

Reason and autonomy in *Grundlegung* III

Much of the difficulty of Kant's *Grundlegung* arises from shifts of framework within each chapter. The third chapter shifts *from* a metaphysic of morals *to* a critique of pure practical reason. This complicated shift is hardly discussed in the text, yet it is the crux of the chapter. In this essay I draw on other Kantian texts, in particular on the *Critique of Pure Reason*, to elucidate the point and outcome of the shift of framework. Much that I shall say is tentative: a proposal for reading the chapter rather than detailed commentary. I shall try to make it definite enough for criticism to have points of focus.

In the preface to *Grundlegung* we are told that "a metaphysic of morals has to investigate the Idea and principles of a possible *pure* will" rather than "the activities and conditions of human willing as such" (*G*, IV, 390). For much of Chapter II and in Chapter III up to page 450, Kant stays within the framework of a metaphysic of morals. He presents the Categorical Imperative and analyzes the connection between freedom and morality in abstraction from claims about actual rational beings.¹ Even if it convinces, we will still be unsure about human rationality, human freedom or human duties. No doubt we will share the common rational understanding of morality of Chapter I; but Kant will not have vindicated this common understanding. The shift to a metaphysic of morals in Chapter II may turn out to be a blind alley for human beings. A metaphysic of morals may lack application to human beings. To show that the discussion of Chapter II and of the first part of Chapter III is relevant to *us*, an entirely different framework is needed.

The brief prefatory comments on critique of practical reason offer a tantalizing clue. Kant writes that a complete critique of practical reason "should be able at the same time to show the unity of practical and theoretical reason in a common principle, since in the end there can only be one and the same reason" (*G*, IV, 391). A successful critique of practical reason should apparently show that there is a single supreme principle for practical and for theoretical reason. The *Grundlegung* is clear

This chapter originally appeared in *Grundlegung zur Metaphysik der Sitten: Ein kooperative Kommentar*, ed. Otfried Höffe, Vittorio Klostermann, Frankfurt, 1989.

1 The examples of duties are an exception. They evidently mention specifically human possibilities and institutions. This need not trouble us if we think that the examples are illustrative. Illustrations are commonly determinate in ways that the texts they illustrate are not.

I. REASON AND CRITIQUE

enough that the supreme principle of practical reason is the Categorical Imperative. Could the Categorical Imperative be the supreme principle of all reason? Does this thought even make sense?

I shall propose a reading of *Grundlegung* III that presents the Categorical Imperative as the supreme principle of all reason. It has the corollary that freedom and autonomy are at the heart not just of morality but of all reasoning. This reading will, I hope, help to show what can and what cannot be done to vindicate reason, and how reason and autonomy are connected. The claim that the Categorical Imperative is the supreme principle of human reason, I shall argue, both offers a coherent view of Kant's larger enterprise and sheds some light on the last chapter of *Grundlegung*.

The agenda is simple, if a bit presumptuous. First I shall recapitulate the conditional claims of the early part of Chapter III, and note what remains to be established if these conditionals are to be relevant to us. Then I shall rehearse Kant's account of the connections between critique, autonomy and the authority of reason, drawing largely on the *Critique of Pure Reason*. These considerations will be used to throw light on the critical passages of Chapter III on the two standpoints (*G*, IV, 450 onward). The initial claims about freedom, autonomy and morality are reconsidered in the light not of the preceding account of morality but of Kant's conception of critique and of the authority of reason. To present this line of thought compactly I have avoided all but the most glancing engagement either with the *Critique of Practical Reason* or with the secondary literature. The result is less a commentary than a proposal for organizing a commentary.

The gap between rationality and freedom

The analysis of freedom at the start of the third chapter makes only conditional claims. Kant characterizes will as "a kind of causality belonging to living beings so far as they are rational" (*G*, IV, 446). This definition leaves it open whether the human will is free and how freedom may be connected to rationality. Kant distinguishes causality that works by natural necessity or the influence of "alien causes" from causality that works by freedom.² Freedom is initially defined negatively as the property by which a will is "able to work independently of *determination* by alien causes" (*G*, IV, 446). The distinction between natural necessity and freedom is clearly drawn; but until we understand what an "alien" cause is, the criteria for its application are unclear. In particular Kant offers no argument to

2 It is important not to attribute to Kant the claim that we can distinguish events that are naturally caused from others that are not. He holds that all events are part of some objective sequences and can be given a naturalistic explanation. His view is that naturalistic explanation cannot be the whole story – not that there are some events about which it has nothing to say. See Chapter 4 in this volume.

