Ethics Vindicated

Kant's Transcendental Legitimation of Moral Discourse

Ermanno Bencivenga

OXFORD
UNIVERSITY PRESS

2007

CONTENTS

Note on Texts xiii

Chapter 1. Problems for Ethics 3

Chapter 2. The Framework 6
1. Transcendental Philosophy 6
2. Transcendental Idealism 13
3. Appearances 16
4. Apperception 20

Chapter 3. Freedom 23
1. Overdetermination 23
2. Autonomy 29
3. Rationality 34
4. (Un)intelligible Mysteries 43
5. Relentless Criticism 49

Chapter 4. Values 57
1. A Spectator 57
2. Evil 62
3. Responsibility 68
4. The Geography of (Non)sense 76

Chapter 5. Imperatives 85
1. Laws for Imperfect Beings 85
2. Virtue and Goodness 96
3. The Dynamics of Hope 101

Chapter 6. Ordinary Morality 111
1. Analytic and Synthetic Methods 111
2. The Problems Resolved 113
3. Philosophy and Freedom 115

Notes 123

Index 191

ONE

PROBLEMS FOR ETHICS

The alleged subject matter of ethics is human conduct. Not human behavior; not everything humans do. But, specifically, what they do of their own choice, because they *want* to do it; what they do *freely*. And here ethics faces a first monumental problem: its alleged subject matter runs the risk of vanishing into thin air, of turning out to be purely delusional.[1] For humans are natural beings, hence what they do, *everything* they do, is as much a necessary consequence of preceding events and conditions as the "behavior" of oceans and avalanches; and, one necessary step after (or rather, before) another, it is a consequence of events and conditions well beyond the scopes of their lives.[2] That I "chose" to write this book is a consequence of moves my parents made, and their parents, and their parents' parents, long before I ever came into the picture. Of course, I wanted to write it, but how much of a causal factor is that? That I want to do something, too, follows from things other people did and I could not have wanted to see done (since I was not there); hence, it cannot be a manifestation of my freedom. Thus, human conduct is nonexistent, indeed inconceivable; and ethics is left with nothing to deal with. Nor is the occurrence of indeterministic events (as, say, quantum mechanics sanctions it) going to provide any relief here; even if we are willing to admit that something *A* might happen as the result of pure chance, it would make no sense to claim that *A* is totally random *and also* the outcome of an individual's free choice.[3]

But assume that we successfully meet the monumental challenge above: that somehow we find room for genuine human actions—where an action is the unit of (human) conduct. Immediately we encounter another threat that is just as deadly. For what ethics is supposed to do with its subject matter is *judge* it, *evaluate* it, assess it on the basis of its own standards. If I do something freely, ethics will not content itself with relating it to other things I or others did equally freely, or with elaborating a taxonomy of what I or others typically choose to do, or even with bringing out how far I or others in fact (freely) approve of what I have done. Any such pursuit would belong to an empirical discipline like psychology, or sociology, or statistics; and ethics is no empirical discipline. Its concern is not with understanding what anyone does, but with determining how *good* it is: not good *for* someone, or *for* some purpose or other, but good, period, unconditionally good, good in a totally absolute sense—one that is independent of what happens and might well be in conflict with all that happens. How is this kind of judgment legitimate? How is it more than the expression of individual preferences? What is the place of values in a world of facts?

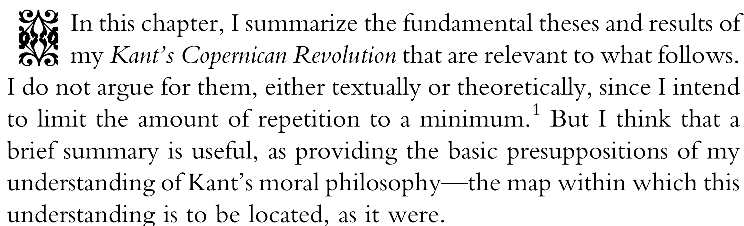
And it is not over. Aesthetics, too, is regarded by many as an evaluative discipline, and as one that might be based on equally ambitious (and unrealistic) expectations. Landscapes, people, and works of art are often assessed by comparing them with standards of beauty which, we might imagine, nothing fits perfectly, with ideals that everything falls short of. And we might occasionally feel nostalgic for such ideals, and desperately long for their realization; but it is unlikely we would go any further. Ethics, on the other hand, does not only evaluate conduct; it also *prescribes* it—when it regards something as good, it also judges it *necessary* that people do it. However beyond our resources good conduct might be, that is precisely what ethics imposes on us: what it says we *ought to* bring about. Ethics is a normative discipline, in a much stronger sense than any other. It is often claimed that scientific theories are normative because the world they "describe" is highly idealized: much more simple and elegant than the real world ever is. And yet, we do not ordinarily think that an astronomical theory tells the universe what to do, or that a sociological theory does that with communities, crowds, or institutions. Ethics, on the other hand, tells us—each of us—what to do; which makes its status, once again, uncertain (what kind of "discipline" is this, that pretends to shape what it applies to? does it amount to mere wishful thinking?) and leaves it totally mysterious where the *authority* that does the prescribing is to issue from. We are

familiar with officers prescribing behavior to their subordinates, and with laws doing so in a state; in all such cases, we can think that the prescribing is authoritative because of the (physical or political) *power* the prescriptive agencies have. But what kind of power makes ethical injunctions authoritative? What sense does it make for ethics to claim, as it often does, that, unless they are consistent with its injunctions, even the officers' or the laws' commands carry no weight, have no real authority?

