7

The Inner Voice

Kant on Conditionality and God as Cause

Rachel Barney

Kant's account of the highest good looks like important common ground with ancient ethical systems. For he holds that the value of happiness is conditional on its being conjoined with moral virtue in the highest good; and something like this *conditionality thesis*, though framed in different terms, is a shared axiom of ancient thought. The Epicureans, for instance, see virtue as a necessary condition of or means to happiness (*eudaimonia*), since only the virtuous person can avoid self-inflicted pains and minimize unavoidable ones (*Letter to Menoeceus* 132). Plato argues repeatedly that non-moral goods depend for their value on virtue or practical wisdom: only the wise person can actually be benefited by natural goods such as health, wealth, and power (*Meno* 87c–89a, *Euthydemus* 278e–282a). The *Republic* supports conditionality from a different direction, arguing that the just person is necessarily happier than the unjust. Aristotle agrees that the good man is one for whom non-moral goods *are* good (*EE* VIII.3), and further argues that happiness *consists* in virtuous activity, though it may be enhanced by non-moral goods (*EN* I.7). The Stoics take a harder and simpler line, insisting that living virtuously and living happily are the same thing, regardless of external circumstance.[1] For all these philosophers, then, happiness, *eudaimonia*, is contingent on the possession of virtue. Moreover, one way or another, all these positions meet Adeimantus' demand to Socrates in *Republic* II, the challenge which motivates the central argument of the *Republic*. Since the poets praise justice by dwelling on its rewards, Adeimantus complains, 'no one has ever adequately described what each [of justice and injustice] itself does of its own power by its presence in the soul of the person who possesses it, even if it remains hidden from gods and humans' (366e).[2] So he challenges Socrates to do exactly that: to show 'what each itself does, because of its own powers, to someone who possesses it, that makes injustice bad and justice good' (367b). So the thesis vindicated by the ancient systems is not merely that virtue and happiness tend to coincide: rather, the one is the cause of the other by a mechanism *internal* to the human psyche, independent of external contingencies.[3]

---

[1] See the texts at Long and Sedley 1987, sections 58, 60, and 63.

[2] Trans. G. M. A. Grube, rev. C. D. C. Reeve, with revisions, in Cooper 1997.

Now Kant emphatically rejects these ancient versions of conditionality, sometimes in rather acid terms: 'it must appear strange that philosophers of both ancient and modern times have been able to find happiness in very just proportion to virtue in *this* life (in the world of sense) or at least have been able to convince themselves of it' (5:115, cf. 4:442: '[the] pretence that well-being always adjusts itself to well-doing is contradicted by experience').[4] His view is that conditionality is empirically false—so obviously false that only self-deception could get anyone to believe in it. The reference to philosophers of modern times is presumably to optimistic moralists like Shaftesbury. Regarding Plato and Aristotle, Kant has remarkably little to say and the depth of his erudition is uncertain; but his general charge against the ancients collectively is clear, namely that they did not properly recognize the fundamental difference between moral and non-moral value (5:64).[5] They failed to give moral value the independence and authority appropriate to it; but they also failed to recognize that happiness is a very different thing, and travels independently. Here the Epicureans and Stoics repeatedly figure as a pair of matching case studies (5:111–12, 5:126–7; Infield 1981, 7). The Epicureans go wrong by reducing virtue to prudence, and treating it as instrumental to happiness; the Stoics, conversely, must be reducing happiness to a kind of moralistic self-satisfaction, since they treat it as an epiphenomenon of virtue.

Interpretation of such criticisms is complicated by the fact that Kant's happiness (*Glückseligkeit*) is a very different thing from ancient *eudaimonia*—and, as Kant emphasizes, is not the complete good for a human being. So we might think that the relevant Kantian successor concept is not happiness, but the highest good, which after all is how Aristotle introduces and defines *eudaimonia* in *EN* I. And since Kant agrees that *this* must incorporate moral virtue, one might see here an opening for reconciliation: perhaps his disagreement with the ancients is just an artefact of terminological confusion.

But this conciliatory suggestion is misguided. For Kant does not just fail to use the ancient concept of *eudaimonia*; he deliberately rejects it as confused.[6] Indeed his own conception of happiness can be neatly presented as a point-by-point repudiation of the Aristotelian one. Happiness is a subjective state, *not* an objective one; it is individual and empirically determined, *not* determined by our universal human nature; it has to do with pleasure and desire satisfaction, *not* the fulfilment of our rationality (nor does it have any special connection to the 'higher' rather than the 'lower' pleasures); far from being constituted by virtuous activity, it incorporates *no* moral component (except insofar as contentment depends, in the case of virtuous agents, on the consciousness of acting well (5:116–18, cf. 5:38–9)); and it is *not* the highest human good.

---

[3] Epicurus might seem to be an exception here, since on his view the value of virtue is purely instrumental; nonetheless, it is a necessary and sufficient condition of happiness (*Letter to Menoeceus* 132, *Principle Doctrines* 17, texts 21B and 22B in Long and Sedley 1987).

[4] Cf. Kant's snarky comments on the Stoics and Epicureans in particular (5:111–12, 5:115–16, 5:126–7).

[5] Irwin patiently works through Kant's criticisms of the ancients in Irwin 1996; cf. also Engstrom 1996, which finds important common ground with Aristotle.

[6] Part of the problem is that Kant does not bother to state the ancient theories in anything resembling their own terms. Hence otherwise baffling claims to the effect that they 'realized that mere happiness could not be the one highest good' (Infield 1981, 6); and, of the Stoics, 'they really left out of the highest good the second element (personal happiness)' (5:127). Here 'happiness' is used as if it were simply synonymous with 'pleasure' or 'the satisfaction of inclination', which can only make nonsense of the ancient views.

These contrasts are so complete that they might seem to overshoot, and to end up *supporting* the conciliatory suggestion that Aristotle and Kant are simply talking about two different things. But Kant's *Glückseligkeit* and Aristotle's *eudaimonia* are clear theoretical counterparts inasmuch as they share a core functional role: each is introduced as our natural and necessary end, the hazily grasped collective object of our everyday self-seeking desires and inclinations. That is what both accounts are accounts *of*; and their differences are symptoms of real philosophical disagreement rather than a terminological mismatch. The grounds of the disagreement are in a way obvious and in a way obscure: for Kant simply takes it for granted that happiness is to be understood in individual, subjective, and hedonistic terms. In doing so, I think, he means to cast himself as very much a modern man—a realistic observer who rejects any high-minded naivety about the nature of human desire, on the same sort of brute empirical grounds as he rejects any optimistic delusions about the relation of virtue to happiness. Be that as it may, it is clear that Kant means to reject not only ancient conditionality, but the very concept of *eudaimonia* on which it depends.

And yet, notoriously, Kant in the end reaffirms conditionality on his own terms: for the highest good *is* constituted by a necessary relation between happiness and virtue. My hope in this chapter is to come to an understanding of Kant's belated rehabilitation of this ancient thesis, and to see how it is like and unlike its ancient counterparts (cf. Engstrom 1996). To do so will require working through some notoriously puzzling moves in Kant's argument, each the subject of an enormous scholarly literature; I will skate quickly past a number of interpretative controversies in order to discuss, rather selectively, the features of his view which directly or indirectly illuminate his conception of conditionality. I will begin by sketching Kant's accounts of happiness and the highest good, and then consider how *conditionality*—in its Kantian version a normative relation, specifying that happiness *ought* to track virtue—comes to be realized in actual *proportionality* between an agent's virtue and her happiness. This involves examining the relevant postulates of practical reason: why and in what sense *must* we, according to Kant, suppose the realization of proportionality to be possible? And second, why must we postulate God to bring it about? At the end, I will turn to Kant's claim that the highest good is the object of our hope; for here, I will argue, is where the radically revisionist nature of Kantian conditionality can be most clearly seen.

I The Obscure Object of Desire

It is notoriously difficult to integrate Kant's characterizations of happiness into a single, coherent conception.[7] His starting point, as I've noted, is with the idea of a collective object of our inclinations, given by nature. According to what looks like an official definition in the second *Critique*:

> Happiness is the condition of a rational being in the world, in whose whole existence everything goes according to wish and will. It thus rests on the harmony of nature with his entire end and with the essential determining ground of his will. (5:125)

As such, happiness is naturally the object of our strongest and deepest inclination (4:399). It is characteristic of *rational* beings precisely because it is an idea, over and above the objects of our particular inclinations (5:430). By the same token our desire for it is a function of self-love rather than our merely animal selves. But it is nonetheless a given, its pursuit part of our natural endowment (6:7 n., 6:387–8).