3. REASON AND AUTONOMY IN *GRUNDLEGUNG* III

show that human beings, or other supposedly rational beings, have wills that can work independently of alien causes.

The second move of the chapter connects the “negative” with a “positive” conception of freedom (*G*, IV, 446). Positive freedom is more than independence from alien causes. It would be absent in lawless or random changes, although these are negatively free, since they depend on no alien causes. Since will is a mode of causality it cannot, if free at all, be merely negative free, so it must work by nonalien causality. Free will (if it occurs) must be (capable of being) positively as well as negatively free; it must be a capacity for *self-determination* or *autonomy* (*G*, IV, 447). Once again the distinctions are well drawn, but the criteria for their application are unfixed.

The third move of the chapter identifies the positive conception of freedom – self-determination or autonomy – with conformity to the Categorical Imperative (*G*, IV, 447). Kant’s reasoning is compressed, indeed sketchy; but the thought, at least in its moral interpretation, is familiar. It was presented in more explicit form in Chapter II, in the discussion of the Formula of Autonomy, which runs, “Never to choose except in such a way that in the same volition the maxims of your action are also present as universal law” (*G*, IV, 440).³ The thought is roughly that *if* there are beings who can choose freely, in the sense that their choosing can be lawlike yet not determined by alien causes, they must be capable of imposing lawlikeness on their actions, that is, of acting on universalizable maxims. However, the crucial connections between rationality and autonomy are not made explicit at the beginning of Chapter III. Kant assumes a background understanding that the supreme principle of practical reason links the requirement of autonomy to an ability to choose universalizable maxims.

Kant views the initial claims of Chapter III as “mere analysis of the concept of freedom” (*G*, IV, 447). All that he claims to have shown is that *if* rational beings are free, they are not simply negatively free, but (capable of) autonomy, and that *if* (capable of) autonomy, they are subject to the principle of morality, that is, to the supreme principle of practical reason. These supposedly analytic claims can seem deeply unconvincing. Why is not conformity to reason, even to a supposed supreme principle of practical reason, just another mode of heteronomy? Why does Kant take conformity to desires for heteronomy, and conformity to reason for autonomy? Why are desires but not reasons to be seen as “alien”? Does not calling the supreme principle of practical reason an “imperative” suggest that reason too is alien? Autonomy in many modern accounts is construed empirically as action on reflectively endorsed desires, or as avoiding specific sorts of social and personal

3 The treatment of autonomy in Chapter II is, however, incomplete: Kant states at *G*, IV, 440, that proof that autonomy is the sole basis of ethics needs a critique of practical reason and “does not belong in the present chapter”.

I. REASON AND CRITIQUE

dependence.⁴ What reason is there to prefer the stricter and stranger Kantian conception of autonomy?

These questions are not answered immediately. We can, however, discern aspects of an answer. Most pertinently we can begin to see why, if autonomous action is to be independent of everything “alien”, it must be action determined by reason. Independence from “alien” causes, taken strictly, must be independence from contingent or variable events, including those that are intimately part of agents. Such events depend on the play of natural forces or (possibly) on random occurrences, and so ultimately on indisputably “alien” causes. Maxims of autonomous action, by contrast, must hold equally for all rational agents, whatever their peculiar contingent and variable characteristics, and so must be universalizable.⁵ Their authority, if they have any, cannot derive from any contingency of human life but only from the requirements of reason, whatever those may be.

From these claims, however, we can draw only a conditional conclusion about reason: If reason has any authority, that authority will hold indifferently for all rational beings. We do not learn *what* the principle(s) of reason are or *why* they have authority. Equally, we can draw only a conditional conclusion about human rationality and freedom: *If* we are rational in the required sense, then we are also free and so capable of autonomy and bound by morality. If we are not rational in the required sense, but only in some other (e.g., purely instrumental) sense,⁶ then there will be a gap between our rationality and our freedom, and the Kantian conception of autonomy will be irrelevant to us.

Kant is well aware that the gap between rationality and freedom has not been closed. We have still, he points out, “to prove that freedom too is a property of the will of all rational beings” (*G*, IV, 447). Unless and until this can be done he can only *assert* or *maintain* but not demonstrate or ground human freedom. Kant uses the first person in two striking and uncharacteristic assertions⁷ to show that these claims are ones that he cannot vindicate (at least at this stage): “Now I assert [*Ich sage nun . . .*] that every being who cannot act except *under the Idea of freedom* is by this alone – from a practical point of view – really free” (*G*, IV, 448). The emphasis is doubled: “And I maintain [*Nun behaupte ich . . .*] that to every rational being possessed of a will we must also lend the idea of freedom as the only

4 This view can be traced back at least to J. S. Mill. It has recently been refined by Harry Frankfurt, “Freedom of the Will and the Concept of a Person”, since when it has become standard fare in English-language writing on autonomy.

