le: Aalen, 1963
Wolff, Christian 1764-66: *Ius naturae methodo scientifico pertractatum*, 8 vols. Frankfurt a.M./Leipzig; reprodução no mesmo: *Gesammelte Werke*, ed. por J. École entre outros, Hildesheim, 1969
— 1740: *Institutiones iuris naturae et gentium*, Halle/Magdeburg; alemão: *Grundsätze des Natur — und Völkerrechts*, trad. por G. S. Nicolai, Halle 1754; reimpressão no mesmo: *Gesammelte Werke*, ed. por J. École e outros, Hildesheim, 1965

Self-understanding and philosophy: the strategy of Kant's *Groundwork**

Paul Guyer
University of Pennsylvania

The topic of this paper is the relationship between "common" and "philosophical" "rational cognition of morals" (*sittlichen Vernunfterkenntnis*) that Kant has in mind in the *Groundwork of the Metaphysics of Morals.* One striking fact about the argument of the *Groundwork* is that philosophy appears in it in the guise of both "popular moral worldly wisdom" — here it may be best to translate *Weltweisheit* literally rather than as just an archaic word for "philosophy" — on the one hand and that of Kant's own "metaphysics of morals" and "critique of pure practical reason" on the other. More precisely, the *Groundwork* starts off by deriving its initial formulation of the categorical imperative from a "common rational cognition of morals" that every reasonable person is supposed to recognize as his or her own, and then relates both "popular worldly wisdom" and a genuine "metaphysics of morals" and "critique of pure practical reason" to this "common rational cognition of morals." What is the relation of these two forms of philosophy to each other and to the pre-philosophical self-understanding from which Kant's argument seems to begin?

One interpretation might be this: Kant thinks that every normal human being innately possesses or naturally acquires an understand-

* Este artigo também será publicado no volume *Philosophie in Synthetischer Absicht – Synthesis in Mind*, editado por Marcelo Stamm, Stuttgart: Klett-Cotta, 1998. Gostaríamos de agradecer ao editor pela autorização para que o artigo pudesse ser publicado no primeiro número da revista Studia Kantiana.

ing of the demands of morality, which however can be corrupted by bad but popular philosophy or "worldly wisdom." Such philosophy, a kind of empiricism in the sense that it bases its prescription of how moral agents ought to behave on observation of how humans actually do behave, could corrupt our innate sense of duty in two ways: it could confuse us about the principle of morality by presenting our own happiness rather than duty as the object of morally worthy action; and it could give us an excuse for lapsing from the stern demands of duty by appealing to determinism as an excuse for our own moral frailty or evil. Kant's strategy, then, would be to counteract the deleterious effects of this "popular worldly wisdom" by showing, in his "metaphysics of morals" (that is, Section II of the *Groundwork*) that happiness can never be the object prescribed by the categorical imperative that we are all disposed to acknowledge, and by showing, in his "critique of pure practical reason" (that is, Section III of the *Groundwork*), that we really are always free to live up to the stern demands of morality no matter what our prior experience and history might seem to have determined us to do. In this way, a sound philosophy would save our innate self-understanding as moral agents from corruption by bad philosophy.

Such an interpretation of Kant's strategy in the *Groundwork* is almost right, but not quite. The risk to our moral self-understanding does not arise from without, from the wily artifices of corrupt philosophers who appear out of nowhere to darken our paths; the source of sophistry and corruption lies within us, in possibilities inherent to our own nature, which can in turn co-opt our own faculty of reason to produce a form of philosophy that would appear to justify our lapses from duty unless countered by a sounder philosophy that is itself another product of our own reason accessible to any of us. Just as our theoretical reason is inherently liable to irresolvable dialectical disputes or antinomies until we fully understand the proper conditions for its use — dialectical disputes that find expression in speculative philosophy but are by no means mere artifices of speculative philosophers[1] — so our practical reason is inherently liable to undermine our common rational cog-

nition of morals by a dialectic that is entirely natural to it, which is thus not caused by but merely expressed in popular moral philosophy or worldly wisdom, and which can only be resolved by a sounder philosophical reflection on the nature of our practical reason and the conditions of its use of which we are also capable. Kant's moral philosophy is not intended simply to rescue our moral self-understanding from bad philosophy contingently imposed upon us by bad philosophers. It is rather intended to give us the fuller self-understanding that we need in order to save our understanding of the moral law and its demands from our own self-misunderstanding and the bad philosophy that we create for ourselves.

This subtle strategy might not be entirely obvious from Kant's first statement of the need for his moral philosophy in the Preface to the *Groundwork*:

> A metaphysics of morals is therefore indispensably necessary, not merely because of a motive to speculation — for investigating the source of the practical basic principles that lie *a priori* in our reason — but also because morals themselves remain subject to all sorts of corruption as long as we are without that clue and supreme norm by which to appraise them correctly. For, in the case of what is to be morally good it is not enough that it *conform* with the moral law but it must also be done *for the sake of the law,* without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also often produce actions contrary to the law. Now the moral law in its purity and genuineness (and in the practical this

1 This is, after all, why Kant always treats the proponents of the dialectically opposed theses and antitheses of his antinomies, Leibniz and Wolff or Locke and Hume, with the utmost respect, rather than dismissing them as fools: they are only giving voice to natural illusions of human reason which cannot fully be dispelled until the nature and conditions of the use of human reason are fully understood.

is what matters most) is to be sought nowhere else than in a pure philosophy... (*G*, 4:390)[2]

Initially, this passage simply leaves open the source of any tendencies to the corruption of morals. It then perhaps goes on to suggest that the problem is that while we may have a natural desire to conform to the requirements of morality, whatever they might be, without a clear recognition of the principle of morality and its requirements, particularly its requirement that we act for the sake of the moral law itself, we are open to all sorts of corruption. Thus we might suppose that Kant means to argue that our natural disposition to conform to the demands of morality has to be supplemented by a clearly formulated principle of morality, and that it is up to his moral philosophy to offer a sound one to compete with, and triumph over, the less sound ones offered by more popular worldly wisdom or philosophy.