As the sum of our inclinations (4:399), happiness consists, it seems, in the satisfaction of desire: it is what we get when we get the other, more determinate things we want (cf. A806/B834). But in some contexts Kant readily substitutes talk of pleasure—particularly when emphasizing the individual, unpredictable character of both: 'Where one places his happiness is a question of the particular feeling of pleasure or displeasure in each man' (5:25). This substitution is problematic, of course, since only on the basis of some very implausible stipulations can pleasure and the satisfaction of desire be identified. But Kant seems to use the two interchangeably, perhaps even to deliberately conflate them: I will suggest a bit later what his point in doing so might be.

Whether as pleasure or satisfaction, happiness is, Kant emphasizes, inherently subjective, and thus individual and unpredictable. It can only be empirically determined—a crucial result because happiness is thus formally disqualified as a source of laws for the will (5:25–6). Thus prudence, the virtue of practical reason in its capacity as the servant of self-interest, operates without laws (5:25–6, 5:36, cf. 4:418, 5:430). Further, the desire for happiness has its roots in our perpetually *un*satisfied natures as creatures of needs (5:25). To be human is to forever generate new desires, so that there is never any point at which we can relax in satiation and treat happiness as an *acquis* (5:430). And yet permanent and complete attainment is part of its very concept: 'all the elements which belong to the concept of happiness are without exception empirical—that is, they must be borrowed from experience; but . . . none the less there is required for the Idea of happiness an absolute whole, a maximum of well-being in my present, and in every future, state' (4:418).

---

[7] Cf. (in addition to the more general studies cited in the References) Wike 1987 and 1994, O'Connor 1982, Hill 1999, Watson 1984, and especially Reath 1989 for different ways of making sense of Kant's description of happiness, and debate as to which conception is most central and/or defensible.

So there is something paradoxical, phantasmagoric—and essentially unrealizable—about happiness. Given that our inclinations are boundless, conflicting, ever-changing, and only imperfectly knowable (4:418), the perpetual complete satisfaction of *all* of them is not a genuine possibility: but that is what happiness is. This confirms the strong suggestion of infantile fantasy in the definition of happiness cited earlier, as the condition of a being for whom 'everything goes according to wish and will'. It is only in fantasies and fairy tales that *everything* goes according to my wishes; and commands to elves and genies always backfire in the end. In real life, happiness must outrun what my will can determine: it depends on 'how things turn out', on unpredictable factors and forces beyond my control—hence the 'hap' in happiness, the *Glück* in *Glückseligkeit* and probably the *daimon* in *eudaimonia*.

This phantasmagoric quality tells us something important, I think, about the *status* of happiness so conceived (that is, happiness as total pleasure and/or satisfaction). It is not only an 'ideal of imagination' (4:418), as Kant says; it is, so to speak, not a concept for which he takes responsibility. It is *our* concept, not his, which he is trying to capture—the fuzzy, incoherent fantasy-end we hazily envisage when we lunge after the objects of our inclinations. This would be why Kant has no interest in trying to define it more precisely as pleasure or desire-satisfaction, or some coherent amalgam of the two.[8] It is perhaps best seen as an impossible combination of *two* fantasies: as a condition in which the perfect satisfaction of my will results in total and perpetual pleasure. The important thing is that we recognize our own self-interested strivings in some depiction along these lines.

But this cannot be the whole story. For the desire for happiness is both natural and necessary to us, as Kant emphasizes (5:25). For us to renounce it is impossible. And practical reason must be able to work within the limits of the possible—as we will see again when Kant turns to argue that we must affirm the possibility of the highest good. So there must be something else that happiness could be: the fantasy must have a realizable counterpart, one which inherits its standing as a natural end but can be endorsed and aimed at by practical reason. This is what Kant alludes to when he qualifies a reference to 'all inclinations taken together' with: 'which can be brought into a fairly tolerable system, whereupon their satisfaction is called happiness' (5:73). Here and at other points Kant speaks of happiness as something systematized by reason out of the raw materials of our inclinations—a maximal set of jointly realizable inclinations, perhaps, or a life plan more deeply shaped by prudence, self-knowledge, and experience. It is happiness so understood which is pursued by self-love under the guidance of prudence (4:416 n., 6:46 n.). This happiness too is an

---

[8] Alternatively, as Sergio Tenenbaum has suggested to me, we might take Kant to assume that it is always a representation of some object as pleasant which makes it the object of my inclination, so that for 'everything to go according to my wish' *is* for me to be pleased: pleasure and desire-satisfaction coincide except in cases of error. This would not settle the question, though, of which is constitutive of happiness, or what to say about cases in which the two come apart. Cf. Reath 1989 for a subtle reading of the role of pleasure in Kantian moral psychology.

'Idea', and might be unattainable in its perfect form, as is perfect virtue in this life, but we can still rationally strive to maximize it.

When Kant speaks of happiness as a component of the highest good, it is this happiness-as-rational-system which he has in mind. Kant at this point converges with a more traditional, objectivist understanding of happiness as the real long-term welfare—as Kant puts it, the 'lasting advantage' (4:416 n.)—of an agent. But there is no retreat here from Kant's understanding of happiness as consisting in pleasure or subjective satisfactions: rather, he is making the further claim that a rationally endorsable, enduring system of such satisfactions is constitutive of an individual's good.

Kant is not as explicit as we might like about the precise relation between happiness as fantasy and happiness as system. But it clearly would be wrong to think of these as theoretical competitors for the title 'happiness'. Nor, on the other hand, are they just different 'senses' of the word, harmlessly distinct once disambiguated. (Kant certainly shows no interest in any such disambiguation.) Rather, these ways of thinking about happiness are complementary, differentiated by their relation to practical reason. The systematic end is the imaginative ideal brought into focus: it replaces the fantasy in guiding our actions insofar as our practical reason is developed.

II Conditionality in the Highest Good

Once we have a grasp of happiness as system, we can be led to see that it is incomplete as an end of practical reason, and can only be valued in conjunction with moral virtue. Kant appeals to our moral responses here: 'a rational and impartial spectator can never feel approval in contemplating the uninterrupted prosperity of a being graced by no touch of a pure and good will, and . . . consequently a good will seems to constitute the indispensable condition of our very worthiness to be happy' (4:393). The pleasure and pain (the *Wohl* and *Weh*) of moral agents only has objective value and disvalue in proportion to their moral condition:

> When . . . someone who delights in annoying and vexing peace-loving folk receives at last a right good beating, it is certainly an ill, but everyone approves of it and considers it as good in itself even if nothing further results from it; nay, even he who gets the beating must acknowledge, in his reason, that justice has been done to him, because he sees the proportion between welfare and well-doing, which reason inevitably holds before him, here put into practice. (5:61)

As this shows, 'happiness, though something always pleasant to him who possesses it, is not of itself absolutely good in every respect but always presupposes conduct in accordance with the moral law as its condition' (5:111). Conversely, virtue, though good absolutely, is *not* 'the entire and perfect good as the object of the faculty of desire of rational finite beings. For this, happiness is also required', and that from an impartial and not merely a subjective, self-interested point of view (5:110).[9] These arguments converge from both sides to show that 'happiness, though it indeed constitutes the second element of the highest good, does so only as the morally conditioned but necessary consequence of the former' (5:119). Thus constructed by this asymmetrical conditional relation, the highest good emerges as the final end of pure practical reason.

Kant's arguments appeal both to the intuitions of an impartial observer and to the principle that the highest good must be characterized by *sufficiency*, leaving out no part of what we may rationally desire. But as several scholars have brought out, the highest good as end can also be seen as a necessary construction from the moral law.[10] To be governed by the moral law is to adopt the standpoint of virtue; and only to the extent that we adopt that standpoint are we rationally entitled to pursue our own happiness as end. Thus the moral law gives back to us, in rationalized form, the end which it had seemed to take away. And so virtue is necessary for happiness to be genuinely good—i.e. for it to be the object of my rational will. As Stephen Engstrom sums up: 'Kant's idea that the universalized claim is necessary and sufficient to validate self-love's claim that personal happiness is good amounts to the idea that virtue is a necessary and sufficient condition of the goodness of happiness' (1992, 762).

At this point we need to be careful not to be misled by a superficial resemblance to certain ancient ideas. Anyone familiar with ancient versions of conditionality might be tempted to take this point about 'the goodness of happiness' as amounting to something like Aristotle's claim that natural goods such as wealth or health *are* good only for the good man (*EE* VIII.3). But there can be no real analogy here. For Kant is talking about happiness itself: our natural self-interested end, not the raw materials whose acquisition and use might tend to bring it about (the 'natural goods' or 'goods of fortune' of the ancients), and which therefore might or might not be 'good for us'. In fact it is not clear that Kant has, or could have, the Platonic-Aristotelian category of 'good for *x*' as opposed to good *simpliciter*: in any case he does not invoke it here, either by saying that the villain's happiness is good for him but not good *simpliciter*, or by saying that it is not even good for him. Instead he insists only that it is not good absolutely in all respects and without qualification (5:60, cf. 5:62). The villain's pleasures and satisfactions are, it seems, no less constitutive of his happiness than the good person's: they just cannot be rationally willed by anyone, even himself.