Ethics has more specific, local problems than the three I mentioned. As with any other human endeavor, its practitioners disagree in subtle and important ways on the details of their positions and arguments. But those three problems are its most basic ones, in the literal sense that, unless they are resolved in a positive way, ethics has no base at all. They are the preliminary to the entertaining of any substantive ethical views, the *conditio sine qua non* for the very legitimacy of moral discourse. Immanuel Kant's philosophy, as I see it, is a sustained, bold, and successful effort aiming at such resolution. In order to reach its goal, it is forced to many digressions, some of them enormously long and complex, and of enormous independent interest. But we do not want to miss the forest for the trees, because there is something of great consequence at stake.[4] The challenges to the credibility of ethics that Kant was facing in the eighteenth century had been raised before, in different languages using different metaphors, and are still being raised, in yet newer jargons; Kant himself would say, indeed, that there is no escaping their constant recurrence. They are an essential component of our form of life: of the irremediably conflictual existence we lead. But it is just as essential to our life that they be answered, and that the answer be loud and clear. The point of this book is to spell out Kant's answer,[5] in a language that speaks to our times, so as to, once again, patiently, attend to the interminable, unavoidable task of establishing the dignity and autonomy of our moral standards.

TWO

THE FRAMEWORK



In this chapter, I summarize the fundamental theses and results of my *Kant's Copernican Revolution* that are relevant to what follows. I do not argue for them, either textually or theoretically, since I intend to limit the amount of repetition to a minimum.[1] But I think that a brief summary is useful, as providing the basic presuppositions of my understanding of Kant's moral philosophy—the map within which this understanding is to be located, as it were.

1. Transcendental Philosophy

Most of Kant's predecessors thought that philosophy could and did establish factual truths; for example, that it could and did establish that God exists, or that the soul is immortal, or that the world is infinite (or finite). As established philosophically, such truths were proved by a priori arguments; hence in fact (one thought) *more* than their factual truth was proved. They were proved to be necessary; and that they were true was then supposed to follow as a trivial consequence.[2] Thus Anselm's and Descartes's ontological arguments proved that God must exist, that He *cannot but* exist, that His nonexistence is inconceivable; from which it was only (it seemed) a small step to conclude that He does exist.

Kant's transcendental philosophy, on the other hand, has no factual import whatsoever. It gives no information about the real world where

we lead our ordinary, everyday life; it cannot add (or detract) anything to (or from) it. The existence and nature of what belongs to the real world is decided by our ordinary experience, inclusive of our empirical sciences. The philosopher is to receive this material and not to challenge it in any way.[3] His task is rather to understand it, to explain how it is possible.[4] More precisely: to provide a conceptual scheme, or logical space, within which the terms used in describing ordinary experience are given definitions generally consistent with that use.[5] It is not for philosophy to decide that, say, we know midsize objects like tables, chairs, and trees; that we do is part of life, and what particular experiences of tables, chairs, or trees count as cognitive is decided by ordinary people in ordinary epistemic contexts, by using their ordinary empirical criteria.[6] But philosophy needs to so determine what objects and knowledge are as to make it possible for us to sometimes know these objects—as to make sense of the claim that we sometimes do (and that the empirical criteria will sometimes attain their intended goal).[7] It is a scandal, Kant thinks, when philosophy cannot account for such claims and is forced to conclude (say) that we are not just empirically wrong in believing we know this or that, but conceptually wrong in believing we can know anything at all—or in believing we can know midsize objects, as opposed to the contents of our own minds.

The real world contains objects and events, and the most important relation among them is causality: how an event brings about another event, how the existence or the behavior of an object determines the existence or the behavior of another one. The logical space contains concepts, and the most important relation among them (as indeed was suggested above) is definition: how a concept is articulated in terms of other concepts, how an understanding of the former is provided in terms of (an understanding of) the latter. So the logical space is like a dictionary, where the concept of an oak is defined in terms of that of a tree, and the concept of an acorn is defined in terms of those of an oak and a fruit.[8] And the crucial demand to be put on this particular dictionary—it is worth repeating—is that its definitions be serviceable in our ordinary dealings, that they do not make it impossible for us to operate with them in ways that fit our general expectations. If I were to define an experience as cognitive, say, when it is particularly vivid, that would contradict my general expectation that many dreams are going to be much more vivid than most waking experiences while not being cognitive (whereas some of those experiences are).