The conclusion of Section I of the *Groundwork*, however, suggests the more subtle strategy I have ascribed to Kant. Here Kant writes:

> Reason issues its precepts unremittingly, without thereby promising anything to the inclinations, and so, as it were, with disregard and contempt for those claims...But from this there arises a *natural dialectic*, that is, a propensity to rationalize against those strict laws of duty, and to cast doubt upon their validity, or at least upon their purity and strictness, and, where possible, to make them better suited to our wishes and inclinations, that is, to corrupt them at their basis and to

2 Passages from Kant's writings will be cited with the volume and page number of the Academy edition, *Kant's gesammelte Schriften*, edited by the Royal Prussian (later German) Academy of Sciences (Berlin: Georg Reimer [later Walter de Gruyter], 1900-). Translations from the *Groundwork*, abbreviated "*G*," are from Immanuel Kant, *Practical Philosophy*, edited and translated by Mary J. Gregor (Cambridge: Cambridge University Press, 1996). Since this and other volumes in *The Cambridge Edition of the Works of Immanuel Kant* include the pagination of the Academy edition, the pagination of the translation does not need to be cited.

> destroy all their dignity — something that even common practical reason cannot, in the end, call good. In this way *common human reason* is impelled, not by some need of speculation (which never touches it as long as it is content to be mere sound reason), but on practical grounds themselves, to go out of its sphere and to take a step into the field of *practical philosophy*, in order to obtain there information and distinct instruction regarding the source of its principle and the correct determination of this principle in comparison with maxims based on need and inclination, so that it may escape from its predicament about claims from both sides and not run the risk of being deprived of all genuine moral principles through the ambiguity into which it easily falls. (*G*, 4:405)

The dialectic of practical philosophy is not academic but natural: that is, the corruption of "common human reason" is threatened not from without, but from within, and "popular worldly wisdom" is not an external threat to "common human reason" but is itself an expression of something natural to human beings that can be resolved only by the self-understanding afforded by a "complete critique of our reason." In particular, common human reason tends to confuse its natural recognition of the genuine principle of morality with "maxims based on need and inclination." Only a clear distinction between the fundamental principle of morality that we all intuitively recognize from any maxims based on need and inclination, but at the same time an equally clear understanding of the proper role of need and inclination in the conditions of human agency, will enable us to save ourselves from the dialectic of practical reason that is as natural to us as is our recognition of the fundamental principle of morality itself.

This program of self-understanding is carried out in the *Groundwork*, I suggest, in the following steps.

1) In Section I, Kant argues that a genuine even if not entirely explicit understanding of the fundamental principle of morality is reflected in our common conceptions of good will and duty and in the

moral judgments that we make about particular cases of human action, especially when those cases are presented to us in ways that do not immediately involve our own interests. From our common conception of good will and duty and from such particular cases, a clear formulation of the genuine principle of morality can be extracted (4:402). This clear recognition of our duty, however, is threatened by two factors that are as natural to our condition as is the recognition of our duty itself. First, it is entirely natural for us each to seek our own happiness, and thus the risk of substituting an imperative to seek our own happiness for the imperative to perform our duty is equally natural to us. We also try to dignify this tendency by adopting a philosophy that seems to entail hedonism. Second, as far as we can see, human beings frequently succumb to this confusion; thus insofar as we try to base our moral principles on actual examples of human conduct, and moreover even try to dignify this procedure by thinking of it as dictated by what appears to be a respectable empirical philosophy, we will tend to substitute the principle of happiness for the genuine principle of duty.

2) In Section II of the *Groundwork*, Kant argues that this natural danger can be avoided only by making completely explicit the fundamental principle of morality that is merely implicit in our initial moral self-understanding. In the fuller development of his moral theory, however, he will also have to show how the interest in happiness, precisely since it is entirely natural to us and neither should not nor cannot be extirpated, is to be incorporated into the object of morality in the form of the highest good, the realization of happiness conditioned by the worthiness to be happy. If the principle of happiness were just a threat from bad moral philosophy, not an ineliminable feature of human nature, the theory of the highest good would not be a necessary part of Kant's moral philosophy.