---

[9] A fuller account (complete with a kind of 'original position') comes in the *Religion* (6:5–6). Kant's sufficiency arguments proceed by the same kind of thought experiment as we find in Plato's *Philebus* (20c–21a) and Aristotle's *Nicomachean Ethics* (X.6–7). To assess a candidate for the highest good, imagine a life which enjoys it to the fullest extent, but is stripped of any *other* conceivable good; and ask whether that life would be complete and sufficient in its desirability. Engstrom goes so far as to say that 'the consummate good is the self-sufficient good, and thus Aristotle's and Kant's formal characterizations of the highest good are fundamentally the same' (1996, 122).

[10] See the very different constructions of Engstrom 1992, Wood 1970, 55–60, Reath 1988, 598–9, and Guyer 2000, 339–45.

A third difference will emerge as crucial later on. This is that Kantian conditionality, unlike the ancient species, is an 'internal' relation only insofar as it is normative. That is, it follows from the nature of happiness and virtue that the former *ought* to track the latter—or, in other words, that it is good only to the extent that it does so. But the realization of conditionality in the world, in the form of happiness proportional to the virtue of its subject, is not caused by any internal or direct causal relation between the two—as it is in the ancient systems, in which virtue is psychologically constitutive of happiness or at least a necessary condition for it. Rather, the relation is brought about only externally, by a rational will determined to make it so.[11] (I will have more to say about this in section IV.)

Kant's account of the highest good is complicated by two ambiguities. One is between the highest good as a social or secular and as a divine achievement (cf. Reath 1988). The former can in turn be seen in either of two ways: we can either hope for future human progress towards the highest good or, somewhat mysteriously, we can view our actions now in relation to 'a *corpus mysticum* of the rational beings' in society (A808/B836). Either way, it is a human achievement: rational beings 'would themselves be the authors both of their own enduring well-being and of that of others' (A809/B837). But Kant himself seems to treat this conception as incomplete or preliminary, using it to prime us for the 'theological' one.[12] What aids the transition is that, even in presenting the secular conception, Kant stresses that the actions of this community of rational agents are 'just as if they had proceeded from a supreme will' (A810/B838). That is, it is only *insofar* as a community of rational wills approaches the condition of a unified divine will that it begins to produce the highest good (cf. 6:97–8).

Second, Kant oscillates between two somewhat different formulations of the highest good. For Kant sometimes speaks of the highest good in terms of *perfect* virtue and the *perfect* happiness appropriate to it, either in the individual or in 'the most perfect world'—the 'maximal' conception, as Beck calls it (1960, 268ff.)—and sometimes as the conjunction of *any* degree of virtue (or the lack of it) with the *proportionate* degree of happiness. There is no conflict between the two, insofar as maximization is a special case—the ideal or optimal case—of the proportional principle. The difference is presumably a matter of the emphasis appropriate to different contexts. Maximization is salient when we consider our own defective progress towards virtue: it grounds the postulation of immortality in an eternal hereafter, in which we may advance towards perfectability without end (5:122–3). Proportionality in general is salient to the question of whether the conditionality relation which structures the highest good can be realized at all, and by what cause: it grounds the postulation of a divine agent (5:124–5, A810–13/B838–41). Since the latter formulation brings Kantian conditionality into sharper relief, it will be my focus for the rest of this chapter.

---

[11] For the purposes of this chapter I for the most part bracket the question of whether we should think of the highest good as being realized in this world, the afterlife, or both—a puzzle which is intertwined with the distinction between the social or secular and the theological conception of the highest good, the unclear division of labour between human and divine agency, and the distinction between the phenomenal and noumenal realms. As with the secular/theological distinction itself, Kant is remarkably casual on this question: e.g. 'either in this or in another life' (A812–13/B840–1). What is important is the negative point that the proportionality essential to the highest good is *not* to be found in the empirical present: where we *do* envisage it—coming into existence in a future society, or hidden in a noumenal *corpus mysticum*, or postponed to the infinite afterlife—is apparently a matter of detail, irrelevant to the value of the attitude we are required to take to it (namely, hope).

[12] As Reath notes, in both the first and the second *Critique* Kant *shifts* from the secular to a theological conception: he 'begins by viewing the Highest Good as an end to be promoted in this world. But when he imagines its realization, he conceives of it as a state of affairs in another world that would result from divine agency' (1988, 608).

Kant's insistence on the necessary proportionality of the highest good is fraught with philosophical difficulties, and a number of interpreters have been inclined to marginalize it or read it in deflationary or non-literal terms.[13] Some of the objections to it are epistemic: how can we promote the highest good if this requires us to judge the virtue of others, which depends on motivations to which even they have only unreliable access (4:406–7)? Kant might respond that at most this makes the pursuit of proportionality difficult rather than impossible. It is, after all, in the interests of proportionality that one gives the vexatious person a right good beating, and Kant exhibits no deep qualms about such everyday exercises in just deserts. Of course, virtue is not a fixed quantity, and one might object that we should be trying to increase the vexatious person's moral worth rather than trying to reduce his happiness to fit our assessment of his current state.[14] But there are epistemic and practical difficulties here as well: and Kant *is* deeply sceptical of the idea that the virtue of others could be part of my ethical task (6:386).

Still, it might seem ethically repellant to suppose that for us to *reduce* the happiness of others could ever be a requirement of Kantian practical reason. Intuitively, one imagines the agent deeply concerned with proportionality as unpleasantly moralistic and self-righteous, and very likely self-deceived or self-serving. (As Nietzsche said, 'Mistrust all in whom the impulse to punish is powerful', *Also Sprach Zarathustra* II.29.) And in Kantian terms it seems to violate the injunction that I should promote the ends of others, as part of treating them as ends in themselves—even granted that there must be limits to how far I should actively collaborate in misguided ends

---

[13] See e.g. Reath 1988, 611–13, Beck 1960, 268ff., Engstrom 1992, 768–73, Friedman 2008.

[14] Deterrence and moral education might even be added benefits of the beating. But Kant is clear that it elicits our approval not for these reasons, but simply as an exercise in proportionality: 'everyone approves of it and considers it as good in itself *even if nothing further results from it*; nay, even he who gets the beating must acknowledge, in his reason, that justice has been done to him, *because he sees the proportion between welfare and well-doing*...' (5:61, my emphases). As Sergio Tenenbaum points out to me, the case confirms that, in order to satisfy the norm of conditionality, proportional outcomes require causal connection as well as coincidence. It would hardly be equally satisfactory if the vexatious person suffered from an equally painful toothache—or so it seems to me, at any rate, though intuitive reactions here seem to vary.

(6:388).[15] To limit anyone's attainment of licit ends in accordance with desert is surely, as Lewis White Beck memorably puts it, 'the task of a moral governor of the universe, not of a laborer in the vineyard' (1960, 245). Moreover this could hardly be intended as a new moral requirement; for Kant himself says that the highest good is not the object of a *further* duty. As he puts it, in a slightly mysterious formulation, 'harmonizing with this end does not increase the number of morality's virtues but rather provides these with a special point of reference for the unification of all ends' (6:5).

However, it does not follow that proportionality can be *none* of our ethical business. No doubt, given the epistemic difficulties and the ethical pitfalls, it will rarely fall to us to deliberately curtail the happiness of the bad. For us to do so, however, is no violation of the moral law, or even an addition to it, given the construction of the highest good from the moral law: for the happiness of an agent is only in fact good, and thus a rational end, in proportion to his virtue. Kant is emphatic that we must recognize this even in the first-personal case: 'Reason does not approve happiness (however inclination may desire it) except insofar as it is united with worthiness to be happy, that is, with moral conduct' (A813/B841). Our intuitive recoil from the prospect of administering proportionality is natural and right, given the universality of human frailty and imperfection: as Hamlet says, 'Use every man after his desert, and who should 'scape whipping?' (II.ii). But it is equally repulsive to suppose that, even in the most extreme cases, we have a duty to help the deeply wicked live happily ever after—as we would were it not for the limitations imposed by the norm of conditionality.