5 This is not to say that they demand *uniformity* of action; *universalizable* principles demand uniform action only for like cases. Conditional universalizable principles demand action only of those who meet the condition.

6 In *Grundlegung* I Kant argued that there cannot be beings with merely instrumental rationality (*G*, IV, 396). This argument is conducted within the framework of common rational knowledge of morality. Its conclusion remains in doubt since that framework is later shown to be insufficient.

7 Cf. *G*, IV, 428: “Now I say [*Nun sage ich . . .*] that man, and in general every rational being, *exists as an end in himself . . .*” This too is a moment where Kant needs to anticipate the results of a critique of practical reason.

3. REASON AND AUTONOMY IN GRUNDLEGUNG III

one under which he can act" (G, IV, 448). Assertion is not argument; but it may be enough for practical purposes. If indeed we can act and judge only "under the Idea of freedom", we must insist upon looking at ourselves as free whether or not we can show we are. In particular we must look at our reasoning itself as autonomous: "Reason must look upon itself as the author of its own principles independently of alien influences. Therefore as practical reason, or as the will of a rational being, it must be regarded by itself as free" (G, IV 448).

However, the practical indispensability of viewing ourselves as free is no proof of human freedom. Could not the fact that we must think we are free when we act and judge be a profound illusion? Many have thought that it is. Kant reminds us that "alleged experiences of human nature" (G, IV, 447) cannot prove human freedom. So long as the connection between reason and freedom is not established, the links between freedom, autonomy and morality are merely a series of analyses and the moral skeptic need not be shaken. The skeptic may hold that reason is inert, available for theoretical purposes but a mere slave of the passions in practical matters. Kant is adamant about the limitations of his analytic arguments. Freedom has not been demonstrated: "we have been quite unable to demonstrate freedom as something actual in ourselves and in human nature" (G, IV, 448). Nor has morality. We have no insight into "*how the moral law can be binding*" (G, IV, 450). For all we know, our inability to act or judge except under the Idea of freedom reflects some quirk or illusion of human nature and nothing fundamental about rational beings as such. The analysis had led us into "a kind of circle, from which, as it seems, there is no way of escape" (G, IV, 450). Since analysis traps us in a circle, no move is possible without a radical shift of argumentative strategy.

Critique, autonomy and the authority of reason

It is only in the passages after page 450 that Kant argues that human beings are free agents, in a sense that makes their actions imputable and the Categorical Imperative relevant to them. Neither his articulation of our common moral consciousness in Chapter I, nor the abstract analysis of principles of obligation for rational beings that a metaphysics of morals offers, has shown, *or could have shown*, that *we* are agents of the required sort. Yet unless we are free agents in the required sense the entire moral theory will lack practical significance for us. To go further we must enter the different framework of a critique of practical reason.

A critique of practical reason could take us no further than a metaphysics of morals if it were simply a more abstract way of arguing. However, this is not the difference between them. Rather it is a matter of starting point. A critique of (practical) reason reconsiders the *standpoint* from which argument is conducted. Kant writes that "One shift [*Auskunft* – perhaps "departure" or "way out"], how-

I. REASON AND CRITIQUE

ever, still remains open to us" (G, IV, 450). The only way out of the circle is to "enquire whether we do not take one standpoint when by means of freedom we conceive ourselves as causes acting *a priori*, and another standpoint when we contemplate ourselves with reference to our actions as effects which we see before our eyes" (G, IV, 450). The critical turn considers how we can shift our very starting points. Only by making such a shift can we exit from the circle of analysis.

Here we see immediately the connection to the fundamental movement of the *Critique of Pure Reason*, the Copernican turn, which Kant also characterizes as a matter of viewing things from a double point of view (CPR, Bxix[n]). To undertake a critique (whether of pure reason, or specifically of pure practical reason) is not to attain an unconditional transcendent vantage point, but to take a critical view of the starting point, or condition, of one's own previous thinking. Critique rests on the ability to shift starting points, and so to gain some distance and independence from a given starting point.