To put it in yet another way, the logical space is like one of those computer programs that help you organize your finances. Whether

you are rich or you are broke is not for the program to decide: that is decided by the empirical data of your assets and liabilities. But it is a condition of the program working properly that its instructions make consistent room for the empirical data: that by following them you do not find yourself running in circles, or getting contradictory outcomes, or being entirely mystified as to where a given asset or liability is supposed to be listed—or whether it is to be listed *anywhere*.

In Kant's own terms, transcendental philosophy is—like all philosophy, that is, all "cognition from concepts" (A837 B865)[9]—entirely constituted of analytic judgments (since "from concepts no synthetic propositions can be derived," N278 XVIII 298). Because this claim seems to contradict some of Kant's own statements,[10] and certainly does contradict a substantial amount of Kantian lore, it will be useful to articulate it further, with specific regard to the ethical works. Take the *Groundwork*, then, where we are told: "That . . . [the principle of autonomy] is an imperative, that is, that the will of every rational being is necessarily bound to it as a condition, cannot be proved by mere analysis of the concepts to be found in it, because it is a synthetic proposition; one would have to go beyond cognition of objects to a critique of the subject, that is, of pure practical reason, since this synthetic proposition, which commands apodictically, must be capable of being cognized completely a priori" (G89 IV 440). Leaving aside details to be discussed later, the general structure of the situation seems clear: Kant is concerned with a proposition of great significance for him, and one that he explicitly designates as synthetic. Fair enough. The whole issue, however, revolves around what it means for him to be concerned with it.

An earlier statement puts us on the right track: "[the] categorical imperative or law of morality . . . is an a priori synthetic practical proposition; and since it is so difficult to see the *possibility* of this kind of proposition in theoretical cognition, it can be readily gathered that the difficulty will be no less in practical cognition" (G72 IV 420; italics added). And later he insists: "*How such a synthetic practical proposition is possible a priori* and why it is necessary is a problem whose solution does not lie within the bounds of the metaphysics of morals" (G93 IV 444). So Kant has a problem, and the best way to understand exactly what that problem is will be to look at how he resolves it: "categorical imperatives are possible by this: that the idea of freedom makes me a member of an intelligible world and consequently, if I were only this, all my actions *would* always be in conformity with the autonomy of the will; but since at the same time I intuit myself as a member of the

world of sense, they *ought* to be in conformity with it; and this *categorical* ought represents a synthetic proposition a priori, since to my will affected by sensible desires there is added the idea of the same will but belonging to the world of the understanding" (G100–1 IV 454). As we will see, the categorical imperative is independent of experience, hence a priori; and I have no difficulty accepting the claim that it is synthetic. What Kant is arguing here, however, is not that this imperative applies, but that it *can* apply. The argument he offers would have no hope of establishing the former, since the intelligible world he invokes to prove his point (whatever his point might be) is one about which (again, as will be detailed later) he must admit we have no knowledge. Therefore, Kant cannot even be attempting to prove the categorical imperative itself; what he is after is a modalized version of it, in which it is preceded by a possibility operator.[11] More precisely, since an imperative expresses the necessity of a certain kind of behavior, the characteristic modality, here as in a number of other crucial cases in Kant, is $\Diamond\Box$: *possibly necessary*.[12] It is *possible* that space *must* be three-dimensional; it is *possible* that every event *must* have a cause;[13] it is *possible* that one *must* act with total disregard for one's idiosyncratic makeup and situation. And this modalized version, as I pointed out in *Kant's Copernican Revolution*, is (with a qualification to be made in the next paragraph) well inside the scope of (analytic) cognition from concepts.[14]

And now for the troubling news. A consequence of Kant's characterization of (transcendental) philosophy is that the latter is not a cognitive enterprise: no knowledge can issue from it. Not in his view, at least, since for him a cognition (*Erkenntnis*—the unit of knowledge)[15] entails the interaction of concepts (or general representations—that is, representations that could in principle apply to more than one thing, though of course some *in fact* apply to only one thing or to none at all) *and intuitions* (or singular representations—constitutionally directed to a single thing); but intuitions do not enter in that reconnaissance of logical space transcendental philosophy consists of—though of course *the concept of* an intuition does. (Transcendental) philosophy can only establish *logical* possibility: it can prove that the description of something, as far as we can tell, is not incoherent. This proof, however, is a function of how detailed the description is and of how deeply we went into the analysis of the terms involved in it—or in what direction: Russell's paradox was close to the surface of Frege's set theory, except that Frege had chosen to look (very intently, and in great depth) away from it.[16] *Real* possibility, on the other hand (or

possibility, period: what is more than an *appearance* of possibility), requires access to a real (singular) example of what we are talking about (to a corresponding intuition), and no such example is forthcoming within the transcendental (that is, conceptual) reflection where transcendental philosophy is developed.[17] *Talk* of examples does; but this talk is incapable of establishing its own consistency, however consistent it might *sound*.[18] In critiquing itself, reason discovers its limits; rational discourse is inspiring and edifying but incapable of proving the veridicality of its tenets. Which is not all bad, as these limits ''make room for *faith*'' (Bxxx); that is, allow for a more nuanced and accurate understanding of the complexity of our form of life.