What these shifting, uneasy intuitions help to make clear is that the construction of the highest good from the moral law, which involves a reconception of virtue as the *worthiness* to be happy, is also the construction of something not previously visible in the Kantian system: namely the moral norm of *justice*, fairness, or desert.[16] As Wood puts it, 'Pursuit of the second component of the highest good is, in effect, beneficence limited by justice' (1992, 402). What Kant's preference for describing the highest good in terms of proportionality brings out is his intense if otherwise under-articulated commitment to this norm—to justice or fairness *tout court* we might say, quite apart from any specialized juridical or political concept—a commitment which he takes all of us to share. For Kant it is a central feature of the moral life that gross failures of proportionality are all around us, and that this is naturally felt to be intolerable. In elaborating on the 'moral proof' of God in the third *Critique*, he says:

---

[15] As Ralf Bader has pointed out to me, here and in the next paragraph I gloss over a number of problems about how proportionality, particularly as presented at 5:61, is to be reconciled with Kant's duty of beneficence and its derivation from the moral law. See the careful discussions of Engstrom 1992, 768–73 and Friedman 2008, 327–30.

[16] My talk of justice as the Kantian norm achieved by proportionality has, admittedly, no real counterpart in Kant's own official formulations of the highest good. But I cannot think of any more suitable name for what our 'inner voice' demands and divine proportionality achieves. And Kant does speak of the villain as acknowledging the *justice* of his beating. Perhaps he avoids thematizing proportionality in these terms in order to reserve 'justice' for the very different concept elaborated in his political works: I will have nothing to say about how these concepts might be related.

> This moral proof by no means [offers] a newly discovered basis for proving [the existence of God], but at most a new elucidation of that basis . . . Once people began to reflect on right and wrong . . . they inevitably had to arrive at this judgment: that in the end it must make a difference whether a person has acted honestly or deceitfully, fairly or violently, even if to the end of his life he has received no good fortune for his virtues and no punishment for his crimes, at least none that we could see. It is as if they heard an inner voice that said: *This is not how it should be.* (5:458)[17]

Or as Mark Johnston puts it (speaking of Kant, but in the context of motivating his own inquiry into the prospects for an afterlife):

> When faced with the contrast between the professional torturer who dies calmly in his sleep at a ripe old age surrounded by his adoring family, and the nurse who, for her whole adult life, cared for the dying only to herself die young and alone from a horribly painful and degrading illness, people do tend to fall into despair over the importance of goodness. Unless, that is, they have hope or faith. (Johnston 2010, 11)

Kant's postulation of real proportionality, and a God who effects it, is the rationalized expression of this inner refusal to accept our experience of the world, with its monstrous injustices, as the totality of what there is. To say that the achievement of the highest good must be possible is to identify with this inner voice, and acknowledge it as the voice of reason. The next step is to see how the demands of the inner voice are to be met.

III The Highest Good and Divine Agency

Kant takes the rational requirement to pursue the highest good to imply a *possibility requirement*: in order to adopt the highest good as the object of our will, we must hold its realization (and proportionality in particular) to be possible. As Kant puts it, 'pure practical reason must necessarily think [the highest good] as possible because reason commands us to contribute everything possible to its realization' (5:119). This possibility requirement in turn leads (quickly if somewhat mysteriously) to a further requirement, in relation to which proportionality is again emphasized.[18] Since the world of experience reveals no coincidence between virtue and happiness, I must postulate a further realm and a cause which *is* able to realize it—namely God.[19] This pair of steps are summed up as follows:

---

[17] A complication I here set aside (but which strikes me as important) is that in discussing the Spinozist, Kant presents the empirical realm available to him as not only unjust but *meaningless*: not only do the righteous suffer deprivation, disease, and death like anyone else, 'they will stay subjected to these evils always, until one vast tomb engulfs them one and all (honest or not, that makes no difference here) and hurls them, who managed to believe they were the final purpose of creation, back into the abyss of the purposeless chaos of matter from which they were taken' (5:452).

[18] As well as the requirement that we postulate the immortality of the soul, which is driven by the 'maximization' version of the highest good (5:122).

> The same law must also lead us to affirm the possibility of the second element of the highest good, i.e., happiness proportional to that morality; it must do so just as disinterestedly as heretofore, by a purely impartial reason. This it can do on the supposition of the existence of a cause adequate to this effect, i.e., it must postulate the existence of God as necessarily belonging to the possibility of the highest good (the object of our will which is necessarily connected with the moral legislation of pure reason). (5:124)[20]

Note Kant's emphasis on proportionality in this context, repeated at 5:125: 'the existence is postulated of a cause of the whole of nature, itself distinct from nature, which contains the ground of the exact coincidence of happiness with morality'.

The great puzzle here, of course, is what exactly is supposed to generate this pair of postulates. At times it seems that Kant is relying on what we might call the *argument from frailty*: we cannot be steadfast in our commitment to the moral law without some hope that the highest good will be realized. This argument amounts to an empirical psychological prediction: we will lack the fortitude to steadfastly do the right thing if we are convinced that the injustice and moral chaos we see around us is all there is. Practical reason demands that we take the steps necessary to prevent ourselves from lapsing from the moral law: hence it demands that we postulate, first, the possibility of the highest good; and second, in order to secure that possibility, the existence of God.

---

[19] I here mean to largely bypass certain questions about the *status* of the postulation of God (see Kuehn 1985 for discussion, in terms of a helpful sorting into 'traditionalist' and 'modernist' or 'existentialist' interpretations, and his own 'deductive' alternative: cf. also Scutt 2010.) I take it that Kant's argument is not intended simply as rhetorical, aimed at persuading us that God exists. But neither does it straightforwardly prove *that God exists*—a truth, if it is one, belonging to speculative and theoretical reason. What it aims to prove is rather the more complex proposition *that it is rational for us to postulate that God exists*, which, rather problematically, cannot be reduced to, and does not even entail, the proposition that he does. It is this irreducibility which I take Kant to be granting when he emphasizes the first-personal and subjective (not to be confused with merely personal and optional), and above all strictly practical character of the postulate (e.g. 5:124–5). At times he even phrases this postulation as a kind of *performance* of commitment: 'I will that there be a God…I stand by this and will not give up this belief' (5:143).

[20] An oddity of the argument which has received relatively little discussion is that Kant here says only that we must hold the highest good to be *possible*; but to do so evidently requires supposing that God is *actual*. This asymmetry is puzzling: as Wood notes, 'if God's existence is both necessary and sufficient for the actuality of the highest good, then belief in the highest good would seem equivalent to the belief that God is possible' (1992, 405). As Wood notes, Kant does occasionally retreat to this weaker position. But his standard formulation is unambiguously stronger: 'Therefore, it is morally necessary to assume the existence of God' (5:125). However, it is not hard to make an intuitive case for this stronger stance. First, when we are considering practical possibilities (the possibility, for instance, that some action of mine will be successful), we find it natural to describe outcome *z* as possible only if its preconditions *x* and *y* are actual. (For instance: it is possible for me to fly, if I have [actually] wings and the bone structure of a bird.) And God as divine cause is perhaps best understood as a precondition for my successful participation in bringing about the highest good, as sketched above. Second, if God does *not* exist, the right kind of connection between virtue and happiness can never obtain: the highest good turns out to be outright *impossible*. For a careful reconstruction of Kant's reasoning here, see Guyer 2000.

Several of Kant's expressions of the possibility requirement are very much in this spirit (e.g. 6:5). This reading also dovetails with Kant's suggestions that the worry to which the possibility requirement responds is first-personal: what I must believe possible is that *my* commitment to the moral law will be matched by *my* happiness.[21] Still, the argument from frailty cannot be the whole story. To claim that we *cannot* be steadfast in our moral commitments without hope would be to elevate an empirical psychological generalization to the status of a necessary causal truth, a kind of move which Kant regularly denounces as improper in moral philosophy (e.g. 5:26). The argument from frailty cannot make the possibility requirement more than *nearly* necessary, as a matter of how human motivation typically works; and the chasm between the nearly necessary and the genuinely so is absolute. There remains the possibility of what we might call Spinozist heroism (or, equally well, 'existentialist heroism' or 'film noir heroism'): a commitment to the moral law which remains steadfast (i.e. reliably overrides any conflicting self-interested motivations) despite a complete absence of any hope for the highest good. And in fact Kant's discussion of the Spinozist seems to consider the position of just such an agent:

> Therefore, let us consider the case of a righteous man (Spinoza, for example) who actively reveres the moral law [but] who remains firmly persuaded that there is no God and . . . that there is also no future life: How will he judge his own inner destination to a purpose, [imposed] by the moral law? He does not require that complying with that law should bring him an advantage, either in this world or in another; rather, he is unselfish and wants only to bring about the good to which that sacred law directs all his forces. Yet his effort [encounters] limits: For while he can expect that nature will now and then cooperate with the purpose of his that he feels so obligated and impelled to achieve, he can never expect nature to harmonise with it in a way governed by laws and permanent rules (such as his inner maxims are and must be). (5:452)

Kant presents the Spinozist as confronted with a dilemma: either lapse into disrespect for the moral law, or accept the possibility of divine agency. Since the former is *ex hypothesi* unacceptable to him, we are to see him as perpetually teetering on the brink of acceptance of the postulates:

> And so this well-meaning person would indeed have to give up as impossible the purpose that the moral laws obligated him to have before his eyes, and that in compliance with them he did have before his eyes. Alternatively, suppose that, regarding this [purpose] too, he wants to continue to adhere to the call of his inner moral vocation, and that he does not want his respect for the moral law . . . to be weakened, as would result from the nullity of the one ideal final purpose that is adequate to this respect's high demand (such weakening of his respect would inevitably impair his moral attitude): In that case he must—from a practical point of view, i.e., so that he

---

[21] E.g. 'Not being nature's cause, his will cannot by its own strength bring nature, as it touches on *his* happiness, into complete harmony with his practical principles' (5:124, my emphasis). Cf. 5:129–30, A813/B841.

can at least form a concept of the possibility of [achieving] the final purpose that is morally prescribed to him—assume the existence of a *moral* author of the world, i.e., the existence of a God; and he can indeed make this assumption, since it is at least not intrinsically contradictory. (5:452–3)

If the problem with the Spinozist is not simply that he is psychologically vulnerable to lapsing from the moral law, what exactly *is* wrong with him? Since Kant presents the possibility requirement as a demand made by practical reason, the question is in what way the Spinozist is being irrational. Zev Friedman has argued that the postulates are rationally required for the legitimate authority or objectivity of the moral law, which Spinoza accepts. For 'if the fate of the morally worthy and the morally unworthy does not reflect their relative worthiness . . . the moral law cannot be understood to be objective. Rather, the law must be discarded as an empty projection of the agent's mind' (2008, 326)—a subjective delusion, answering to nothing in the way the world is, like the inner narrative of someone who believes himself to be Napoleon. 'Unless the natural order is understood to be under the hegemony of the moral order the moral law cannot claim objective status' (2008, 342).

Kant does seem to flirt with this argument from the legitimacy of the moral law, as when he says that the impossibility of the highest good would render the moral law 'fantastic, directed to empty imaginary ends, and consequently inherently false' (5:114).[22] For the most part, though, Kant is hostile to any suggestion that the moral law could depend for its authority on anything else at all—including the postulation of God (5:125–6).

I would suggest that both the argument from frailty and the argument from the legitimacy of the moral law are, for Kant, sidelines. Neither is his principal argument—or even the right *kind* of argument—for the possibility requirement.[23] Instead, there is *no* argument: Kant simply asserts the possibility requirement as obvious. Why might it seem so? Probably because it is natural to suppose that one cannot genuinely intend the impossible (*de dicto*).[24] If I believe that X is *impossible*, then X cannot quite be the object of my action, which we need to redescribe in terms of some other end in the vicinity. I think that this is very likely a defensible principle, and a useful constraint on our ascriptions of intention.[25] It seems reasonable, for instance, that someone who votes for a fringe party without having any hope of their victory should be described not as intending to elect them and enact their policies (or to join others in doing so), but rather as sending a certain message, registering a protest, and so on. And even someone fanatical about making healthy choices should not be described as doing so with the *intention* of living forever—unless we want to claim that she is mad—but only as hoping to remain healthy for as long as possible. So if we may take it that Kantian willing is or incorporates intending (I will return later to this assumption), for me to will the highest good without believing it to be realizable is not a genuine possibility. As Kant puts it: 'But the subjective effect of this law, i.e., the intention which is suitable to this law and which is necessary because of it, the intention to promote the practically possible highest good, at least presupposes that the latter is possible. Otherwise it would be practically impossible to strive for the object of a concept which, at bottom, would be empty and without an object' (5:143).

---

[22] Likewise a startling passage from the *Lectures on Ethics*: 'Morality implies a natural promise: otherwise it could not impose any obligation upon us. We owe obedience only to those who can protect us. Morality alone cannot protect us' (Infield 1981, 82; cf. A811–14/B839–42).

[23] What *are* they then? Responses to objections, I would suggest. The argument from frailty turns to Kant's own purposes the protest that the moral law demands an unrealistic degree of unselfishness from us. The argument from the legitimacy of the moral law does the same for the objection that it is unfair to demand that we subject ourselves to unenforced laws. Here as elsewhere (see section V below), Kant's reasoning needs to be read in its dialectical context. Addressed to someone not already committed to the moral law, the possibility requirement could only serve as a corrupting incentive. Addressed to a confirmed believer in Christian theodicy, it is unnecessary. Kant's reasoning is addressed to an interlocutor in roughly the unstable position of the Spinozist, who, while denying the rationality of faith, might well *also* feel threatened with despair or tempted by a Job-like protest against the injustices around us, and so might be responsive to these incentives to the will. (For the general point that his concepts and theses must be grasped in a certain sequence, cf. 5:64; for the idea that one incentive might be legitimately deployed only to balance another, cf. 5:88; for Kant's reading of Job, cf. 'On the Failure of All Philosophical Theodicies', in Despland 1973.)

Thus belief in the possibility of the highest good turns out to be a demand of practical reason because it is a necessary prerequisite for states of the will which are themselves demanded by practical reason—necessary as a matter of the logic of intention, not just nearly necessary given common human motivational weaknesses. We can now see, I think, the nature of the Spinozist's failing. It is not that he is (as per the argument from frailty) doomed to discover at some point that he is too weak to persist in his commitment to the moral law. Nor is it (just) that he is being irrational, refusing to respect the entailment relations among his commitments. Rather, his will is at an impasse: he cannot *really* will the highest good at all, since he considers it unavailable; yet he recognizes that practical reason requires him to do so. He tries to fall back on his commitment to the moral law, suppressing all selfish inclinations; but it is the moral law itself which in turn generates the requirement to pursue the highest good, which he cannot do. Until he assents to the postulates, his reasoning can only run round in a doomed circle, like someone whose most careful calculations keep giving an obviously wrong result.

---

[24] Contemporary philosophers debate whether intending to ϕ requires believing that one will succeed in ϕ-ing: that I must at least consider it a 'live possibility' that I will succeed seems to be uncontroversial. (See Holton 2008, with references.)

[25] Cf. Wood on the argument as *ad absurdum practicum* (1970, 27–30) (though Kant's explanation of it in those terms is, to me, *per obscurior*). Wood also raises the question whether we should read Kant as holding that it is impossible to intend what is believed to be impossible, or only that it is *irrational* (1970, 21–3, cf. 1992, 401); I here opt for the stronger reading. One might object that this is too strong, since the position of the Spinozist seems to be presented as possible, and Kant consistently emphasizes that the postulates are *rational* requirements. However, the distinction is not so clear-cut as we might assume: a stance which is irrational may be impossible to hold transparently, and it is hard to say how transparent the Spinozist's predicament is to him.

On this reading, however, a major difficulty looms: if the possibility requirement is grounded on this principle about the availability of intentions, it does not obviously support the further postulation of God—indeed the two may not even be compatible. For if I must think that the end of my action is possible, this is surely because, more precisely, I must think that the success *of my action* is possible. Suppose that the question is whether I can intend 'flying to the top of the Empire State Building'. If Kant's claim is that I can do so only if I believe that 'flying to the top of the Empire State Building' is possible, surely this is elliptical for *my* flying, the very flying I purport to intend. Whether birds can do it is irrelevant to what *I* can intend. In the same way, that God can bring about a necessary proportional relation between virtue and happiness does nothing to show that I can. Worse, if his actions are sufficient for that end, then my own actions and intentions seem to be otiose at best. Whether and how God's agency leaves any room for my own to be efficacious at all is a sticky question, to which I will return in the next section: at a minimum, we might object that it renders my own agency *unnecessary*, while doing nothing to secure its possibility. The moral agent turns out to be like a child who is encouraged to think he is steering the shopping cart, making the soup, etc. while really the parent does everything.

This problem of the division of labour between the human moral agent and God is bound up with a more general interpretative dilemma. Our own actions, after all, can only aim at bringing about the highest good in a local and imperfect way. So what is it, exactly, that we must believe possible: is it merely that at least some of our actions aimed at that end will meet with success? Or must we believe in the possibility of the *perfect*, total realization of the highest good, which only God or some perfected community of rational human wills can provide?[26] Only the former seems to follow from the principle that I cannot intend what I consider impossible; but only the latter seems to require the postulation of God. The problem can be put as an objection from the Spinozist. Why is it not sufficient for me to believe that the highest good (and more specifically, proportionality) can be realized on some occasions by my actions—that happiness and virtue can be brought to coincide occasionally, that good guys don't *always* finish last? Even the film noir hero hopes for this much; and it is all he needs in order to act.

To answer this, Kant needs a principle to the effect that we can believe that the highest good is possible even locally and imperfectly *only if* we believe that it can be perfectly realized. Now I think that Kant does in fact accept such a principle, and has some defensible (up to a point) reasons for doing so. To bring these out, we need to look more closely at the relation of divine and human agency, and at what Kant means by presenting the highest good as the object of our hope.