In the *Critique of Pure Reason* Kant sees criticism of reason as undertaken from the standpoint of those who find themselves apparently able to reason, yet led into antinomies in their reasoning, from which they are able to escape only by disciplining their reasoning. From the start Kant insists that critique of reason must be a reflexive process: "reason has insight only into that which it produces after a plan of its own" (CPR, Bxiii). The plan that he proposes is to shift perspective: "We must therefore make trial [*Versuch*] whether we may not have more success in the tasks of metaphysics if we suppose that objects must conform to our knowledge" (CPR, Bxvi). The entire "Doctrine of Elements" of the first *Critique* is offered as the first part of such a "trial" or experiment. It offers no account of reason's authority. The whole familiar dissection of sensibility and understanding is intrinsically incomplete. Although we may receive a manifold of intuition that we synthesize under empirical concepts and judge according to the categories of the understanding, these acts of judgment rely on as yet unvindicated uses of reason. After seven hundred pages we have only "an inventory of materials" [*Bauzeug*; elsewhere *Stoff*] (CPR, A707/B735). It is the task of the "Doctrine of Method" to move toward an account of the authority of reason. The next part of the "trial" must be to see what can be said in vindication of reason.

In the "Doctrine of Method" Kant offers a recursive account of the authority of reason. In a post-Copernican world, with no absolute *up* or *down*, we cannot have a foundationalist account of the grounds of reason (whatever that might be). A critical grounding of reason works by showing what it takes to orient thinking and acting by principles that do not fail even if used universally and reflexively. The key idea behind the notion of a critique of pure reason is that we can find standards of reasoning by considering how we can and must discipline our thinking. However, this discipline is not to be thought of as externally imposed: Rather it is *self-discipline* or *autonomy* in thinking.

3. REASON AND AUTONOMY IN *GRUNDLEGENG* III

These points have a firm textual basis. The “Doctrine of Method” initially characterizes discipline as “the compulsion by which the constant tendency to disobey certain rules is restrained and finally extirpated” (*CPR*, A709/B737). Kant notes: “that reason, whose proper duty it is to prescribe a discipline for all other endeavours, should itself stand in need of such a discipline may indeed seem strange” (*CPR*, A710/B738). Yet this strange requirement is the key to Kant’s account of the authority of reason. Our thinking needs “a discipline, to restrain its tendency towards extension beyond the narrow limits of possible experience and to guard it against extravagance and error” (*CPR*, A711/B740), and “the whole philosophy of pure reason has no other than this strictly negative utility” (*CPR*, A711/B740). Reason, the discipline of all disciplines, can only be and must be *self-disciplined*: The subordination of thinking or practice to other supposed authorities (state, church, experts, personal preferences) is not reason, but the abrogation of reason.⁸ Reason’s discipline cannot be alien; it must be autonomous.

The task of closing the gap between reason and autonomy (which the analytic passages left open) can now be seen in a new light. This gap is to be closed not by establishing that human beings are rational, and then proving that they are free, and hence also (given the analytic argument) autonomous. Kant’s strategy is the reverse. He argues not from reason to autonomy but from autonomy to reason. Only autonomous, self-disciplining beings can act on principles that we have grounds to call principles of reason. Reason has no transcendent authority; it can only be vindicated by critique, and critique itself is at bottom no more than the practice of autonomy in thinking. Autonomy does not presuppose but rather constitutes the principles of reason and their authority:

Reason must in all its undertakings subject itself to criticism; should it limit freedom of criticism by any prohibitions, it must harm itself, drawing upon itself a damaging suspicion. Nothing is so important through its usefulness, nothing so sacred, that it may be exempted from this searching examination, which knows no respect for persons. Reason depends on this freedom for its very existence. For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objections or even his veto. (*CPR*, A738/B766)

Because the authority of reason can only be established by autonomous self-discipline in thought and action, attempts to lean on other authorities backfire:

it is indeed absurd to look to reason for enlightenment, and yet to prescribe beforehand which side she must necessarily favour. Besides, reason is already of itself so confined and held within limits by reason, that we have no need to call out the guard, with a view to bringing the civil power to bear upon that party whose alarming superiority may seem to

8 These passages throw light on Kant’s insistence that the kernel of Enlightenment is not reason but autonomy. Cf. *WE*, VIII, 33; and Onora O’Neill, “Enlightenment as Autonomy: Kant’s Vindication of Reason”.