Three remarks are in order before moving on. First, admittedly, the radical distinction implied here between ordinary concerns and philosophical activity seems artificial: ordinary people make constant use of conceptual tools, even complex ones, and some have argued that what tools those are determines what world they live in (Eskimos live in a world in which there is no such thing as *just snow*, and so forth).[19] The very notion of an ordinary person, one might insist, is a philosophical abstraction. Which is a point well taken—except that it does not detract from the substance of Kant's position but only from its superficial rhetoric. Changing some of the rhetoric but none of the substance, we could then say: It is a scandal that our experience, inclusive of our various attempts at making rational sense of it, should invariably make so little sense; and what causes the scandal is one constant feature of those attempts. They present themselves as final and all-inclusive: that reason of ours which relentlessly motivates their recurrence cannot help thinking of itself as self-standing and self-contained, as in need of nothing external for a full resolution of its problems.[20] But such in not the case: our reason is sharply limited precisely in how it can satisfy its own demands, which is revealed in an obvious way by the poor cognitive status of its pronouncements and in a less obvious but ultimately equivalent way by its necessity to always defer to an *other*—to the nonrational or nonphilosophical as such; to what it itself must characterize as nonrational and nonphilosophical—as a source of the wisdom it is forever (and forever unsuccessfully) looking for.[21] As we will see, this conclusion has no negative impact on reason's ambition or on its nobility—the latter is indeed thought to be even higher here because of reason's failure to attain its ambitious goal (because of its faithfulness to its standards, hence to its vocation, in the face of such failure). But it does set Kant in sharp contrast with all those other rational thinkers who thought that intellectuals like themselves—

whether because they had seen the Forms, or because they knew the principles and causes of things, or because they had reached the stage of Absolute Spirit—exhausted the significance of the world in their intellectual activity, and hence should also rule it (or instruct its rulers).[22]

Second, since the other to which philosophy is supposed to defer has a temporal dimension, as well as a development along this dimension, deferring to it entails that philosophy may have to accept as given, and work hard to establish the possibility of, different material at different times; hence that its task may have to be repeatedly redefined, even to a dramatic extent. Which sometimes gets Kant in trouble because, for reasons I discuss in *Kant's Copernican Revolution*, he is constantly tempted to provide more detail for his conceptual accounts, hence to commit himself more to the (scientific or moral) views current at his time (in his terminology, to move from the critique to the system)[23]—as opposed to staying safely within the confines of such highly general statements as have most of a chance of remaining stable over time.[24] And, insofar as the views he refers to are no longer current, he exposes himself to the risk of being ''refuted'' by later developments that have nothing specifically philosophical about them and are entirely irrelevant to whether his transcendental arguments for the possibility *of the earlier views* are correct. Many of the actual refutations people have proposed over the years turn out to be, upon closer examination, way too hurried: that we are now in possession of abstract mathematical theories about more-than-three-dimensional, non-Euclidean ''spaces,'' for example, does nothing to refute the (nonphilosophical) claim held by Kant that the space of our experience[25] is Euclidean and three-dimensional—in fact, it does not even prove that the objects described by those theories are legitimately called ''spaces'': that the use they make of this term is anything more than a suggestive metaphor. But, clearly, the risk is there; and I will not deny that I am far from sharing some of what Kant took from the ''ordinary people'' of his time—say, his unconditional approval of the talion law, of the death penalty, or of property rights based on first occupancy. In this regard, however, he is in no worse position than any other philosopher, who, whether he aspires to ''comprehend his time in thoughts'' or to be sharply critical of it, can certainly have nonphilosophical views many of his readers judge despicable, and can spend a large amount of his time and energy providing a justification for them—while still, perhaps, making philosophical moves that will benefit all future practitioners of philosophy. In fact, I would add, the

very conception of philosophy that creates this problem for Kant also puts him in a *better* position to address it than most of his colleagues, in two ways. On the one hand, because of (his) transcendental philosophy's dependence on a nonphilosophical other, its verdicts cannot be considered absolute but must always be seen as open to revision if and when the nonphilosophical context changes—if and when, say (to consider some quite radical developments), we mutate into beings who visualize in an eleven-dimensional space, or who have intellectual intuition. On the other, this philosophy intends to prove possibilities, not necessities; and one possibility does not rule out another. Thus Kant's (both philosophical and nonphilosophical) views present themselves (despite his occasional statements to the contrary) as less definitive than most others'; and what falsehoods the man Kant may have believed in the sciences or in morality, or what mistakes the philosopher Kant may have made when rationalizing those falsehoods, can be corrected by the very listening attitude, and the very critical activity, which by all means he did not initiate from scratch but of which he first gave us a lucid and articulate account.[26]

