# Ordinary Goodness
– for –
# Ordinary People

## An Essay on the Kantian System

Jonathan D. Shapiro
Word Count: 28,905
July 1, 2020

## ABSTRACT

Kant does not explicitly present an account of what one might call *ordinary goodness for ordinary people*, or the sense of moral worth one has in mind when one judges someone to be a "good person": somewhat worryingly, when he does discuss the moral worth of people in *Religion Within the Boundaries of Mere Reason*, he seems to stray very far from terms which can apply to ordinary people in their everyday lives. Indeed, there is a plausible line of reasoning that owing to, first, Kant's thoroughgoing determinism and, second, his complete rejection of compatibilist conceptions of freedom, Kant cannot take ordinary people to be free and, as a result, morally responsible—*a fortiori*, he cannot have an account of ordinary goodness for ordinary people. I find that, ultimately, Kant can reject such reasoning through what I call his doctrine of *Transcendental Compatibilism*, which consists of three claims: (i) due to Transcendental Idealism, Kant's determinism is compatible with a certain libertarian-style freedom called transcendental freedom; (ii) transcendental freedom is precisely that kind of freedom required for moral responsibility according to Kant's moral theory, and; (iii) ordinary human beings actually are, under normal circumstances, transcendentally free and, as such, are morally responsible despite living in a deterministic world. By understanding this doctrine, one can derive an account of phenomenal moral worth—*i.e.*, ordinary goodness for ordinary people—from Kant's comments on noumenal moral worth. Therefore, in Chapter I, I explore Kant's determinism and, through examining his resolution to the Third Antinomy of the first *Critique*, establish (i) above; in Chapter II, I examine the core of Kant's moral theory, establish (ii) and (iii) above, and respond to three potent objections; finally, in Chapter III, I elucidate Kant's account of noumenal moral worth so to derive a positive characterization of phenomenal moral worth—*i.e.*, a Kantian account of ordinary goodness for ordinary people—therefrom.

# Contents

**Acknowledgments** iii

**A Note Concerning Citations and Abbreviations** iv

**Introduction: A Problem of Goodness** 1

**I Kant's Determinism** 7
1. Objectivity and Constitutivity . . . . . 7
2. Key Presuppositions . . . . . 14
3. The Analogies of Experience . . . . . 17
   *i.* The First Analogy . . . . . 19
   *ii.* The Second Analogy . . . . . 22
   *iii.* The Third Analogy . . . . . 24
4. Resolution of the Third Antinomy . . . . . 26

**II Transcendental Compatibilism** 29
1. The Basic Components of Kant's Moral Theory . . . . . 29
2. The Moral Law and the Singularity of Freedom . . . . . 38
3. The Fact of Reason . . . . . 48
4. Three Objections to Transcendental Compatibilism . . . . . 49

**III The Goodness of Persons** 57
1. Noumenal Goodness . . . . . 57
2. Phenomenal Goodness I: Radical Evil and the Change of Heart . . . 60
3. Phenomenal Goodness II: Virtue and Progress . . . . . 65
4. A Schema for Goodness: Conscientiousness and Reflectiveness . . . . 68

**Conclusion: Ordinary Goodness for Ordinary People** 70

**Bibliography** 73

# Acknowledgments

There are many people without whom this thesis would never have been written—or else ended up completely differently—but I can only thank so many of them here:

First, Patricia Kitcher, for introducing me to the wacky world of Kant and sparking a life-long passion;

Second, Elisabeth Huh, whose philosophical spirit (in the true sense of the term) and ever-engaging debates—in which we don't exactly *agree* but don't exactly *disagree* either—force me to constantly rethink my positions and reconsider what I had previously overlooked;

Third, Oli Woolley, for taking my mind *off* of Kant when doing so was desperately needed;

Fourth, my father, Mark Shapiro, for supporting and encouraging my curiosity and intellectual development, not just as I worked on the thesis, but throughout my whole life;

And finally, Ralph Walker, whose advice and criticism greatly improved every aspect of this thesis and, furthermore, without whose work I would never have even considered this topic in the first place.