I. REASON AND CRITIQUE

us to be dangerous. In this dialectic no victory is gained that need give us cause for anxiety. (CPR, A747/B775)

A large part of the “Doctrine of Method” is rebuttal of supposed alternative accounts of the authority of reason. The rationalist conception of reason as mathematical method is doubly wrong: Philosophy cannot be done *more geometrico*, nor mathematics *more analytico*. Equally, an eristic conception of reason as intrinsically polemical cannot be vindicated. The only hope, if we are to discern principles that have authority either for thinking or for acting, is a discipline, a “negative instruction” that rejects as pseudorational principles that make thinking conditional on some “alien”, rationally ungrounded authority. Such principles have authority only on condition that the “alien” authority on which they depend is accepted. The only unconditional principles that have authority are those that appeal to nothing except disciplined autonomy in thinking. Criticism, which has destroyed the authority of traditions, can sustain the authority of reason not by *polemic* with supposed rival authorities (churches, civil powers, etc.), let alone by dependence on discredited authorities, but only by relying on principles that demand no more than that criticism be conducted in and survive their own terms. Such principles have no transcendent authority; we grant them provisional authority and rely on them because they survive all self-criticism and depend on nothing alien. The authority of reason differs from that of civil and other powers, whose antagonisms are destructive. Reason is self-disciplining, but not self-destroying.

In *What Is Orientation in Thinking?* Kant articulates the fundamental strategy of autonomy in thinking. He begins with the claim that there is no alien, transcendent authority in thinking. The choice is between autonomy and lawlessness, reason and madness: “Freedom in thinking means the subjection of reason under no other laws than those it gives itself. Its opposite is the maxim of a lawless use of reason” (WOT, VIII, 145; 303–4). But lawlessness ends in servitude: “if reason will not subject itself to the law it gives itself, it will have to bow under the yoke of laws which others impose . . .” (WOT, VIII, 145; 303–4). These dilemmas are dramatic; but they do not show anything about the “law” that reason can give itself. Which principles have unconditional authority in thinking and acting? Where could self-discipline either in thinking or in acting begin? At least one negative point can be made. One corollary of refusal to bow under an alien yoke is that what count as the principles of reason cannot hinge on variable and contingent matters, all of which, however intimately human, are alien causes. Would-be reasoners must then at least adopt the strategy of not acting on principles that accept such alien authorities. They must act only on maxims on which others whose contingent circumstances and characters differ can also act.

In adopting this merely negative strategy for guiding thought and action, reasoners – or potential reasoners – adopt the principle that we know in other contexts as the Categorical Imperative. The Categorical Imperative is the supreme

3. REASON AND AUTONOMY IN *GRUNDLEUNG* III

principle of reasoning not because it is an algorithm either for thought or for action,⁹ but because it is an indispensable strategy for disciplining thinking or action in ways that are not contingent on specific and variable circumstances. The Categorical Imperative is a fundamental strategy, not an algorithm; it is the fundamental strategy not just of morality but of all activity that counts as reasoned. The supreme principle of reason is merely the principle of thinking and acting on principles that can (not “do”!) hold for all. Hence to reason is “to ask oneself, with regard to everything that is to be assumed, whether he finds it practicable to make the ground of the assumption . . . a universal principle of the use of his reason” (*WOT*, VIII, 146n; 305n). Both in thinking and in acting the self-discipline of reason is a matter of asking whether the ground of the assumption can be a universal principle. The supreme principle of reason both emerges from and disciplines human thought, action and communication. There is no gap between reason and autonomy because the authority of reason is grounded in autonomy.

Critique and the two standpoints

This account of the connections between critique, autonomy and reason provides a background for understanding the shift to a critique of practical reason in *Grundlegung* III. After page 450 Kant adopts a critical approach to the enterprise of specifically practical reasoning. He focuses no longer on the analytic connections between concepts but on the distinct starting points that are open to us. He begins with the reminder that empirical knowledge, both of the world and of ourselves, is knowledge of appearances and is incomplete. Appearances cannot be everything. Hence we not only may but must form the conception of an intelligible, that is, nonempirical, “world”. However, he also reiterates that this intelligible world cannot be an object of knowledge (see also *CPR*, A255/B311). He apparently repudiates both the thought that such a nonempirical world is illusory and the thought that it is a distinct, supernatural reality. He eschews both empirical idealism and transcendental realism, and speaks rather of the two standpoints from which a rational being can regard himself: “*first* – so far as he belongs to the sensible world – to be under laws of nature (heteronomy); and *secondly* – so far as he belongs to the intelligible world – to be under laws which, being independent of nature, are not empirical but have their ground in reason alone” (*G*, IV, 452). Kant’s claim is that human agents (and presumably any other

9 But surely, it may be objected, the principles of reason should be algorithms. They should include – perhaps consist of – elementary rules of inference and algorithms of practical reason, if there are any. This objection forgets that algorithms are not self-deploying. When we reconsider rules of inference we are considering an aspect of reasoning that can be abstracted; but abstracted rules of inference and decision procedures alone cannot guide a single thought or act. The authority of algorithms presupposes rather than provides the authority of reason.