My final remark builds on the previous ones and sets the stage for the next section. Kant was not just interested in describing the conceptual space of his time; he wanted to revolutionize it. And such revolutions often have empirical consequences—which once again makes the neat separation between transcendental and ordinary concerns look too simple. The first person who thought of equities as assets did not add a dime to anyone's wealth; but eventually, *because* people thought of equities as assets, many of them had more money to spend. Kant often tries to minimize the impact that his novel philosophical views can have on everyday life—most typically when he is defending himself against the censors' attacks.[27] But such defenses are disingenuous, and at other times he clearly manifests the hope that, in the long run, how we think of things will change how we live: "it could well happen that the last would some day be first (the lower faculty [of philosophy] would be the higher)—not, indeed, in authority, but in counseling the authority (the government). For the government may find the freedom of the philosophy faculty, and the increased insight gained from this freedom, a better means for achieving its ends than its own absolute authority" (R261 VII 35).[28] And yet, though these links complicate the relation between the two levels (in ways that will turn out to be crucially relevant to the Kantian analysis of morality), they do not deny their distinctness. That people come to have different basic conceptions is a fact, and as such it can certainly have causal influence

in the empirical world, but what conceptions those are, and how they are related to one another, is independent of who holds them, or of whether anyone holds them at all.

2. Transcendental Idealism

Within transcendental philosophy, various positions are possible depending on what concept(s) is (are) considered primitive in logical space. The two positions around which Kant's Copernican revolution unfolds are *transcendental realism* (TR) and *transcendental idealism* (TI); that revolution is the transition from TR to TI, the "experiment" (Bxvi) of adopting TI instead of TR.

TR is the structuring of logical space *implicit* (for Kant) in the tradition: most likely, traditional philosophers would not have described what they were doing by using this language (or anything equivalent), but describing it that way best makes sense (Kant thinks) of their practice and of its outcomes. That is, it is most useful to characterize them as *thinking* in terms of objects: as taking the concept of an object (*res*) to be the fundamental one, and every other concept to be dependent on it—and most often definable (possibly after numerous steps) by an eventual reference to it.[29] Thus, a sailor is a human who works on a boat, and a human is a rational animal, and an animal is a self-moving living object; hence, a sailor is a rational self-moving living object who works on a boat (and rational is an object that can think and argue, and a boat is an object that holds humans and merchandise and crosses oceans, and an ocean is an object . . .). What an object is, on the other hand, a transcendental realist cannot say: he can use synonyms (a being, an entity, a thing, the bearer of properties), but an informative definition is out of the question, for no fault of his—primitive concepts cannot be defined.

TI, on the other hand, is the structuring of logical space that takes the notion of a representation (*Vorstellung*) to be fundamental, and every other concept (including the concept of an object—"what may be contained in my concept of a thing . . . [, what] belongs to its logical essence," TA89 IV 294) to be dependent on it—and most often definable in its terms.[30] Because TI is to this day a minority position, it is held to higher standards; and no sooner do people hear it described as I just did than they start asking, "What is meant by a representation?" or even worse, "*Whose* representations are we talking about?" And then they might even conclude that the position is not new after all, because

a representation is nothing but a state (or a property) of a mind, which is an object like any other. Such irrelevant questions and criticisms must be resisted, while admitting the initial awkwardness of an unfamiliar way of thinking; one must firmly reject any tacit commitment to the very realist vocabulary that is being challenged while gently guiding interlocutors, by appropriate examples and rhetoric, to seeing things in a manner consistent with TI.[31]

Representations, here, are *no one's*; not, at least, to begin with. For at the stage where we are—at the very origin of logical space—there is nothing other than representations, hence nothing for them to belong to.[32] Eventually, after objects are defined, and some of them are characterized as minds, it will be possible to ascribe some representations to them; but it will take a lot of work to get there, if indeed we ever do. Which suggests that the word "representation" is an unfortunate choice, since it seems to imply that something is present to something else (indeed, something that *was* present once and is now present again—in this sense, "*Vorstellung*" is a little better, as it evokes no repetition), and what could both of these "somethings" be other than objects? But we cannot help that: there is no neutral language in which the various setups of logical space can be entertained and compared with one another. What language is available is always the expression of a given setup, and the current setup (current at Kant's time, and also at ours) is TR; so our language is reflective of this dominance, and the only possible tack for would-be revolutionaries is a translation (of their vision into the existing words and phrases) that is also (inevitably, like all translations) a betrayal—the forcing of new wine into old bottles, at risk of exploding them, and of spoiling the wine.

In TR, where representations are conceptually dependent on objects, they cannot, however, be regarded as conceptually dependent on (let alone as definable in terms of) the objects they are (allegedly) representations *of*, since (1) often there is no such thing and (2) whether or not there is one is often an empirical matter, which we cannot allow (not prima facie, at least) to have an impact on conceptual issues: if the dependence must be ruled out in *one* case, it must be ruled out in *all* cases that are only empirically different from that one (at least initially, and open to possible revision by the regimenting project to be mentioned later). So here it is crucial that, for example, one cannot understand my imagining a winged horse as a case in which an imagining relation occurs between myself and a winged horse (for nothing is a winged horse, hence an imagining relation with "it" is a relation with

nothing—or not a relation at all); to give a conceptual account of this experience (as well as of the empirically different one of imagining the current Pope), the transcendental realist must regiment it in some way, and claim that its logical form is not what it appears to be.[33] In TI no such problem arises, and a representation is always of something: "All representations, as representations, have their object" (A108). It continues to be the case, however, that a representation cannot depend conceptually on what it represents (as the realist thinks of it)—this time because of the conceptual priority of representations. In fact, the very unfailing success representations have in "hitting upon" an object proves this "relation" to be a trivial, purely verbal, one: to speak of the object *of* a representation is only to bring out, in somewhat different terms, the representation's representational character.