# A Note Concerning Citations and Abbreviations

Following standard practice, all citations to the *Critique of Pure Reason* have the form 'A$x$/B$y$', where $x$ and $y$ represent the pagination in the first and second editions, respectively, unless the passage appears in only one of the two. All citations to Kant's other works have the form '$x.y{:}z$', where $x$ represents the name of the work as abbreviated in the table below and $y$ and $z$ represent the standard volume and pagination, respectively, from the Prussian Academy Edition.

| | |
|---|---|
| *An Answer to the Question: What is Enlightenment?* | WE |
| *Anthropology from a Pragmatic Point of View* | A |
| *The Conflict of the Faculties* | CF |
| *Critique of Practical Reason* | CPR |
| *Critique of Pure Reason* | CR |
| *Critique of the Power of Judgment* | CJ |
| *The End of All Things* | ET |
| *First Introduction* to the *Critique of Judgment* | FI |
| *Groundwork of the Metaphysics of Morals* | G |
| *The Jäsche Logic* | JL |
| *Lectures on Pedagogy* | LP |
| *Metaphysical Foundations of Natural Science* | MF |
| *The Metaphysics of Morals* | MM |
| *On a Discovery Whereby Any New Critique of Pure Reason is to be Made Superfluous by an Older One* | OD |

| | |
|---|---|
| *On the Common Saying: That May be Correct in Theory, but It is of No Use in Practice* | TP |
| *Prolegomena to Any Future Metaphysics that will be able to Come Forward as Science* | P |
| *Religion Within the Boundaries of Mere Reason* | R |
| *Toward Perpetual Peace* | PP |
| *What Does It Mean to Orient Oneself in Thinking?* | OT |
| *What Real Progress has Metaphysics Made in Germany Since the Time of Leibniz and Wolff?* | PM |

In addition, citations to books by Henry Allison and Onora O'Neill use the abbreviations in the following table.

| | |
|---|---|
| *Acting on Principle* | AP |
| *Constructing Authorities* | CA |
| *Constructions of Reason* | CoR |
| *Kant's Theory of Freedom* | KTF |
| *Kant's Transcendental Idealism* | KTI |

Finally, the following abbreviations will be used throughout—though I introduce them in-text upon the phrase's first use, I include this table for ease of reference.

| | |
|---|---|
| Categorical Imperative | CI |
| Fact of Reason | FR |
| Formula of Autonomy | FA |
| Formula of Humanity | FH |
| Formula of Kingdom of Ends | FKE |
| Formula of Laws of Nature | FLN |
| Formula of Universal Law | FUL |
| Law of Natural Causality | LNC |
| Ordinary Goodness for Ordinary People | OGOP |
| Ought Implies Can | OIC |
| Principle of Hypothetical Imperatives | PHI |

# Introduction
## A Problem of Goodness

> It is impossible to think of
> anything at all in the world,
> or indeed even beyond it,
> that could be considered
> good without limitation
> except a **good will.** (G.4:393.)

IMMANUEL KANT

It is not an uncommon conclusion that, whatever virtues Kant's moral theory may have, it ultimately fails to mesh well with ordinary moral experience, not necessarily because its requirements are morally unpalatable—though some certainly think this—but because it supposedly fails to provide adequate meaning for important moral concepts. For example, Barbara Herman bemoans that Kant's moral theory lacks a conception of *character* that is so important for virtue ethics, taking it upon herself to *introduce* such a conception of character into Kant's theory.[1]

A particularly acute problem, in my own opinion, is that Kant's moral theory does not appear to have a place for the perfectly ordinary judgment that so-and-so is a *good person*.[2] Such a judgment involves what I call *ordinary goodness for ordinary people* (OGOP): such goodness can be predicated of people in their everyday lives and the sense of goodness involved is ordinary in that everyday experience can provide

[1] 37.