---

[26] This dilemma is not resolvable into the distinction between the 'maximal' and 'proportional' formulations of the highest good: the problem is specifically whether I can make *proportionality* my end while considering it possible only locally.

IV Hope and Justice

The postulation of God does much more than assure us that, in the long run, virtue and happiness will coincide.[27] It attributes that coincidence to an infallibly reliable cause,[28] one rationally guided by the norm of conditionality. Thus the relation between virtue and happiness turns out to be a necessary and a causal one—not one of 'internal' or immediate cause and effect, however, but one mediated by the divine rational will. This necessary causal relation, Kant emphasizes, is itself part of what we will when we make the highest good our object (5:111, 5:113–15). I will my happiness not only proportionally, *insofar* as I am worthy, but as *an effect* of that worth—and not, say, the other way around.

This kind of causally structured object of the will is familiar enough, as are cases where the causality must be mediated by the will of another. Think of a student, Annie, who works hard on a paper. Annie wants two things: to write a good paper and to receive a good grade. If she writes a good paper and gets a C, she will be pleased in one respect and disappointed in another. If she writes a bad paper and gets an A, she will again be partly pleased but also disappointed—disappointed in herself and perhaps in her professor as well. Now imagine that she writes a good paper and receives an A; but she then learns that the paper was never read by Professor Careless, who assigns grades by a roll of the dice. In this case, both of Annie's desires, as initially specified, are satisfied. Yet surely this outcome will be at least as insufficient and frustrating as the other two. So something was missing from the initial characterization: Annie did not want two things, to write a good paper and to get a good grade, but a single complex thing: to get a good grade *as a result of* having written a good paper. The conditionality specified by the causal link is itself a part of the object of desire.

What we want in cases like this is to be rewarded for acting well: what we want, in short, is *justice*. And justice is not done if mere coincidence happens to bring about a proportional result. Even in the movies, we do not want the villain to suddenly die from an allergic reaction or a subway accident unless we can somehow interpret it as *telling*, and a form of cosmic justice: his death must somehow be seen to result from his own wicked actions and character. Moreover, Kant may reasonably insist that the causal link demanded here is a necessary and universal one. Like the assassin who shouts, 'Sic semper tyrannis!', what I will in willing the highest good outruns anything I can myself perform, for I will my actions *as* applications of a rule.[29]

---

[27] Cf. Engstrom: 'the highest good is distinct not only from any condition in which happiness is not proportional to virtue, but also from any condition in which the proportionality is fortuitous' (1992, 750). Cf. Reath 1988, 598.

[28] Despite Kant's avowed circumspection about entering into traditional speculative theology, it turns out that many of God's standard predicates can be deduced from the role assigned to him by practical reason (5:140).

[29] This is arguably a feature of all rational action according to Kant, since we cannot help but act on maxims which are general in form. A fuller discussion of the issues here would have to say more about the

This requirement that the link between virtue and happiness be a matter of necessary and universal causality is already enough to require the postulation of a superhuman agency; and it also plausibly belongs to the content of the highest good that this causal link be *rational*. We find it natural to think of justice as something *done* or administered, rather than as a merely natural event, something that happens; and Kant is in a good position to explain why that is. For in willing proportionality, we will that virtue be rewarded with happiness qua moral worth or desert. And worth and desert are properties which only reason can pick out and respond to—to expect them to operate as natural forces is to misunderstand what they are. (As Iris Murdoch puts it: 'There is no triumph of good, and if there were it would not be a triumph of good.')[30] This is for Kant one of the most basic respects in which the ancient philosophers went wrong: if they were right that virtue was the natural, 'internal' cause of happiness, it could not be strictly *as* virtue (that is, commitment to the moral law) but only as something else (psychological health, perhaps); and the highest good would actually be precluded by *that* causal relation. Not only would corrupting incentives to virtue be pervasive and inescapable, there would be no room for any rational will to respond to virtue as such, and reward it as it deserves. From the Kantian perspective, the internal conditionality insisted on in the ancient ethical systems can only look like magic, an attempt to procure through our own powers what is properly the gift of another. It is a form of conditionality which leaves no room for justice to be done.[31]

All this entails that the highest good is impossible without God—not only in its perfect realization, but even locally. (Thus the distinction between the 'local' and 'perfect' versions of his argument drawn above, which seemed to show his argument to be equivocal, is in the end, Kant may argue, not significant.) For a merely local occurrence of proportionality can only count as an instantiation of the highest good to the extent that it can be seen as part of a necessary, universal, rationally willed

---

nature of Kantian willing, its relation to intending, and whether it can entail that local action for the highest good involves intending its perfect realization. I here offer what seems to me a less proprietary 'sufficiency' argument: local instances of the highest good would be missing something valuable, and would thus fail to *count as* instances of the highest good, if they were not the necessary effect of a rationally governed system.

[30] Strictly speaking, as Murdoch's character Bradley Pearson puts it, in *The Black Prince* (108).

[31] This is perhaps an overstatement. In some versions of ancient conditionality, notably Aristotle's, room is left for divine luck (*EE* VII.2), or for 'goods of fortune' (*EN* I.10–12) which affect our degree of happiness directly, not just by helping us to exercise virtue. But the goods of fortune are not generally presented as literally the result of divine intervention, and the gods typically befriend the virtuous anyway. Unambiguous references to divine interventions in human happiness in the ancient philosophers are fleeting, and generally look like attempts at harmonization with pre-philosophical *endoxa* (e.g. *EN* 1179a22–32). There *is* a deep commitment to divine providence in the Socratic tradition (Plato, *Apology* 31a–b, 41d, Xenophon, *Memorabilia* I.4), which flowers fully in the *Timaeus*, *Laws* X, and Stoic theology. But for the most part this attributes to God the establishment of a natural cosmic order in which (a) all is for the best, and (b) goodness and wickedness *naturally* bring about—or indeed *are*—their own reward and punishment in the long run (e.g. *Theaetetus* 176e–177a, *Timaeus* 42a–d, *Laws* 903b–905d). So divine intervention at the individual level must either be understood allegorically or be taken to operate *through* human character.

causal system. Only a supreme rational will can establish *that*; and such a will is in turn sufficient for the perfect realization of the highest good.

We can now get a better sense of how human and divine agency must be related in the highest good—namely, by a kind of *collaborative* agency. Only a tiny fragment of the highest good can be directly produced by my will; what I will in willing it *as* the highest good is that my actions form part of a greater, indeed a perfect whole. This kind of collaborative and other-referring willing seems to be what Kant means by *hope*. Its centrality is brought out by the status of hope as one of just three fundamental attitudes of reason, with the highest good as its distinctive object.[32] The highest good as a whole must be the object of *both* my will and my hope because it can only be realized by a collaboration between my own will and a rational will external to my own. (This is the truth behind the folk understanding that happiness depends on the hap—on how things turn out beyond our own control.) It is important that this attitude, demanded by the possibility requirement, is not a state of knowledge; this would exclude hope, and its certainty would corrupt moral agency with incentives. Indeed I would go further, and suggest that Kantian hope also should not be understood as a belief, or as any epistemic state or propositional attitude at all. (It is clearly not *reducible* to a belief that something is likely to happen—with what degree of probability, exactly?—together with a pro-attitude towards its happening. Such pro-attitudes might have an infinite number of miscellaneous objects, whereas Kantian hope has a single defining one.)[33] Belief in the normal sense is *not* subject to our will, and cannot be morally required. Hope in the relevant sense is not a propositional attitude or cognitive state but a practical stance, a condition of the will: it is something we allow or deny ourselves, and Kant's claim is that practical reason requires us to allow it.[34]

This practical stance is in a sense a negative one: hope is a *refusal* to despair, or to view one's own commitments as doomed or absurd. But we can also describe Kantian hope positively, as the state of the will in which we choose to view our actions as part of collaboration with a will greater than our own. Hope is what a player in a band, a string quartet, or a soccer team depends upon in acting. It is a commitment to a particular form of *seeing as*—namely, to seeing my commitments and actions as contributions to a common work commanded by God. Hence Kant's frequent use in this context of the language of seeing as, viewing, representing, recognizing, and so forth.[35] This is how the possibility requirement and the postulation of God can be genuinely new ethical requirements without adding to the *content* of my duty: they do not demand further actions, but require me to *view* all my actions in a certain light, as part of a collaboration with a rational will other and greater than my own. What makes the postulation of God required, then, is that without reference to that greater will, I cannot not perform the actions required of me under the descriptions which the moral law commands that I will them. For the moral law commands that I will the highest good as such, and to do that is to will my actions *as* contributions to a system of just rational causality. In the absence of God, the moral agent would not be like the man who thinks he is Napoleon, but like a sworn tax collector or border guard in a state that has fallen into anarchy. The problem is not that he will in fact fail in collecting taxes, but that he cannot even will his action under the description 'collecting taxes on behalf of the government', once he knows that there is no longer any government. (So long as the bad news does not reach him, there is no ethical problem, any more than there was for the Japanese army stragglers who never learned of the end of the war: hence the adequacy of a strictly subjective postulate.)