3) Section III of the *Groundwork* then takes up the second threat to our natural recognition of duty, a threat which is just as natural to us as the interest in happiness and just as much as that needs to find its proper place in a complete self-understanding of our moral agency. This is the threat of determinism, or as Kant himself calls it "pre-determinism,"[3] the doctrine that our actions at any given moment are thoroughly necessitated by events at prior moments, from which it seems to follow that it is not always in our power to live up to the stringent demands of duty, which must therefore be weakened. Such a doctrine of determinism is clearly a sophistry of our own reason by which we can excuse our failure to act as we know we should, and needs to be answered by a critique of pure practical reason that will show that we do indeed always have the power to act as duty requires no matter what our past might seem to predict. However, the doctrine of determinism is not just a rationalization of our moral weakness offered to us by popular philosophy (although of course historically it was a prominent doctrine of the empiricist philosophies of Locke and Hume), but is itself a genuine aspect of our self-understanding, indeed the indispensable foundation of our theoretical understanding of the world of nature and our place in it. Thus what is necessary is not a *refutation* of determinism, but rather a proper *situation* of it in the fuller self-understanding that we can reach through a sound philosophy — which must also lie ready in ourselves. This is of course what Kant attempts to provide through the transcendental idealism which he invokes in Section III of the *Groundwork* and then again in the *Critique of Practical Reason.*[4] Just as Kant's theory of the highest good is his recognition that we cannot simply dismiss the principle of happiness but must incor-

3 See *Religion within the Boundaries of Mere Reason*, 6:49-50n.

4 Dieter Henrich has, discussed the vexed issue of the relation between Kant's treatment of freedom in the *Groundwork* and in the second *Critique* in his famous paper, "Die Deduktion des Sittengesetzes: Über die Gründe der Dunkelheit des letzten Abschnittes von Kant's *Grundlegung zur Metaphysik der Sitten*," in Alexander Schwann, ed., *Denken im Schatten des Nihilismus: Festschrift für Wilhelm Weischedel* (Darmstadt: Wissenschaftliche Buchgesellschaft, 1975), pp. 55-110. An English translation of most of this paper appears in Paul Guyer, ed., *Kant's Groundwork of the Metaphysics of Morals: Critical Essays* (Lanham: Rowman & Littlefield, 1998), pp. 303-41. In the present paper, I will assume that Kant's transcendental idealist conception of freedom is the same in both works, although the arguments by which he introduces it are different; and even then, I will suggest below, there is a crucial structural similarity between the two works' arguments for transcendental idealism, as Kant tries to argue that the transcendental idealist understanding of freedom is in fact just as natural to us as the natural theory of determinism which is the internal threat to morality that a fuller self-understanding must resolve.

porate it into our full understanding of the object of morality, so he also recognizes that we cannot simply dismiss determinism as a groundless threat to our sense of duty but must rather show its proper place in relation to the indisputable fact of our freedom in the transcendental idealism that gives fullest expression to our self-understanding as moral agents.

I will hardly have room here to explicate and argue for these three claims in the detail they require. I will just be able to present some of the key evidence for these claims and to comment on some of the issues that they raise.

1) From the outset of the *Groundwork*, Kant insists that everything essential in moral philosophy is readily accessible to the ordinary human being. The Preface maintains that "in moral matters human reason can easily be brought to a high degree of correctness and accomplishment, even in the most common understanding" (*G*, 4:391). The argument of Section I then takes the form of deriving the first formulation of the fundamental principle of morality from an analysis of the concept of a good will that is taken to be commonly acknowledged, where the common possession of this concept is itself confirmed by our common response to hypothetical examples of the performance of duty, such as the case of the man who has been created without sympathetic inclinations or lost them through his own misfortunes yet who can nevertheless act virtuously out of his respect for duty (*G*, 4:398). We can consider this style of argument to be continued in Section II of the *Groundwork* when Kant appeals to commonly accepted examples of duty — now ranged in four classes — to confirm not the common concept of duty itself but rather the first and second formulations of the categorical imperative to which this concept of duty gives rise (*G*, 4:421-3 and 429-30). Throughout, Kant's strategy is to show that the moral principle that he proposes — which is hardly supposed to be a new invention, "as if, before him, the world had been ignorant of what duty is"[5] — is implied by the commonly shared conception of duty and expressed in commonly shared judgments about particular cases of dutiful action.

5 *Critique of Practical Reason*, 5:8; translation from Gregor, *Practical Philosophy*.

Kant does not make his assumption clear when he first introduces the analysis of the concept of a good will: here he just says, without any methodological comment, that in order "to explicate the concept of a will that is to be esteemed in itself...we shall set before ourselves the concept of **duty**" (*G*, 4:397). Upon having derived his initial formulation of the only possible principle for the determination of the good will from this concept, however, he states that "Common human reason also agrees completely with this in its practical judgments and always has this principle before its eyes" (*G*, 4:402). A page later, he reiterates that "we have arrived, within the moral cognition of common human reason, at its principle, which it admittedly does not think so abstractly in a universal form, but which it actually has before its eyes and uses as the norm for its appraisals" (*G*, 4:403). The fundamental principle of morality is implicit in our common conception of duty and in our common judgments about duties, and even if not already explicitly formulated by every normal human being it will still be immediately recognized and acknowledged when presented to any normal human being in its explicit form.

As I am here more concerned with the form of Kant's account than with its content, the details of his analysis and its confirmation can be recalled briefly. Kant analyzes the concept of a good will by means of three propositions that are obviously supposed to be acknowledged by anyone with common moral cognition: i) good will consists in acting from duty rather than from inclination (*G*, 4:397); ii) "action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the *principle of volition*" (*G*, 4:399-400); and iii) "*duty is the necessity of an action from respect for law*" (*G*, 4:400). The first proposition of the analysis in particular is confirmed by an appeal to an example: we all recognize that there is no manifestation of good will and thus no special moral worth in a grocer's maintaining a policy of honesty for the sake of his

own long-term interest or in somebody preserving his life merely out of inclination (*G*, 4:397), but we do recognize good will and thus moral worth when somebody "preserves his life without loving it" or continues to act benevolently even though his mind has been "overclouded by his own grief" (*G*, 4:398). Our judgments of moral worth in such cases can only be explained by our assumption that moral worth lies in the performance of actions out of the motive of duty rather than out of inclination or self-interest. The first phase of Kant's analysis of the concept of a good will in terms of the concept of duty is thus confirmed by commonly accepted moral judgments.