[2] All discussion of 'goodness' throughout is of specifically *moral* goodness; as such, I freely drop the qualification.

reasons to (not) predicate it of someone. However, insofar as Kant addresses the moral worth of people, his discussion strays far from everyday experience:[3] whatever conception of goodness Kant may have, it seems quite remote from OGOP.

Indeed, a fairly plausible line of reasoning suggests Kant cannot have *any* account of OGOP: despite his thoroughgoing determinism, he flatly denounces all compatibilist notions of freedom as "nothing better than the freedom of a turnspit, which, when once… wound up, also accomplishes its movements of itself",[4] concluding that such notions are completely insufficient for moral responsibility. But, so the objection goes, this implies that people in their everyday lives, as fully determined, cannot possibly have the kind of freedom required for moral responsibility—as such, ordinary people are not morally responsible and, *a fortiori*, no concept of goodness can ever apply to them. This is not to say that there can be no moral responsibility *tout court*: noting Kant's transcendental distinction between phenomena and noumena, the objection applies only to *phenomenal* and not to *noumenal* people, so it remains possible that noumenal people, as morally responsible, may be good. This would explain why Kant's discussion of goodness seems to employ noumenal terms,[5] but it fails to relieve the objection's key worry: is there any hope for a Kantian account of OGOP?

Some—perhaps even most—commentators would think not. For example, Herman, although sympathetic to the general claims of Kant's moral theory, announces that we need a "Kantian ethics liberated from its noumenal baggage" in order to develop a moral theory that can accommodate character[6]—and, by extension, OGOP. But since "noumenal baggage" grounds Kant's entire critical system, this implies Kant's moral theory can only be saved by fundamentally rejecting his overall system; hardly a vote of confidence.

I demur: not only *can* Kant have an account of OGOP, I take it he actually *does*—indeed, *must*. That is, despite no apparent effort to draw all the threads together, Kant provides all the resources necessary for a positive account of OGOP—though, perhaps surprisingly, one can only draw it out by diving head-on into his "noumenal baggage".

[3] *Cf.* R.
[4] CPR.5:97.
[5] *Cf.* R.
[6] 36.

To rebut the objection that Kant cannot have an account of OGOP, one must understand, first, Kant's particular account of determinism, and, second, the freedom he finds necessary for moral responsibility: these are, respectively, discussed in Chapters I and II. More precisely, I aim to develop, on Kant's behalf, a doctrine I call *Transcendental Compatibilism*, consisting of the following three claims:

(i) Due to Transcendental Idealism, Kant's thoroughgoing determinism (as established by CR's Analogies of Experience) is fully compatible with a particular libertarian-style freedom called *transcendental freedom*, rendering it theoretically possible for rational beings to be *both* transcendentally free *and* completely determined by laws of nature;

(ii) Transcendental freedom is both necessary and sufficient for moral responsibility, and;

(iii) Some rational beings—including ordinary people under normal conditions—*actually are* morally responsible and thus both transcendentally free and completely determined by laws of nature.

It is important that all three claims are established: (i) alone leaves open whether transcendental freedom is sufficient for moral responsibility, while (i) and (ii) without (iii), though sufficient to answer the compatibilist question of whether moral responsibility is compatible with determinism, are insufficient for any Kantian account of OGOP. Thus, I establish (i) in Chapter I and (ii) and (iii) in Chapter II, where I then respond to three potent objections which, together, constitute the above objection that a Kantian account of OGOP is untenable. Finally, in Chapter III, I use the insights of Transcendental Compatibilism to derive, from Kant's comments on noumenal moral worth, an account of phenomenal moral worth, *i.e.*, a Kantian account of OGOP.

❧ ❧ ❧

Before I begin my argument proper, I highlight six key points.