I. REASON AND CRITIQUE

finite rational agents) not only *may* but *must* adopt both standpoints, and must shift between them: “when we think of ourselves as free, we transfer ourselves into the intelligible world as members . . . when we think of ourselves as under obligation, we look upon ourselves as belonging to the sensible world and yet to the intelligible world at the same time” (G, IV, 453). Neither standpoint is eliminable; the standpoint of the intelligible world is indeed required to grasp that of the sensible world: “*the intelligible world contains the ground of the sensible world and therefore also of its laws . . .*” (G, IV, 454). Kant insists that, despite the apparent contradiction, the two standpoints are compatible: Nature and freedom are equally necessary and “not merely *can* get on perfectly well together, but must be conceived as *necessarily combined* in the same subject” (G, IV, 456).

These passages are highly controversial. Kant himself concedes that reason is in a tight corner. One tempting reading suggests that in talking of “intelligible” and “sensible” worlds Kant assumes the very transcendent metaphysics that he otherwise repudiates, and so lapses into rationalism. Yet if his views of self-knowledge undercut the possibility of knowledge of an ontologically distinct intelligible world, how can he offer any two-worlds story? Surely such an account can be neither true nor knowable from a Kantian perspective, and in any case offers an ontologically extravagant and psychologically implausible account of atemporal agency. There is much to be said against such readings both on textual and on wider grounds. Although the metaphors of the intelligible and the sensible world continually invite an ontological interpretation that fits their Platonic and Leibnizian heritage, Kant repeatedly distances himself from such interpretations, and insists that he hinges his argument on *two standpoints* and not *two worlds*. He puts the matter unambiguously: “The concept of the intelligible world is thus only a *point of view* which reason finds itself constrained to adopt outside appearances *in order to conceive itself as practical*” (G, IV, 458; cf. 462 and CPR, A5/B9, A255/B310ff.). Indeed, it is not only the intelligible world that cannot be a transcendent reality. The core of transcendental idealism is the denial that the sensible world either is or reveals a transcendent reality. Our understanding of Kant’s use of the term *world* must be firmly restrained if we are not to slide into transcendent realism. Nature is “the sum of all appearances” (CPR, B163), and “appearances do not exist in themselves” (CPR, B164).

Yet there are also strong reasons to interpret the term *world* as referring to an ontologically distinct realm. For how else are we to make sense of the crucial notion of a standpoint? A standpoint, surely, is a position in some space. Without the space there will be no standpoints. How can the intelligible world be *only* a point of view?

It will not be adequate to read Kant as holding that the two standpoints, far from being or depending on ontologically distinct worlds, are mere illusion, so that the duality of standpoints reflects a merely subjective necessity – a psychological compulsion. This approach would abandon all hope of making sense of the

3. REASON AND AUTONOMY IN *GRUNDLEUNG* III

critical portions of *Grundlegung*. Here and elsewhere Kant repudiates empirical idealism as firmly as he repudiates transcendent realism.

A careful reading of the critical sections of *Grundlegung* III suggests a way between the view that the intelligible and sensible standpoints are transcendent realities and the view that either is illusion. Kant uses the vocabularies of “sensible world” and “intelligible world” and of the naturalistic and practical standpoints more or less interchangeably. But it is the latter vocabulary that is used to anchor the former. Rather than taking the two standpoints as standpoints within two worlds, hence needing an independent account of those two worlds, we are to take the two worlds as differentiated by the inability of human reason to form a comprehensive picture of its objects. From the first sentence of the first *Critique* we are warned of the predicament of a reason that aspires to tasks that it cannot achieve: Reason’s failure is that it cannot give a unified account of nature and freedom. *The metaphor of the intelligible world signals the finitude, not the transcendence, of human reason.*

Kant in fact argues for three points. He claims that both standpoints are indispensable; that the theoretical standpoint, and the knowledge of nature it permits, are unavailable without the practical; and finally that the two are compatible.

Why does Kant hold that neither way of viewing ourselves is dispensable? (Many dualities in human self-consciousness are quite dispensable.) Why, in particular, does he not think (with many others) that the naturalistic standpoint is the only important one? We are perhaps not inclined to reconsider the status of the naturalistic standpoint when we read Chapter III of the *Grundlegung*. Since the initial, analytical passage of the chapter proposes interesting conditional connections between freedom, autonomy and morality, we are (quite reasonably) eager to discover whether the antecedents of those conditionals are true. We want to know whether we human beings are free agents with the capacity for autonomy, so that morality with all its implications binds us. In reading the third chapter we readily take for granted that we must look upon ourselves as belonging to the sensible world, and assume that we must vindicate freedom in theoretical terms. A return to the perspective of the first *Critique* can be revealing at this juncture because the *naturalistic* standpoint both on ourselves and more generally is in question there. Neither the status nor the scope of causal explanation is established at the start of the first *Critique*.