After completing his analysis of duty, Kant then derives the formula *I ought never to act except in such a way that I could also will that my maxim should become a universal law*," which he will designate as the first formulation of the categorical imperative after he has introduced the concept of such an imperative in Section II, from the fact that since this analysis has "deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law" (*G*, 4:402). The validity of this formula is then again confirmed by an example: if we consider whether we may make a promise without the intention of keeping it in order to get out of a current difficulty, we all realize that the relevant question is not whether it is possible or prudent to do so, but rather simply "would I indeed be content that my maxim (to get myself out of difficulties by a false promise) should hold as a universal law (for myself as well as for others)...?" (*G*, 4:403). Kant's previously cited claims that his formulation of the principle of morality is reflected in the practical judgments of common human reason immediately precede and succeed the exposition of this example (*G*, 4:402 and 403). Kant's argument in *Groundwork* I thus has the following form: our common conception of good will as manifest in the performance of action from duty, which is supported by examples of virtuous action that we all recognize, combined with an analysis of the concept of duty that we all accept, gives rise to a formulation of the fundamental principle of morality, which, even if we do not explicitly recog-

nize it in its most abstract form, is in fact the basis for the particular moral judgments that we make, as can again be confirmed by an appeal to any example of a duty that we all acknowledge.

Kant sums up this first stage of his work by saying that "there is no need of science and philosophy to know what one has to do in order to be honest and good, and even wise and virtuous" (*G*, 4:404). If this is so, why does the *Groundwork* need its Sections II and III? Kant's answer to this question is not that we need philosophy simply in order "to present the system of morals all the more completely and comprehensibly and to present its rules in a form more convenient for use," but rather that "innocence...is easily seduced," because the "human being feels within himself a powerful counterweight to all the commands of duty," namely, "the counterweight of his needs and inclinations, the entire satisfaction of which he sums up under the name happiness" (*G*, 4:404-5). The next stage of the argument of the *Groundwork*, which occupies Section II, must then be to distinguish clearly the principle of morality from the principle of happiness; in Kant's moral philosophy more broadly considered, however, the next stage of the argument must be not merely to distinguish the principle of morality from the natural pursuit of happiness but also to show how happiness, as the ineliminable natural end of human beings, does properly fit into the complete object of morality. Before turning to this next stage of Kant's argument, however, several comments on the character of its first stage are in order.

First, there is a question about what sort of moral principle could be derived by the appeal to common concepts and appraisals that Kant presents in *Groundwork* I. Thus far, I have referred to both the fundamental principle of morality and the categorical imperative without distinction, but of course, these are not exactly the same: as Konrad Cramer has argued, the fundamental principle of morality can be considered a pure synthetic *a priori* principle, applicable to any and all rational beings, while the categorical imperative is an impure synthetic *a priori* principle, the form in which the fundamental principle of morality presents itself to beings like us, who empirically know ourselves to have in-

clinations and interests that may conflict with compliance with the fundamental principle of morality, and thus may experience this principle as a constraining obligation — a categorical imperative — in a way that beings without such conflicting incentives would not.[6] Shouldn't a derivation of a moral principle that appeals to commonly shared concepts such as those of good will and duty and to commonly shared responses to particular examples of duties and dutiful sorts of persons yield at best an impure formulation of the fundamental principle of morality in the form of a categorical imperative applicable to beings like us only, rather than the fundamental principle of morality itself in its pure form? Indeed, shouldn't a derivation of a principle of morality in any form from common concepts and judgments yield only something empirical, not any sort of *a priori* principle at all, that is, a principle that is universally and necessarily valid for any species of rational agents, let alone all rational agents? To answer this question, we need to distinguish carefully between a derivation of the *formulation* of the principle of morality (in any form) and a derivation of its *validity*. It cannot be Kant's position that we derive the *validity* of the moral law by any sort of empirical method from commonly accepted concepts and judgments. Rather, his view is that by *reflection* on our common concepts of good will and duty and on common moral judgments of particular examples of duties and dutiful persons we can see that we already acknowledge the validity of the moral law even in its purest form, its form as the fundamental principle of morality, as well as in its form as the categorical imperative, and even if we have not previously explicitly formulated the principle in any abstract terms at all. We immediately see that our recognition of the principle is what explains our acceptance of the concepts and judgments that we all do accept; we don't only come to accept the principle because of our response to particular cases. As Kant says, "Nor could one give worse advice to morality than by wanting to derive it from examples. For, every example of it represented to me must itself first be appraised in accordance with principles of morality, as to whether it is also worthy to serve as an original example" (*G*, 4:408).[7]

But at this point a second question about an argument involving appeal to examples arises. At the outset of Section II of the *Groundwork*, Kant inveighs against any attempt to derive the fundamental principle of morality in any form from *actual examples* of human conduct:

> If we have so far drawn our concept of duty from the common use of our practical reason, it is by no means to be inferred from this that we have treated it as a concept of experience. On the contrary, if we attend to experience of people's conduct we meet frequent, and as we ourselves admit, just complaints that no certain example can be cited of the disposition to act from pure duty; that, though much may be done *in conformity with* what *duty* commands, still it is always doubtful whether it is really done *from duty* and therefore has moral worth....In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty. (*G*, 4:406-7)

Doesn't this blunt statement completely undermine any attempt to derive anything about the moral law from examples of any kind?