First, I will treat three of Kant's distinctions—that between appearances and things-in-themselves, that between phenomena and noumena, and that between empirical and intelligible objects—as equivalent throughout; additionally, I take it

that "objects of experience" are simply phenomena (and thus appearances/empirical objects).[7] Although I recognize that Kant does on some occasions claim that these distinctions are not exactly equivalent, he fairly regularly uses them interchangeably and so, while one certainly can carefully separate the three distinctions, I find that, for my purposes, this would require an unnecessary level of detail.[8]

Second—perhaps most importantly—I presuppose a "two-aspect" interpretation of Transcendental Idealism throughout, specifically the kind championed by Allison.[9] Though ultimately interpretations of Transcendental Idealism must be established through argument, I rely on Allison's here as, given space constraints, it would be otherwise impossible to present my own arguments concerning a Kantian account of OGOP. However, since, so far as I can see without offering any explicit arguments, Transcendental Compatibilism is only viable on a two-aspect view,[10] my arguments for Transcendental Compatibilism function as an indirect argument for the two-aspect view.

To clarify, I summarize this two-aspect view as follows: while some noumena are logically possible but, since they violate the necessary conditions for the possibility of experience imposed by the mind, cannot appear as phenomena (*e.g.*, round squares, biangles, things that are both completely red and completely blue[11]), one can consider phenomena (*e.g.*, chairs, birds, and thoughts) not as they are insofar as they are objects of experience but, abstracting from human sensibility, as they are in themselves, *i.e.*, as noumena. As such, the distinction between phenomena and noumena is not an *ontological* one between kinds of objects but rather an *epistemolog-*

[7]It is important to distinguish, as Kant does, between appearance in the sense of phenomena and *apparency* or mere seeming-to-be: the phenomenon is, so to speak, an *empirical thing-in-itself* which can *seem* to be different from how it actually is—*e.g.*, as a result of an optical illusion. *Cf.* A45-6/B62-3, A258/B313-4; PM.20:269.

[8]Many commentators treat these distinctions as interchangeable only implicitly, though Allison makes it explicit—*cf.* "Non-Spatiality" 316n13.

[9]*Cf.* KTI. Allison is not alone in subscribing to a two-aspect view—*cf.* O'Neill, COR 60; Korsgaard x-xi.

[10]For example, it seems to me that, on the predominant two-world interpretation, Transcendental Compatibilism would require some sort of preestablished harmony—but, as that is precisely the kind of doctrine Kant wants to avoid, a genuine Kantian would have to reject Transcendental Compatibilism (or else abandon the two-world interpretation).

[11]More precisely, these are what such noumena *would appear as* if, *per impossibile*, they could appear to us as objects of experience.

*ical*—indeed, *transcendental*—one between ways in which objects can be considered: a phenomenon is thus numerically identical to itself *qua* noumenon.[12]

Third, though I reject a two-worlds interpretation of Transcendental Idealism, I nevertheless refer to the noumenal and phenomenal *worlds*, not to denote ontologically distinct planes of existence, but simply to mean the sum-total of either noumena or phenomena.[13] I similarly discuss noumenal or phenomenal *objects*, though, strictly speaking, there are only objects *simpliciter* which can be considered either phenomenally or noumenally.

Fourth, I mainly ignore the modal categories—*i.e.*, possibility, actuality, and necessity—when discussing the categories in general;[14] as Kant himself acknowledges, the modal categories are unique in that, just as a judgment's modality adds nothing to its content,[15] the modal categories add nothing to an object's concept,[16] explaining Kant's famous claim that existence is not a real predicate.[17]

Fifth, I presuppose the principle that *ought implies can* (OIC) as it is an eminently Kantian principle Kant himself uses.[18]

Sixth, I will assume that, for Kant, Transcendental Idealism consists of the following claims: that space and time are our *a priori* forms of intuition and we are therefore discursive, not intuitive, intelligences;[19] that all phenomena are temporal, some phenomena are spatial, and all noumena are non-spatiotemporal;[20] that phenomena are transcendentally distinct from noumena; that every phenomenon is

[12] *Cf.* Allison, KTI 3-73; "Non-Spatiality" 317-8. That there may be noumena without corresponding phenomena thus expresses the epistemological fact that there may be things of which humans cannot, even in principle, have any experience.