---

[32] A805/B833.

[33] If hope is a matter of belief, it seems that my belief must be an absolutely certain one, since the postulates are necessary conditions of the indisputable moral law. But Kant does not speak of it in those terms, and it is hard to imagine how certainty about divine proportionality could avoid corrupting our will with incentives.

[34] I intend this paragraph as a somewhat tentative gesture towards a reading of Kantian hope. A full discussion would have to consider the relation of hope to religious faith [*Glaube*], which does seem to be an epistemic or doxastic state. If hope is in the end identical or reducible to *Glaube*, the Spinozist objection above seems to me inescapable.

This ultimately collaborative conception of moral agency helps to explain why Kant feels entitled to offer both the secular and theological conceptions of the highest good, as if they were interchangeable. What is distinctive about hope is that part of what I will is deferred to the will of some independent Other; exactly *what* Other is less significant, and perhaps makes no difference to the moral quality of the attitude itself. Whether I place my hope in the future of mankind, or the *corpus mysticum* of the righteous, or God, is comparatively speaking a question of detail.[36]

In short: the commitment to the highest good required by practical reason is best understood as a commitment to seeing our own actions as collaborative contributions to a rational system of justice, a system in which the demands of the 'inner voice' are satisfied at last. Now although I cannot argue the case here, it seems to me that Kant is right to treat the inner voice, and its demand for proportionality, as a central fact of moral experience, one to be respected by any account of the highest good. It also seems to me defensible for him to hold that to will justice, as such, is to will a universal system of necessary causality mediated by rational wills, and that I must see my own contributions to that system as those of a merely junior partner. The challenge which remains, though, is for Kant to show that hope in this practical 'seeing as' sense is sufficient to require postulation of the existence of God—inescapably an epistemic or (more precisely) doxastic commitment as well as a practical one. At this point we can once again imagine an objection from the indefatigable Spinozist: why can I not undertake to think and will collaboratively, independent of any *belief* that collaborators exist? As a musician, I can play the second violin part of a symphony, as such, alone in the studio; as a private detective, I can follow the police code of conduct even if I have been struck from the force; as Don Quixote, I can obey the norms of chivalry in the wrong time and place; and as a Spinozist, I can follow the commands of a religion in whose God I do not believe. As Puddleglum the Marshwiggle says, at the darkest moment of *The Silver Chair*: 'I'm on Aslan's side even if there isn't any Aslan to lead it. I'm going to live as like a Narnian as I can even if there isn't any Narnia.'[37] This kind of *als ob* commitment seems to be a live psychological possibility, perhaps even a heroic one: if so, it shows that a Spinozist Marshwiggle could adopt the collaborative stance without adopting any epistemic postulates at all. Friedman reads Kant as, in effect, claiming that this could only be the stance of a madman. But this is, I think, to ignore the distinction between local and universal, and perhaps between intending and Kantian willing: even if I will my action collaboratively and universally *als ob*, I can at the same time intend and act locally in the world as it is. The assassin need not, in fact, think that there is any chance that all tyrants *will* be assassinated, when he utters his cry and draws his sword. He need only hope that his own particular action has some chance of success, and offer it in a spirit of contribution to a common project with whatever like-minded spirits may be out there—whether there are any or not.[38]

---

[35] E.g. 'the sensible world, viewed, however, as . . . a *corpus mysticum* of the rational beings in it' (A808/B836); 'To view ourselves, therefore, as in the world of grace . . . is . . . a necessary idea of reason' (A812/B840); 'we are necessarily constrained by reason to represent ourselves as belonging to such a world' (A811/B839); 'Religion is the recognition of all duties as divine commands' (5:129).

[36] Hence, given the validity of the secular conception of the highest good, Kantian ethics leads to politics as directly as to religion: cf. Engstrom 1996, 132–3, who rightly notes here a parallel with Aristotle.

V Conclusions

In the trajectory of argument I have sketched, Kant is, to a perhaps greater degree than is usually appreciated, writing not just a logically structured exposition of a theory, but also a narrative, with a moral-educational and psychagogic dimension. In this he is following, consciously or not, some powerful ancient models. Like Plato in the 'ascent' passages of the *Symposium*, *Phaedrus*, and *Republic*, or Cicero when he delineates the stages of Stoic moral progress in *De Finibus* III, Kant is depicting the stages through which the progressing moral agent must pass, so that we rehearse them imaginatively in following his argument: seeking happiness as an 'imaginative ideal', to be brought into focus as a rational end; grasping and committing to the moral law; coming to see both happiness and virtue as parts of the highest good, related by the norm of conditionality; affirming that this highest good is possible; and, finally, coming to hope for its achievement by collaboration with divine agency. Each step involves a rethinking and repositioning of concepts and principles previously grasped incompletely or in isolation. Above all, this ascent amounts to an ordered sequence of instructions on exactly how—in what spirit—we are to think about happiness. For each step in moral progress involves a reconception of what happiness is and how it is to be attained: from infantile fantasy to rational system, from the object of a selfish maxim to a universalized one, from our natural end to only a conditional part of the highest good, from the object of my rational will to the object of my hope. This final step is in a way the most radical rethinking of all: for at this point I come to think of my happiness as something I am to make myself worthy of, so that it may be conferred by divine justice, rather than as something to be directly procured by my own efforts.

---

[37] Lewis 1965, 156–7. Puddleglum has at this point been half-persuaded by the Witch's magic that there is no world above the subterranean cave in which he and the children have been trapped. He begins his peroration by admitting as much: 'Suppose we *have* only dreamed, or made up, all those things—trees and grass and sun and moon and stars and Aslan himself. Suppose we have. Then all I can say is that, in that case, the made-up things seem a good deal more important than the real ones. Suppose this black pit of a kingdom of yours *is* the only world. Well, it strikes me as a pretty poor one...' (156). This is very much the situation of the Spinozist forced to recognize the empirical world as a meaningless chaos (cf. n. 17 above), or Friedman's moral agent whose actions and self-conception find no support in the empirical world. Lewis seems to propose that one can occupy that position transparently and rationally, and that it may even be heroic to do so.

[38] To spell out this objection a bit more: if Kantian hope is the non-doxastic 'seeing as' which I have so far sketched, it does not require the postulation of God, but only a commitment to will *als ob*. If on the other hand Kant means to insist that I must postulate the existence of God in order to act even with some local instantiation of the highest good as my end, more argument is needed.

Though I have cited some ancient precursors, this particular narrative of moral ascent is not closely analogous to anything in ancient philosophical thought. And yet Kant's version of the highest good is intended as a rediscovery of some very old ways of thinking about happiness and virtue. He says so himself, when in the passage quoted earlier he cites the 'inner voice' which says, *This is not how it should be*, and attributes it to the very dawn of moral judgement and reflection.[39] Kant is right about that: the inner voice can be heard, for instance, in Hesiod's didactic poem *Works and Days*, and the argument there is worth juxtaposing with his own. For the most part, Hesiod simply affirms that Zeus rewards the righteous and punishes the unjust—all around us, in the empirical world, though with a rough justice. Cities wickedly governed suffer famine, plague, sterility, and war; wrongdoing always rebounds against the wrongdoer in the end, thanks to the justice of Zeus. But then we get an abrupt change of tack, quickly covered with a reassurance:

> The eye of Zeus sees all, and understands all,
> And when he wishes, marks and does not miss
> What sort of justice a city holds within.
> But as things are I would not myself be just, nor have my son
> Be just among men: for it is bad
> To be a just man where felons rule;
> But I trust wise Zeus to save me from this pass.[40]

---

[39] Given this passage, it is odd that Reath claims that 'the secular version is predominant in the third *Critique* and later works' (1988, 601).

[40] Lines 267–73, Wender trans., with revisions: in the last line, the construction is more literally 'I do not expect that' Zeus will allow this. 'Expect' (*elpizô*) can also mean 'hope', and is related to the noun 'hope' (*elpis*).