To answer this question, we need to distinguish between the use of *actual* and of *hypothetical* examples of moral conduct. Kant's initial argument in Section II is that we cannot be sure that any *actual* conduct, that of others or even our own, has been performed out of the pure motive of duty, and thus we would be hard-pressed to derive a fundamental principle of morality from actual human behavior in the face of

6 Konrad Cramer, "Metaphysik und Erfahrung in Kants Grundlegung der Ethik," in Gerhard Schönrich and Yasushi Kato, eds., *Kant in der Diskussion der Moderne* (Frankfurt am Main: Suhrkamp, 1996), pp. 280-325; originally published in *Metaphysik und Erfahrung: Neue Hefte für Philosophie* 30-31 (1991): 15-68.

7 See also Kant's discussion of Christ as a model for our own morality in *Religion within the Boundaries of Mere Reason*, 6:62-4.

such uncertainty; indeed, we might even take him to go on to argue that we can be reasonably sure that almost all actual deeds, whether our own or others, have been motivated by inclination and self-interest, thus that if we attempt to formulate a fundamental principle of morality by induction from actual conduct we shall almost certainly come up with the *wrong* principle. In particular, we know that in actual cases of actions in any way affecting our own interests, our judgments are likely to be distorted by self-love (*G*, 4:407). But the examples in *Groundwork* I are not actual examples of human conduct, but hypothetical cases for moral judgment; and in the case of such examples what is at issue is only the question of how we judge that agents in such cases *ought* to be appraised, not whether we ourselves or anyone else ever actually lives up to such judgments. Kant's claim is precisely that in the appraisal of hypothetical cases and situations of human action, where the threat of self-love can be set aside, we all immediately recognize how human agents ought to be motivated and to behave, even if we are not sure that any of us has ever actually been motivated in that way. And the basis of such acknowledgments of the principle of morality, Kant insists, is not experience but pure practical reason. Examples need to be adduced for the confirmation of our common concepts of good will and duty because pure practical reason commonly expresses itself in the judgment of particulars rather than in abstractions, but not because these concepts rest on experience rather than pure reason.

2) Kant's argument in *Groundwork* II is that the true principle of morality can never be discovered by examples from ordinary experience; rather, it requires "pure rational concepts" and a "metaphysics of morals" (*G*, 4:410), although once "the doctrine of morals" has been "first *grounded* on metaphysics" it can be "provided with *access* by means of popularity" (*G*, 4:409). By a "popular philosophy" (here he does use the word *Philosophie* instead of *Weltweisheit)* he means simply a method "that goes no farther than it can by groping with the help of examples" (*G*, 4:412). Thus, he does not explicitly identify "popular philosophy" with a specific school of academic moral philosophy, such as the moral-

sense school as an applied form of academic empiricism. In fact, he clearly means to include the perfectionism of Wolff and his followers as well as the moral-sense philosophy of Hutcheson and Hume under this rubric—what we get if we attempt to discover the principle of "morality in that popular taste" is a hodgepodge of principles identifiable with those of all the popular schools of moral philosophy: "One will find now the special determination of human nature (but occasionally the idea of a rational nature as such along with it), now perfection, now happiness, here moral feeling, there fear of God, a bit of this and a bit of that in a marvelous mixture" (*G*, 4:410). But the overall argument of Section II is certainly a polemic against what Kant assumes would be inevitably suggested by basing our conception of the fundamental principle of morality on observation of actual examples of human motivation and behavior, namely, the idea that what morality prescribes is the pursuit of happiness as such — as Hume puts it, the cultivation of qualities useful and agreeable to ourselves and others.[8]

Kant carries on his polemic against any such principle in several stages. First, he derives the concept of a categorical imperative in general from what he clearly assumes to be the common understanding that the fundamental principle of morality must be an objectively necessitating principle, that is, a principle necessitating certain principles of action for all relevant agents (*G*, 4:413-14), and then argues that a simple principle of pursuing happiness could never give rise to a categorical imperative but only a hypothetical one. Such a principle would be a hypothetical one not because it is entirely contingent whether anyone adopts happiness as an end — on the contrary, Kant recognizes it as a fact of nature, a "natural necessity" (*G*, 4:415), that everyone does adopt happiness as an end — but rather because of the following sorts of con-

8 David Hume, *An Enquiry concerning the Principles of Morals*, Section IX, Part I; in the second edition of Hume's *Enquiries* by L.A. Selby-Bigge (Oxford: Clarendon Press, 1902), p. 268 (the pagination remains the same in the third edition of Selb-Bigge, revised by P.H. Nidditch in 1978).

siderations: it is contingent what *particular* ends it is the satisfaction of which would constitute anyone's happiness; it is contingent whether the various particular ends the satisfaction of which would constitute a single person's happiness are conjointly realizable (*G*, 4:418); and, as Kant adds in the *Critique of Practical Reason*, it is also contingent whether two or more different persons' conceptions of happiness are conjointly realizable (5:28). For these sorts of reasons, then, although it is not exactly contingent that anyone has happiness as an end, it certainly would be contingent whether anyone has as his end a particular conception of happiness that could rationally be pursued in the actual circumstances of his life.