The naturalistic standpoint is in question in the first *Critique* because it leads to antinomies when extended indefinitely. When we try to treat it as a comprehensive viewpoint, it proves self-undermining. We cannot form a coherent account of a spatiotemporal totality that is causally ordered. Although the empirical use of reason must strive for completeness in naturalistic understanding, the completeness is only a regulative ideal, and not attainable (CPR, A565/B593). A naturalistic account of the world lacks closure. Causal explanations of the natural world work only on condition that naturalism and causal explanations are not the whole

I. REASON AND CRITIQUE

story. The only way to adopt the naturalistic standpoint without being led into antinomies is to acknowledge the restricted scope of naturalistic explanation, which provides the framework for understanding the world as it appears to us, but cannot explain that the world should so appear. Hence naturalistic explanations cannot cover things as they are in themselves; if they could, the antinomies would be fatal to the coherence of naturalistic explanation. Consideration of the antinomies and the consequently restricted scope of naturalistic understanding suggests why Kant should have held that the ultimate grounds of “the sensible world” must lie beyond that “world”. However, we have no intellectual grasp of things as they are in abstraction from the naturalistic standpoint. The “intelligible” grounds of nature elude our theoretical inquiries.

These considerations may show why Kant thought that those who view matters from a naturalistic standpoint must acknowledge some other standpoint. But do they vindicate human freedom and agency? Have we grounds for asserting practical as well as transcendental freedom? It may be the case that (as a matter of anthropological fact) we conceive ourselves from two standpoints and that (as a matter of necessity) the naturalistic standpoint provides an incomplete view. If so, why should we think a specifically practical standpoint indispensable? Why should we “transfer ourselves to the intelligible world” or think of *ourselves* as able to act in independence of alien causes? Why should not the duality of standpoints be a merely anthropological phenomenon, unconnected with the deeper reasons why our theoretical knowledge is incomplete? Why should we understand *ourselves* as part of the “intelligible world” that naturalistic understanding presupposes? Might not that “intelligible world” be not merely inaccessible to human cognition, but also irrelevant to human action, for which (despite our fragmented self-understanding) we can best account in naturalistic terms?

Once again it may be helpful to look at the claims of *Grundlegung* III in the light of the first *Critique*, where claims to know that nature is causally ordered are in question. The argument of the second analogy may provide a key to the claim that the practical standpoint is no illusion. Kant argues that the basis for claiming that events are causally ordered – that they have a necessary order – is that we can distinguish merely subjective from objective sequences of perceptions. This distinction cannot be based on any direct apprehension of objective time, which we lack. Rather, the distinction between objective sequences (which are causally explicable) and other sequences (which are not) must be drawn *within* experience. The elements of merely subjective sequences could have been perceived in another order: In Kant’s example a sequence of house perceptions might be perceived in various orders. In objective sequences the order of perception is (in certain respects) outside the control of the perceiver: In Kant’s example a sequence of perceptions of a moving ship cannot be (wholly) rearranged by the perceiver (*CPR*, A192–3/B237–8). We distinguish a sequence of perceptions as objective and open to causal explanation when we cannot vary their order. We understand

3. REASON AND AUTONOMY IN *GRUNDLEGUNG* III

what it means for the order of perceptions to lie beyond our control when we contrast cases where we have (some) control over the sequence of perceptions. If we could not make this contrast, we could not distinguish empirical reality and fantasy. In a fantasy world all sequences are dreamlike; we cannot distinguish self from world, or our own doings from what happens; *and the basis for naturalistic explanations of an objective world is lost*. When succession is objective, by contrast, “I cannot arrange the apprehension otherwise than in this very succession” (*CPR*, A193/B238); and we realize that “this compulsion is really what first makes possible the representation of a succession in an object” (*CPR*, A197/B242).

The second analogy, in short, argues that the grounds of objective sequence, hence of naturalistic explanation, lie in the possibility of distinguishing what we control from what is beyond our control. Agency is here taken to be the presupposition of causal judgment. The second analogy invokes practical freedom – not the mere spontaneity of understanding, which is revealed *within* theoretical knowledge – to account for the possibility of causal explanation.¹⁰ It follows that if the theoretical standpoint is indispensable to us, the practical standpoint must also be so. This argument from the second analogy provides the additional move that explains why Kant maintains in *Grundlegung* III not only that the sensible *world* has its ground in the intelligible *world*, but specifically that *we* cannot think of ourselves merely as members of the sensible world. The reason that we must “transfer” *ourselves* to the intelligible world is that if we are not agents, we will never have reasons to think of ourselves as confronting a natural world that is causally determined, and so resists our control. If we take the “speculative standpoint” for granted, we will not see that if we are to have empirical or scientific knowledge, we not merely may but must be free agents. For we will fail to note that the naturalistic view is available only to those who are free and capable of autonomy.