Abruptly juxtaposed with Hesiod's lengthy, repetitive promises of divine justice, this brief glimpse of the alternative has the air of the mask slipping, and a private doubt suddenly exposed. But this glimpse is itself a didactic strategy, and a clarification of what precedes it. For here we see that justice is, in Hesiod's world as in Kant's, deferred to the future tense. Of course there are important differences: Hesiod never considers deferring justice all the way to an afterlife, or to an intelligible order distinct from the sensible one. (The closest he comes is to defer retribution to a man's descendants.) Kant begins precisely where Hesiod refuses to go, with an acceptance that we cannot perceive any justice or proportionality in the empirical order. But there is nonetheless a formal and ethical similarity. In the end, Hesiod—like Kant—demands of us not an empirically grounded belief in divine justice but a first-personal commitment to hope, the only alternative to which would be a despairing refusal of moral demands as rationally unacceptable. Thus for Hesiod as for Kant, hope (which coincides with trust in divine justice) becomes a moral requirement.[41]

This kinship can help us to locate the Kantian conception of the highest good in relation to its philosophical predecessors and alternatives. In accordance with Adeimantus' demand, the conditionality affirmed by the ancient philosophers is a natural and internal relation: virtue constitutes happiness, or causes it, or is at least a psychological precondition for it, independent of outside circumstance. And Adeimantus' demand is provoked by a disgusted *rejection* of what poets like Hesiod have had to say about virtue and happiness (*Republic* 362d–367e). His question to Socrates is, among other things, a request for a story about proportionality which does not depend on divine intervention or any other external 'rewards'. For one thing the poets are, as he complains, inconsistent; sometimes they suggest that the gods favour the wealthy, who can buy them off with rituals and display, and that cult initiates have special advantages in the afterlife (364b–366a). Starting with Socrates, the ancient philosophical systems respond by taking the gods out of the loop—and with them the *daimon* and the 'hap'. By making the relation of happiness to virtue a matter of natural causality or even identity, the philosophers bring happiness fully under the virtuous agent's control.[42]

Kant's embrace of conditionality is a profound rejection of *this* way of conceiving it, and a reaffirmation of the more ancient Hesiodic conception.[43] That is why hope is a central value for both Hesiod and Kant, while it is only a peripheral value in any ancient system of philosophical ethics. And this sheds a retroactive light on Kant's rejection of the very idea of ancient *eudaimonia*: his assumption that *of course* happiness is just a matter of obtaining pleasures or satisfactions is from the start a rejection of the philosophers' strategies of internalization. Kant starts from a modern, pessimistic acceptance of the unmoralized nature of human inclination and satisfaction, and of the moral indifference of the world of experience; but his response is to recover an ethical stance much more ancient than the ancient philosophers.

---

[41] Hope is also significant for Hesiod as the last resident of Pandora's jar, a standing compensation for the evils of human life (*Theogony* 96, cf. West 1978 for discussion).

[42] Cf. Nussbaum 2001, though Nussbaum's contrast between Plato and Aristotle seems overstated in this context (that is, in comparison to what separates both of them from the poets and from Kant alike).

[43] As Kant himself sees: 'it now becomes clear why the Greek schools could never succeed in solving their problem of the practical possibility of the highest good. It was because they made the rule of the use of which the human will makes of its freedom the sole and self-sufficient ground of its possibility, *thinking that they had no need of the existence of God for this purpose*' (5:126, my emphasis). This confirms what we can also infer from the 'inner voice' passage, namely that Kant's postulation of divine agency, and of hope as an ethical requirement, are not considered by him either as needed by or conceivable to Christians alone.

Whether this Kantian response represents philosophical progress, theological regress, or just another turn of the great dialectical wheel is too hard a question for me. Viewed from the outside, what is striking about this debate is that it is one in which *both* sides can criticize the other as naïve. From the Kantian standpoint, Adeimantus' demand is immature and doomed, an invitation to self-deception: only by fudging the empirical data and adopting an artificially internalized conception of 'happiness' can a Plato or Aristotle insist that my virtue will naturally bring its own rewards. But from the Aristotelian standpoint, it is the Kantian who is naïvely optimistic, even primitive in his hope for a divine justice of which the empirical world gives no evidence. It is possible, of course, that both accusations are correct.[44]

References

Beck, L. W. 1960. *A Commentary on Kant's Critique of Practical Reason*. Chicago: University of Chicago Press.

Colli, G., and Montinari, M. (eds.) 1999. *Nietzsche: Sämtliche Werke*. 15 vols. Munich; Berlin; New York: De Gruyter.

Cooper, J. M. (ed.) 1997. *Plato: Complete Works*. Indianapolis: Hackett.

Despland, M. (ed.) 1973. *Kant on History and Religion*. Montreal: McGill-Queens University Press.

Engstrom, S. 1996. 'Happiness and the Highest Good in Aristotle and Kant'. In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by S. Engstrom and J. Whiting, pp. 102–38. Cambridge: Cambridge University Press.

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

Engstrom, S., and J. Whiting, J. (eds.) 1996. *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*. Cambridge: Cambridge University Press.

Friedman, R. Z. 2008. 'The Importance and Function of Kant's Highest Good'. *Journal of the History of Philosophy* 22: pp. 325–42.

Guyer, P. 2006. *Kant*. London: Routledge.

---

[44] For helpful discussion and/or comments on earlier drafts of this chapter I would like to thank Steve Engstrom, Béatrice Longuenesse, Sergio Tenenbaum, Jens Timmermann, the participants in the St Andrews conference in 2010, two anonymous readers for OUP, and above all the editors.

Guyer, P. 2000. ‘From a Practical Point of View: Kant’s Conception of a Postulate of Pure Practical Reason’. In *Kant on Freedom, Law, and Happiness*, pp. 333–71. Cambridge: Cambridge University Press.

Guyer, P. (ed.) 1992. *The Cambridge Companion to Kant*. Cambridge: Cambridge University Press.

Hill, T. E. 1999. ‘Happiness and Human Flourishing in Kant’s Ethics’. *Social Philosophy and Policy* 16: pp. 143–75.

Holton, R. 2008. ‘Partial Belief, Partial Intention’. *Mind* 117: pp. 27–58.

Infield, L. 1981. *Kant: Lectures on Ethics*. Indianapolis: Hackett.

Irwin, T. 1996. ‘Kant’s Criticism of Eudaemonism’. In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, edited by S. Engstrom and J. Whiting, pp. 63–101. Cambridge: Cambridge University Press.

Johnston, M. 2010. *Surviving Death*. Princeton: Princeton University Press.

Kant, I. 1996. *Religion and Rational Theology*, ed. and trans. A. Wood and G. Di Giovanni. Cambridge: Cambridge University Press.

Kant, I. 1987. *Critique of Judgment*, trans. W. S. Pluhar. Indianapolis: Hackett.

Kant, I. 1964. *Groundwork of the Metaphysics of Morals*, trans. H. J. Paton. New York: Harper.

Kant, I. 1956. *Critique of Practical Reason*, trans. Lewis White Beck. Indianapolis: Bobbs-Merrill.

Kant, I. 1929. *Critique of Pure Reason*, trans. N. Kemp Smith. New York: St. Martin’s Press.

Kuehn, M. 1985. ‘Kant’s Transcendental Deduction of God’s Existence as a Postulate of Pure Practical Reason’. *Kant-Studien* 76: pp. 152–69.

Lewis, C. S. 1965. *The Silver Chair*. Harmondsworth: Penguin.

Long, A. A., and Sedley, D. N. 1987. *The Hellenistic Philosophers*. 2 vols. Cambridge: Cambridge University Press.

Nussbaum, M. C. 2001. *The Fragility of Goodness*. 2nd edn. Cambridge: Cambridge University Press.

O’Connor, D. 1982. ‘Kant’s Conception of Happiness’. *Journal of Value Inquiry* 16: pp. 189–205.

Reath, A. 1989. ‘Hedonism, Heteronomy and Kant’s Principle of Happiness’. *Pacific Philosophical Quarterly* 70(1): pp. 42–72.

Reath, A. 1988. ‘Two Conceptions of the Highest Good in Kant’. *Journal of the History of Philosophy* 26: pp. 593–619.

Scutt, M. Z. 2010. ‘Kant’s Moral Theology’. *British Journal for the History of Philosophy* 18: pp. 611–33.

Watson, G. 1984. ‘Kant on Happiness in the Moral Life’. *Philosophy Research Archives* 8: pp. 79–108.

Wender, D. (ed. and trans.) 1976. *Hesiod and Theognis*. Harmondsworth: Penguin.

West, M. L. (ed.) 1978. *Hesiod: Works and Days*. Oxford: Clarendon Press.

Wike, V. S. 1994. *Kant on Happiness in Ethics*. Albany, NY: State University of New York Press.

Wike, V. S. 1987. ‘The Role of Happiness in Kant’s Groundwork’. *The Journal of Value Inquiry* 21: pp. 73–8.

Wood, A. 2005. *Kant*. Malden, MA: Blackwell.

Wood, A. 1992. ‘Rational Theology, Moral Faith, and Religion’. In *The Cambridge Companion to Kant*, edited by P. Guyer, pp. 394–416. Cambridge: Cambridge University Press.

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