After his initial contrast between merely hypothetical imperatives and a categorical imperative, Kant argues that the very concept of a categorical imperative gives rise to precisely the same formulation of the fundamental principle of morality that the previous analysis of the concepts of good will and duty yielded, the principle that one should only act on maxims that can at the same time be willed as universal law (*G*, 4:421). In the discussion of this and the following further formulations of the categorical imperative, especially the principle of humanity as an end in itself,[9] Kant continues to emphasize that the principle of morality is not "a subjective principle on which we might act if we have the propensity and inclination," thus not a principle prescribing our happiness, but an "objective principle on which we would be *directed* to act even though every propensity, inclination, and natural tendency of ours were against it" (*G*, 4:425), thus a principle that apparently ignores all reference to our own happiness. Kant stresses that the categorical imperative abstracts from all "subjective ends," and is thus *formal* rather than *material* (*G*, 4:428); but since human beings cannot act without an end at all, he elevates humanity into *an end in itself*," "which is the supreme limiting condition of the freedom of action of every human being," "an objective end that, whatever ends we may have, ought as law to constitute the supreme limiting condition of all subjective ends" (*G*, 4:430-1). Thus, Kant insists, the end of morality is not happiness, the satisfaction of our

particular, subjective material needs and inclinations, but is rather something else, humanity as such, which is a limiting condition on the pursuit of happiness. This is not the result that we would get by induction from actual examples of human motivation, even if we were to dignify such an induction with the name of philosophy, but is the result that we get from reflection on the pure concepts of a metaphysics of morals that is in fact accessible to each of us.

Now at the height of the polemic against founding a principle of morality on the object of happiness in *Groundwork* II, Kant goes so far as to say not merely that "all objects of the inclinations have only a conditional worth," but also that "the inclinations themselves, as sources of needs, are so far from having an absolute worth so as to make one wish to have them, that it must instead be the universal wish of every rational beings to be altogether free from them" (*G*, 4:428). However, if this suggests that human beings either could or should eradicate all inclinations in themselves, thus eradicating everything the satisfaction of which could produce happiness, and that the goal of morality could or should be pursued by means of such a mass extinction of inclination, then it radically misrepresents what will become the considered position of Kant's moral philosophy. As Kant makes clear in *Religion within the Limits of Mere Reason*, we are not evil because we *have* sensuous inclinations, but because of the fundamental maxim regarding them that we adopt. We are not evil simply because we have such inclinations, first because we "cannot presume ourselves responsible for their existence" (6:35),[10] but even more because "predispositions in the human being" are "*original*," that is, "they

9 For the classical exposition and discussion of the various formulations of the categorical imperative, see H.J. Paton, *The Categorical Imperative* (London: Hutchinson University Library, 1947). There has been much recent discussion of this subject; for my own approach, as well as references to much of the recent literature, see my "The Possibility of the Categorical Imperative," *Philosophical Review* 104 (1995): 353-85.

10 Translation by George di Giovanni, from Immanuel Kant, *Religion and Rational Theology*, translated and edited by Allen W. Wood and George Di Giovanni (Cambridge: Cambridge University Press, 1996).

belong to the possibility of human nature," and — certainly on the teleological view of nature which Kant had long assumed should regulate our reflection on our natural endowments[11] — they must therefore be assumed to be "not only (negatively) good (they do not resist the moral law) but they are also predispositions *to the good* (they demand compliance with it)" (6:28). Further, Kant writes,

> *Considered in themselves* natural inclinations are *good*, i.e., not reprehensible, and to want to extirpate them would not only be futile but harmful and blameworthy as well; we must rather only curb them, so that they will not wear each other out but will instead be harmonized into a whole called happiness. (*Religion*, 6:58)

Other things being equal, the fulfillment of human inclinations can be assumed to be a part of what is good for human beings, which we represent to ourselves by conceiving of it as part of what nature intends for us. We realize our radical possibility for evil only if we "reverse the moral order of [our] incentives in incorporating them into [our] maxims," by placing the satisfaction of all of our own inclinations ahead of our obedience to the moral law. "Whether the human being is good or evil must not lie in the difference between the incentives that he incorporates into his maxim (not in the material of his maxim) but in their *subordination* (in the form of the maxim): *which of the two he makes the condition of the other*" (6:36). We are not evil simply because we satisfy natural inclinations, but only if we make the satisfaction of our own inclinations the sole condition under which we will comply with the moral law rather than making the possibility of our complying with the moral law the sole condition under which we will find it permissible to satisfy our natural inclinations.

On Kant's view, then, there is one sense in which it is natural to place our own happiness before all else and to try to dignify this into a moral principle by purporting to philosophize from actual examples of human conduct, but another sense in which the existence of inclinations

the fulfillment of which would bring happiness is entirely natural and in itself a predisposition to the good rather than to evil. If this is so, then inclinations and happiness as their satisfaction cannot simply be banished from our conception of ourselves as moral agents, but must be given their proper place. This is what Kant suggests in highly abstract form by arguing in the *Religion* that being good lies not in eradicating but in subordinating natural incentives to the moral law, and expresses more concretely in his doctrine of the highest good, the exposition of which is found not only in the summation of each of Kant's three critiques but also in the Preface to the *Religion* itself[12] — a fact which cannot but suggest the absolute centrality of this doctrine for Kant. As Kant expounds this doctrine in the *Religion*, human beings cannot determine their wills to action except by the representation of some particular ends to be achieved by acting. That is, in order to act we must have something specific we intend to do, which can only be some particular action proposed as a way to fulfill some human need or inclination. Morality as "the supreme limiting condition of the freedom of action of every human being" (see *G*, 4:430-1) needs particular courses of action to limit. Or as Kant puts it in the *Religion*, while the motivation to act — respect for duty — and a formal specification of the condition on all maxims of action — the fundamental principle of morality — can be acknowledged by us independently of "the representation of an end which would have to precede the determination of the will," morality must still have "a necessary reference to

