregarding certain other interpretations simply because they rely on texts that are supposedly less Kantian in spirit, although still featured in places where we would expect to find his most considered views. Going forward, it would be best if all interpreters of Kant's doctrine of the highest good were to address this central methodological problem explicitly and thoroughly.

References

Note: All translations of Kant are from the Cambridge Edition of the Works of Immanuel Kant. Excluding the *Critique of Pure Reason*, which is cited by the corresponding A and B-edition paginations, all references to Kant's writings in German are to the volume and page number of the *Akademie Ausgabe* of his works (abbreviated AA).

Bader, R. M. 2015. "Kant's Theory of the Highest Good." edited by Aufderheide and Bader.
Beck, L. W. 1960. *A Commentary on Kant's Critique of Practical Reason*. Chicago & London: Chicago University Press.

---

27 Kant, of course, denies that these rewards and punishments should become our motive, but this is not to deny that they are nonetheless rewards and punishments.

Baumgarten, A. 2013. *Metaphysics: A Critical Translation with Kant's Elucidations, Selected Notes and Related Materials*. Edited and translated with an introduction and notes by C. D. Fugate and J. Hymers. London: Bloomsbury Academic.

Chance, B., and L. Pasternack. 2019. "Baumgarten and Kant on Rational Theology: Deism, Theism, and the Role of Analogy." In *Kant's Lectures of Metaphysics*, edited by C. D. Fugate. Cambridge: Cambridge University Press.

Denis, L. 2005. "Autonomy and the Highest Good." *Kantian Review* 10: 33–59.

Engstrom, S. 1992. "The Concept of the Highest Good in Kant's Moral Theory." *Philosophy and Phenomenological Research* 52: 747–80.

Fugate, C. D. 2014a. "The Highest Good and Kant's Proof(s) of God's Existence." *History of Philosophy Quarterly* 31: 137–58.

Fugate, C. D. 2014b. *The Teleology of Reason: A Study of the Structure of Kant's Critical Philosophy*. Kantstudien-Ergänzungshefte 178. Berlin & New York: de Gruyter.

Kleingeld, P. 1995. "What Do the Virtuous Hope for? Re-reading Kant's Doctrine of the Highest Good." In *Proceedings of the Eighth International Kant Congress, Vol. 1, Part 1, Sections 1-2*, edited by H. Robinson in collaboration with Gordon Brittan, B. Dörflinger, et al., 91–112. Milwaukee: Marquette University Press.

Kleingeld, P. 2016. "Kant on 'Good', the Good, and the Duty to Promote the Highest Good." In *The Highest Good in Kant's Philosophy*, edited by T. Höwing, 33–49. Berlin & New York: Walter de Gruyter).

Klemme, H. F. 2010. "The Origin and Aim of Kant's Critique of Practical Reason." In *Kant's Critique of Practical Reason: A Critical Guide*, edited by A. Reath, and J. Timmermann. Cambridge: Cambridge University Press.

Marwede, F. 2018. *Das Höchste Gut in Kants Deontologischer Ethik*. Berlin & New York: de Gruyter.

O'Connell, E. 2012. "Happiness Proportioned to Virtue: Kant and the Highest Good." *Kantian Review* 17: 257–79.

O'Connell, E. 2014. "Kantian Moral Retributivism: Punishment, Suffering, and the Highest Good." *The Southern Journal of Philosophy* 52: 477–95.

Silber, J. 2012. *Kant's Ethics: The Good, Freedom, and the Will*. Berlin & New York: de Gruyter.

Watkins, E. 2010. "The Antinomy of Practical Reason: Reason, the Unconditioned and the Highest Good." In *Kant's Critique of Practical Reason: A Critical Guide*, edited by A. Reath, and J. Timmermann, 145–67. Cambridge: Cambridge University Press.

Wood, A. 1970. *Kant's Moral Religion*. Ithaca, NY: Cornell University Press.