This is why Kant concludes in the critical sections of *Grundlegung* III that although no (theoretical, naturalistic) *explanation* of freedom is possible, still “to argue freedom away is as impossible for the most abstruse philosophy as it is for the most ordinary human reason” (*G*, IV, 456). It is impossible because freedom is presupposed by theoretical understanding.

Morality, autonomy and reason

Grundlegung III ends at “the extreme limit of all moral enquiry” (*G*, IV, 462). Kant contends that if we press on with a critical consideration of our very starting

10 Cf. *G*, IV, 452, where Kant stresses the distinction between spontaneity of understanding and the higher spontaneity of reason. The former produces only concepts “whose sole service is to bring sensuous ideas under rules”; this spontaneity is shown in the indeterminacy of judgment. The latter “manifests its highest function in distinguishing the sensible and intelligible worlds . . . and so in marking out limits for understanding itself”. The argument of the second analogy spells out some limits of understanding; it shows that practical freedom is presupposed by cognition within which spontaneity of understanding is possible.

I. REASON AND CRITIQUE

points, we find that freedom (and so, by the analytic argument of the beginning of Chapter III), autonomy and morality, which are fundamental to the practical standpoint, are also presupposed by theoretical understanding. This does not constitute an unconditional argument that the practical standpoint is indispensable; but it is a powerful consideration for those who accept the theoretical standpoint.

Critique of reason leads us face to face with the conditional character of all reason: We can demonstrate neither a transcendentally real world nor the actuality of human freedom: "it is an . . . essential *limitation* of the same reason that it cannot have insight into the *necessity* either of what is or what happens, or of what ought to happen, except on the basis of a *condition* under which it is or happens or ought to happen" (*G*, IV, 463). Put in other terms: All reasoning must have some standpoint, and what it establishes is conditional on that standpoint. Kant has not shown unconditionally that we cannot be skeptics about freedom and morality. But he has raised the stakes: If we are skeptics about the practical standpoint, we will also have to be skeptics about knowledge and science.

Yet despite the conditional character of human reason we can see why reason, unlike desires, is not an alien cause. Desires come and go, are contingent, varied and naturally caused. Only while a desire (or other alien cause) lasts does it affect deliberations. To speak of desires as alien is just to stress this contingency. Reason, by contrast, depends on nothing separable from an agent. It is merely autonomy in thinking and acting, considered in the abstract. Without autonomy there can be neither practical nor theoretical reasoning. There is no gap between reason and autonomy, because nothing counts as reason except the principles of self-discipline or autonomy that cannot be wholly dispensed with in any thinking or acting. Of course, *finite* rational beings can reject autonomy sporadically. Both evil and error are to be expected. Hence reason, although grounded in autonomy, presents itself to us as an imperative. It does so because we are finite, not because its authority is external or alien.

The Kantian grounding of reason, as of morality, cannot be foundationalist. Anything that could count as foundations would have to be transcendent, and so alien. Once we make the Copernican turn we cannot expect any such foundations to be available. Kant's strategy is rather to give a constructivist account of the authority of reason, whose supreme principle is no more than the maxim or strategy of refraining from acting or thinking on principles that cannot be adopted by all potential agents, regardless of their variable characteristics.

Postscript: a conjecture

If this is a plausible way of reading *Grundlegung* III, why did Kant apparently come to think otherwise? Why in writing the *Critique of Practical Reason* did he assign to morality and freedom the status of "facts of reason [*Facta der Vernunft*]?"

3. REASON AND AUTONOMY IN *GRUNDLEGUNG* III

He would hardly have forgotten if he thought he had so recently done all that could be done to vindicate morality and freedom! So mustn't he have concluded that the argument had failed and that neither could be vindicated without appeal to an unvindicated, transcendent fact? Perhaps. However, if reason's authority is constructed, in the way I have suggested, then reason is precisely and unavoidably *factum*, and not unvindicated and unvindicable *datum*. Those who see the "fact of reason" passages as reversion to transcendent realism assume that the "fact" is given or posited; but if it is made or constructed, the distance between *Grundlegung* and the *Critique of Practical Reason* may not be so great. A great deal of textual spadework would be needed to show whether this conjecture can be made convincing.