11 Of course, at the time of writing the *Religion* Kant had already defended the adoption of a *regulative* interpretation of a teleological view of nature as a single system directed to our own moral fulfillment in the *Critique of Judgment* (see especially 83-4). But the view that we should conceive of every natural faculty and disposition of our own nature as having a proper and indeed properly moral use was hardly new to the third *Critique*; it is clearly expressed in the 1784 essay "Idea for a Universal History from a Cosmopolitant Point of View," Proposition One (8:18), and in the *Groundwork* itself (4:395-6).

12 In the *Critique of Pure Reason*, in the "Canon of Pure Reason," A 804-19/B 832-47; in the *Critique of Practical Reason*, in the "Dialectic of Pure Practical Reason," especially 5:110-13; in the *Critique of Judgment*, especially in "the moral proof of the existence of God," -87, 5:447-53; and in the Preface to *Religion within the Boundaries of Mere Reason*, 6:4-6.

such an end" because without it we would be "instructed indeed as to *how* to operate but not as to the *whither*" (6:4) — that is, we wouldn't actually have anything particular to do. Particular things to do can only be suggested by nature, not by the pure rational idea of morality itself, and this means that such particular ends of action must be suggested by the various needs and inclinations that we all actually have. What morality imposes is not the eradication of such natural occasions for action, then, but "only the idea of such an object that unites within itself the formal condition of all such ends as we ought to have (duty) with everything which is conditional upon ends we have and which conforms to duty (happiness proportioned to its observance), that is, the idea of a highest good in the world" (6:5). Such a happiness proportioned to duty is not just one's own happiness pursued without regard to any constraints — that would be a goal liable to be incoherent both in itself and with the happiness of others — but is rather the conjoint satisfaction of the naturally good inclinations of oneself and others insofar as that is both licensed by and indeed also prescribed by the goal of adopting maxims that are also fit to be universal law.

Section II of the *Groundwork* thus initiates Kant's complex argument about happiness. The satisfaction of our inclinations, and thus the attainment of happiness, is a natural goal of human beings. Unfortunately, the disposition to place above all else the attainment of our own, individual happiness — or, even more precisely, the attainment of what seems to us at a given moment the means to our own individual happiness — is also a natural tendency of human beings, and one which tries to dignify itself by adopting an empiricist approach to philosophizing in order to dignify the actual conduct of human beings with an air of necessity. That tendency has to be resisted, but it cannot be resisted simply by extirpating all our natural inclinations. That would be both impossible and also incoherent, for it would leave us with no actions to undertake at all. Instead, we must combat our tendency to subordinate morality to our own happiness and to dignify this with the name of (popular) philosophy with a sounder philosophy and the proper subordination of happiness to duty that this philosophy prescribes — as we have always known. We misunderstand the conditions and requirements of our own agency both by subordinating morality to our inclinations but also by proposing to extirpate all our inclinations; we properly understand both our nature and our duty when we condition our pursuit of both our own happiness and that of others by the fundamental principle of morality, as is dictated by the concept of the highest good as the object of morality.

3) The other great mistake that we would make if we were to draw our moral principles solely from the observation of actual human conduct would be to adopt the view that human actions are always entirely and solely determined by previous actions and events, leaving us no freedom of choice when faced with a particular moral issue. Such a view of the limits of human action would damage our original disposition to morality by transforming what we so often observe, namely human behavior falling short of the demands of morality because of frailty, impurity or depravity (see *Religion*, 6:30), into a necessity of human nature, which would then lead us to cut and trim our original recognition of the stringent requirements of morality to whatever weaker principle might seem compatible with such a view of the limitations of human nature. If the actions commanded by morality seem to be "actions of which the world has perhaps so far given no example, and whose very practicability might be very much doubted by one who bases everything on experience," then on such a view "nothing can protect us against falling away completely from our ideas of duty and can preserve in our soul a well-grounded respect for its law" (*G*, 4:407-8). A revision of the principle of morality to reflect the limitations of what human beings can actually do would indeed be the only reasonable response to such limitations, on the principle of rationality that Kant always assumes we all share, that "duty commands nothing but what we can do" (*Religion*, 6:47).[13] On this prin-

13 Kant repeatedly asserts the principle that we must be *able* to do what we *ought* to do in the *Religion*; e.g., 6:62, 63.

ciple, if we cannot do an action, then the principle of morality cannot command it, so the principle of morality must reflect what we can do.