Using more recent terminology than Kant's, the object *of* a representation, in TI's sense, could be called an *intentional* object[34] (and the representation's representational character could be called its *intentionality*); but we need to be clear that intentional objects here are not another *kind* of object as traditionally understood—they are not a species of the traditional genus *object* in the sense in which, say, red or round objects are. They are a step in a new understanding of objects altogether: of objects as conceptually dependent upon representations.[35] And, because of Kant's empirical conservatism (which, as related to the present issue, is relabeled *empirical realism*), they often end up being objects only in a manner of speaking, objects by courtesy. For, being an empirical conservative, Kant does not want to add any new objects to the world, hence in the final analysis he wants to say that the winged horse I am imagining, though an "intentional" object, is *really* no object at all: once again, *there is no such thing*.[36]

The conceptual "construction"—that is, definition—of (real) objects takes place by imposing requirements on representations.[37] Such requirements are best thought of as placed not on individual representations but on systems of representations. Kant calls them *categories*; equivalently, they could be called conceptual criteria of objectivity.[38] A system of representations cannot be called objective (that is, it cannot amount to knowledge, since a cognition is "an objective perception," A320 B376, or an objective "representation with consciousness," A320 B376) unless it is consistent and connected: the latter criterion, which replaces for Kant (in Hume's wake) traditional causal efficacy, amounts to lawlikeness or regularity—that representations follow one another according to rules.[39] Also, the objects of the representations must be identifiable and countable; since our means of identification are space

and time (they are the forms of intuition: the conditions at which only we can represent anything singular—or, in more contemporary terminology, the necessary conditions for us to be able to provide any demonstrative reference), objects must have a spatiotemporal location.[40] And they must be irreducible to any conceptual specification of them, and richer than any such; it must always be possible for us to extend the relevant system of representations, to ''synthesize'' additional features of their objects.[41] And so on. That a representation is objective *means* that it belongs to a system of representations satisfying these criteria; that the intentional object of a representation is a real object, or an object *simpliciter* (an object, period), hence that there really (not just verbally) is a relation between the representation and an object, means that this representation is objective—or cognitive. Which is what Kant expresses, famously, by inviting us to assume ''that the objects must conform to our cognition'' (Bxvi).[42]

3. Appearances

So far, I have spoken vaguely of a system of objective representations. The question naturally arises: how large is this system supposed to be? And the most obvious answer, the only one that would satisfy reason's drive to universality (more about this later), is: a global system, that is, a set of representations to which nothing further could be (consistently) added, that is no proper part of any other such (consistent) set.[43] If this global, universal system were still to be regarded as objective, all representations belonging to it would have to be conceived as *unified*—not as arbitrarily jumbled together but as objectively connected: as representations of elements and aspects of one and the same objective *world*.[44] Kant, however, proves that the thought of an objective world involves us in endless, irremediable contradictions; as a result, the conceptual criteria of objectivity can only be applied meaningfully *in a context*, that is, within a horizon that is not itself interrogated, to which the same criteria are not applied—at least at the moment; they certainly could be applied to it at some other time, provided that what is now the context were to become part of another (uninterrogated) context. To avoid absurdity, and make knowledge possible, overweeningly ambitious reason must give way to *understanding*, that is, to its own projection (the projection of its own criteria) onto a limited, dogged, stepwise, myopic mode of operation.[45]

This is not an empirical issue; it has nothing to do with empirical limitations of ours, which we might think of eventually overcoming. It is a conceptual, or transcendental, issue; it has to do with a conceptual clash internal to the very criteria of objectivity—or rather, to the spatiotemporal conditions of their application. Take identifiability, for example, as translated into being assigned a definite spatiotemporal location. It turns out that such a location can only be assigned to an object relative to other objects, in a context in which other objects are supposed to have already received their own location. If we try to overcome this limit, the very search for identification becomes incomprehensible: it makes no sense to ask where or when an object is, period (in more dignified philosophical terms: where or when it is *absolutely*), or in relation to *the totality* of space and time (otherwise put, it makes no sense to ask where or when *the whole world* is, as opposed to asking where or when something is *in the world*). Similarly, the irreducibility of objects to concepts (of the ultimate subjects of predication to their predicates, as we could also say) clashes with the manifold nature of space and time: with the fact that spatiotemporal objects (what subjects of predication we do in fact encounter—and we can in fact know) are indefinitely divisible, hence there are no ultimate constituents of matter, no "objects" that could not also be seen as complicated (systems of) properties of, or relations among, simpler objects. Or take the lawlike (spatiotemporally determined) connectedness of objects (better: of the events in which they participate): it can only be meaningfully applied to some current objects (or events) if some antecedent objects (or events) are presupposed to which the former are connected. Therefore, we will never be able to reach an absolutely first antecedent for anything in the world (or in the chain of events) and we will never know, in an absolute, definitive sense, *why* that is (or happens)—but only relatively to something else that also is (or happens).