Kant clearly must limit the damage that could be done to morality by the all too common examples of human frailty and the philosophy of determinism that dignifies such examples with the air of necessity. But, as in the case of happiness, he cannot deal with the threat of frailty and its philosophical expression in the doctrine of determinism simply by "extirpating" or *refuting* this doctrine. For determinism is the keystone of Kant's own theoretical philosophy: the condition of the possibility of understanding nature, and of understanding ourselves as creatures in nature, is nothing less than the universal validity of the principle that every event in nature is determined to occur when it does in accordance with a law linking it to a prior occurrence that necessitates what follows. Determinism is not merely a natural attitude for us, but the condition of the possibility of an understanding of nature itself. Thus, just as Kant's normative moral philosophy must deal with the natural interest in happiness not by eradicating it but by assigning it its proper place in the complete object of morality, so his account of the conditions of the practicability of the stringent principle of morality must still find a place for a doctrine of determinism. In Kant's own words, "Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the very same human actions, for it cannot give up the concept of nature any more than that of freedom" (*G*, 4:456).

It is hardly necessary here to go into the details of Kant's way of assuring that there is "no true contradiction" between "freedom and natural necessity": every reader will know that Kant argues that determinism is a necessary condition of assigning a determinate order to events as they occur *in time*, but that since time itself is a feature only of the *appearance* of things, not of those things as they are *in themselves*, it is entirely possible that the real agents of our actions are not situated in time at all, and therefore are not subject to determinism, and so are instead free to act as morality requires regardless of what past experience might predict. Nor do I here want to canvas the well-worn objections to this reconciliation

of freedom and determinism. What I do want to emphasize is that it is part of Kant's view of our own self-understanding, thus of what must be reflected by a proper philosophy, that certainty of our freedom is just as readily and naturally *accessible* to every normal human being as confidence in determinism is: the "rightful claim to freedom of will" is "made even by common human reason" (*G*, 4:457). The assignment of determinism into its proper place in the more complex doctrine of transcendental idealism is not merely the speculative replacement of unsound philosophy by sound philosophy; it is, in Kant's view, itself the proper expression of ordinary human self-understanding.

It might not seem surprising to say this about Kant's defense of freedom in the *Critique of Practical Reason*, where Kant argues precisely that everyone immediately infers his freedom to act as the moral law requires directly from "the *moral law*, of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves)" (5:29) On this account, "*practical reason*," starting from an indubitable consciousness of what the moral law demands of us, infers our freedom always to do what the law demands by the principle that ought implies can, and then imposes the fact of freedom on "speculative reason," which has as it were no choice of its own but to secure (if not explain) at least the possibility of freedom (5:30). But, at least on one standard interpretation, the *Groundwork* reconciles freedom and determinism by a more theoretical or speculative route than the *Critique of Practical Reason*: the *Groundwork* argues that the distinction between appearances and things in themselves is one that is introduced in theorizing about the nature of knowledge, and then carried over to reflection on practical reason, where it can directly establish the fact of our freedom from which in turn the validity of the moral law can be inferred (*G*, 4:451-3). But theorizing about the conditions of the possibility of knowledge can easily look like the furthest thing from an activity of "common human reason," and thus it might well seem surprising to claim that Kant's defense of freedom in the *Groundwork* is intended to be a proper expression of ordinary human self-understanding.

But even in the *Groundwork* Kant claims that "no subtle reflection" is required to make the distinction between appearances and things in themselves, rather "one may assume that the commonest understanding can make it, though in its own way, by an obscure discrimination of judgment which it calls feeling." Even this commonest human understanding, Kant alleges, is aware of the difference "between representations given us from somewhere else and in which we are passive, and those that we produce simply from ourselves and in which we show our activity"; and this is enough to "yield a distinction, although a crude one, between a *world of sense* and the *world of understanding*," a distinction which will in turn allow anyone to conceive of the difference between the appearance of objects and their states that are fully governed by deterministic laws of nature and the spontaneous actions of things as they are in themselves that can only be governed by laws of reason rather than sensibility (*G*, 4:450-1). The *Critique of Practical Reason* may infer the fact of our freedom from our prior acknowledgment of our obligation under the moral law, while the *Groundwork* may infer our obligation under the moral law from the fact of our freedom, which is in turn inferred from the basic structure of human cognition, but the intended epistemology of both arguments is the same: each argument assumes that what it characterizes as the sufficient ground for knowledge of our freedom is just as available to every human being, just as much a part of our self-understanding, as is the basis for the belief in determinism. In both arguments, Kant's philosophical reconciliation of freedom and determinism is supposed to be the expression of common human self-understanding.

This result leads to one last conclusion, which can tie together Kant's apparently optimistic moral writings of 1785 and 1788 with the apparently pessimistic *Religion* of 1793. If transcendental idealism with its reconciliation of freedom and determinism is really the proper expression of ordinary human self-understanding, then the belief in the philosophical doctrine of determinism could not possibly be due to an academic philosophical misunderstanding alone, any more than the elevation of one's own happiness into the unrestricted principle of morality could

be the product of a merely speculative misunderstanding alone: the sounder philosophy which reconciles freedom and determinism, just like the sounder philosophy that subordinates but at the same time incorporates happiness into the complete object of morality, Kant has insisted, is just as available to common human reason as the one-sided philosophies are. Instead, the adoption of the one-sided "worldly wisdom" that would undercut our recognition that happiness is not the sole object of morality and human frailty not an excuse for trimming the demands of morality could only be the *product* or *expression* of the human possibility to be evil instead of good, not the *cause* of this evil. If the proper understanding of our own agency is always available to us, then misunderstanding the possibilities of our agency cannot simply be imposed upon us, but must be self-imposed. We cannot blame philosophy for our own failings, Kant must hold, for the philosophy that can save us from these failings is always already available to us.