None of this would impress the transcendental realist. That criteria of objectivity have no meaningful absolute application is *our* problem, he would say, an epistemological one, and one that does not touch the metaphysical structure of the objects themselves, which are what they are whether they can be identified or not, whether they can be radically opposed to their properties or not.[46] So this contradictory outcome—this "antinomy of reason"—is no reductio of TR, as has often been claimed, but rather an important articulation of TI.[47] What it proves is that within TI the occurrence of knowledge necessarily depends on the act of choosing a context, and of holding on to it for as

long as it is to be relevant[48] (conservation is continuous creation); and this act of choice[49] (of *synthesis*, [50] to use a Kantian term that has already emerged above; and now the significance of this word, of this ''putting together,'' begins to come forth) can only be conceived as spontaneous, as itself not determined by anything else—because without presupposing it no determination is possible, because it is itself the origin of all determination. In TI we cannot think of knowledge as merely receptive, as the passive acceptance of a structure simply ''given'' to it: the world is (to be conceived as) being constituted as much as it is received, within the very same experience of receiving it; that ''adequacy'' to its object which is the ideal of knowledge (''[t]he nominal definition of truth, namely that it is the agreement of cognition with its object,'' A58 B82) is necessarily infected by an activity that makes the object what it is. ''[E]xperience cannot be given but must be made'' (O122 XXII 405).

Empirically, we distinguish objects from appearances, delusions, phantoms; and we do so by regarding the latter (but not the former) as partial, incomplete, unstable, and as dependent on our support for their very being (as well as for being what they are). If in a moment of confusion (or in the grips of a powerful desire) I see a tree as a person dear to me, that ''person'' will not sustain a prolonged inquiry: ''she'' will not display other angles (my perception of her will not be enriched) as I move around her (or rather, around the tree), in fact she might no longer be there if I look at her (that is, at the tree) again after getting distracted for a moment. Her existence is fragile, ready to collapse as soon as I stop offering it my cooperation. What is true empirically of this ''object,'' Kant thinks, is true conceptually of all (empirical) objects: since a spontaneous act must be conceived as originating their being, they all lack the self-sufficiency, the independence objects *ought to* have. Nothing we ever encounter fully matches what we would expect of an object; we only ever encounter faulty objects, objects to a point. Transcendentally, empirical objects are all mere appearances, and it is a *transcendental illusion* to conceive of them as independently real—an inevitable, but still deceptive, mistaking them for what they cannot be. Or, to put it otherwise (and introduce more terminology), the concept of an object is an *idea* of reason, *that is*, a representation for which no adequate realization can be found in experience.

The other, more positive side to this coin is that within TI, where representations are primary (and because thoughts—or concepts—are a kind of representations), we can still *think* of objects in the proper

sense, objects that are what they are entirely of themselves, independently of any external contribution or choice[51]—unreal as these objects are bound to be, thinking of them here (we know) is not thinking *of nothing*—and we may even claim that such thinking (of objects of pure thought, *noumena*) is a necessary consequence of reason's (frustrated) aspiration to fulfilling its standards (to realizing its ideas), and of its perpetual dissatisfaction with what world it is forced to recognize as real. Because of those features of the spatiotemporal framework which we found earlier to clash with our conceptual criteria of objectivity, none of these "objects in the proper sense" can be thought of as spatiotemporal (there are no things in themselves in space and time), hence we can never think of acquiring any information about them. They are nothing but fictions, and yet we do nothing wrong when entertaining them, or even (if appropriate) when judging what *can* be experienced in their light.

There is a tendency within Kant interpretation to overstate this positive side of the critical outcome, which it will be good for me to address here. People get carried away by passages asserting that "[a]s soon as . . . [the] distinction [between appearances and things in themselves] has once been made . . . , then it follows of itself that we must admit and assume behind appearances something else that is not appearance, namely things in themselves" (G98 IV 451), and conclude that the real world is made of things in the proper sense, which we can only know *as they appear to us*. Next thing you know, Kant is turned into an extreme case of Locke, and Schopenhauerian Nirvana is at hand: everything we have access to is a secondary property, but there is still something unspeakable that is the *true* basis of it all. And, in view of what crucial role synthesis has now acquired (and of moral considerations to follow), freedom looks like an attractive candidate for occupying this exalted metaphysical position; so one will declare that the noumenal subject is really exercising its spontaneous will, whatever the case might be for its unfortunate little brother that inhabits the delusive world of experience. In addressing this (exegetical, if not substantial) nonsense, we must remind ourselves of the transcendental nature of Kant's enterprise: all he means by such passages as the one quoted above, and all he *can* possibly mean, is that *calling* something an appearance amounts to also mobilizing *the concept* of something that would not be an appearance and in comparison with which our appearance talk can be made sense of. The thought of a thing in itself is the conceptual ground for thinking of appearances, and of spatiotemporal objects *as* appearances, much like the thought of a perfect,

archetypal human is the conceptual ground for thinking of any concrete human specimens as imperfect[52]—which is not supposed to imply that, because there are imperfect humans, there must also be perfect ones (which we only experience as imperfect?); or that, because there are appearances, there must also be things in themselves. All we can say, in all such cases, is that "our reason . . . [feels] a need to take the *concept* of the unlimited as the ground of the concepts of all limited beings" (R11 VIII 137–38).[53] Reason can provide a verbal articulation for this need because its criteria of objectivity are, at bottom, purely intellectual conditions—the criterion of causal connectedness is, at bottom, the purely intellectual condition of finding a ground for something—and hence can be used independently of spatiotemporal coordinates; though, of course, it is only when they are applied in the context of those coordinates that they acquire as much definiteness as makes it possible to say that *objects* are involved. And, in conclusion, "[t]he thing in itself = x is a mere thought-object" (O184 XXII 421). "[When we] make the distinction between the representation of the thing *in itself* and that of the same thing as *appearance* . . . [then] concepts, not things, are contrasted with one another" (O174 XXII 32–33).[54]

4. Apperception

In looking for a firm basis for their a priori arguments, traditional transcendental realists were typically drawn to the experience of self-consciousness. There, it seemed, one could make contact with something whose existence and properties were beyond doubt: however questionable one finds the outside world, there is no question (or so Descartes and others believed) that I exist, that I think some thoughts, feel some emotions, and so on. I might or might not succeed in building a bridge between such certainties and an equally certain knowledge of what is other than me (Descartes believed this to be possible, others did not); but, whatever the fate of this subsequent operation, *that* I am and *what* I am is to be regarded as settled.

In TI, however, self-consciousness (in Kant's terminology, *apperception*) provides no knowledge. The first-person pronoun "I" must be able to accompany all representations:[55] it is part of the logic of representations (of what representations are necessarily like) that they always have not only an object but also a subject[56]—understanding by the latter: they can always be thought of as "representing" from a

specific point of view. But this point of view, which may be considered responsible for the mysterious act of synthesis that makes objective experience possible, is not itself an object—much as the horizon of experience *as such* never is, and indeed the two are but different angles on the same mystery. If I direct my attention to it, I end up either turning it into an ordinary empirical object ("it is this body, located in this position, seeing things consistently with its location"), which is just as dependent as any other such object on the original positing of a point of view, or being left with something totally indefinite, an "I, or He, or It (the thing), which thinks" (A346 B404)[57]—something to which I cannot attribute any quality; indeed for which I cannot even meaningfully pose any issue of identification or distinction, hence I cannot say *how many of it (them?)* there are. "The subject is not a particular thing but an idea" (O175 XXII 33). "The *consciousness* of myself is logical merely and leads to no object; it is, rather, a mere determination of the subject in accordance with the rule of identity" (O188 XXII 82). "No quantum of substance is possible in the soul. Hence also nothing that one could determine through any predicate and call persistent" (A183 XXIII 31).[58]

Thus, once again, TI ends up seeing things in reverse order from the tradition. The self "revealed" to consciousness can be no starting point for any epistemic construction: its unity is a purely formal one (it signals that experience always comes in a certain form) but is not substantial—it is not the unity of a substance, of an object.[59] If I want to move beyond this purely formal level, I must focus on the spatio-temporal *content* of experience: insofar as such content can be conceived as unified by the categories into a connected world,[60] it will be legitimate to think of that world as issuing from a single point of view. The unity of apperception can be nothing other than (categorial) coherence, hence it is dependent upon the coherence of the world that is apperceived. And, since I already know that the latter coherence can never be completely established (because the notion of "the whole world" is contradictory), I also know that *both* coherences will have to be thought of as always only "in progress": painstakingly constructed (by the understanding) one step at a time, invariably appealing to a context that must be taken for granted, constantly at risk of exploding into *in*coherence (as far as the I is concerned, into schizophrenia; as far as the world is, into an undifferentiated manifold) when the next step is taken.[61]

It sounds like a devastating outcome, and in some sense it is. In *Kant's Copernican Revolution* I described it by saying that my relation to

the subject of self-consciousness is more akin to what we ordinarily call faith than to (anything Kant would consider) knowledge. But remember: knowledge must be limited, and faith is what it must make room for. The basis of knowledge is itself noncognitive; therefore, that I cannot know the self, that I can only think of it, believe that it is, have trust in its action, just makes the self one more *noumenon*—one more of those unknowable (hence unreal, though not for that reason insignificant) entities and processes I must invoke to make sense of what I do know. And, if this faith finds confirmation in some other, noncognitive experiences I also have, its noncognitive character will be no indictment of it, given how little knowledge is able to help itself.