[13] *Cf.* O'Neill, COR 61; Korsgaard 167N10.

[14] *Cf.* Alison, KTI 199N73.

[15] A74/B99-100.

[16] A219/B266.

[17] *Cf.* A598/B626.

[18] *Cf.* A547-8/B575-6; R.6:50.

[19] Strictly speaking, space and time are only *our* forms of intuition and, as other forms may be possible, Transcendental Idealism is more general than as I state it here; however, as this work only concerns the human perspective, the above statement is sufficient. *Cf.* A27/B43, A231/B283, A255-6/B311-2; PM.20:267.

[20] I agree with Allison that the non-spatiotemporality of noumena follows analytically from the concept of noumena and, as such, reject that the so-called "neglected alternative" is a genuine alternative. See "Non-Spatiality; *cf.* KTI 19. Though Falkenstein disagrees, he ultimately finds that noumena must be non-spatiotemporal as well, so I feel no need to address this perennial debate.

grounded by some noumenon;[21] that phenomena alone can be *given* to us;[22] and that our minds impose *all a priori* structure, including but not necessarily limited to (spatio)temporality, on the phenomenal world. Further, any doctrine diverging from Transcendental Idealism in any way falls under the heading *Transcendental Realism*, meaning the two are mutually exclusive and exhaustive.[23] Though I do not assume here that Transcendental Idealism *holds*—whatever that may mean—I take it as a preliminary hypothesis for now.[24]

[21] *I.e.*, there cannot be *appearances* ungrounded by things that *appear*.

[22] *I.e.*, one can directly experience phenomena but not noumena—*cf.* A19-20/B33-4.

[23] *Cf.* Allison, KTI 3-73.

[24] This is precisely what Kant signifies with his famous analogy of his project to a "Copernican Revolution"—*cf.* BXVI-XXII.

— I —

# Kant's Determinism

The first task for a Transcendental Compatibilist is establishing Kant's determinism through the Analogies of Experience; this in turn presupposes his accounts, discussed in §1, of *objectivity*—more specifically, *objective validity* and *objective reality*—and what it is to be *constitutive* (what one might term "constitutivity"). After presupposing various important claims, established throughout CR, in §2, I demonstrate, in §3, that the Analogies are constitutive of experience, thereby establishing Kant's thoroughgoing determinism. Finally, I show, in §4, that Kant's resolution of the Third Antinomy demonstrates the compatibility of his determinism with transcendental freedom, thereby establishing the first claim of Transcendental Compatibilism.

## 1. Objectivity and Constitutivity

It is clear that Kant's determinism involves, on the one hand, the *relational categories*—substantiality, causality, and community[1]—and, on the other, the *Analogies of Experience*: but, insofar as Kant argues for his determinism, what does he take himself to prove regarding the relational categories and the Analogies?

One obvious candidate would be *truth*, which Kant defines as "the agreement of cognition with its object".[2] While this surely cannot be what Kant shows of the relational categories—what is it for, *e.g.*, causality to be true?—it is *prima facie* plausible that he takes himself to show it of the Analogies. However, to show that the Analogies are true is either *too much* or *not enough*. Given the transcendental distinction between phenomena and noumena, one can distinguish between *empirical* and

[1] A80/B106.
[2] A58/B82, *cf.* JL.9:49–50.

*intelligible truth*, insofar as the object is phenomenal or noumenal, respectively. Kant explicitly denies that the Analogies apply to noumena, suggesting that, in so doing, one oversteps what can be legitimately claimed about the Analogies, and so he cannot take himself to prove their intelligible truth;[3] yet, he needs to demonstrate, not empirical truth *simpliciter*, but the *necessary* empirical truth of the Analogies.[4] Therefore, truth is out.

I propose Kant takes himself to show that the relational categories—indeed, *all* the categories—are *objective* and that the Analogies—indeed, *all* the principles—are *constitutive of experience* or simply *constitutive*.[5] As I distinguish between objective *validity* and objective *reality* as two crucial aspects of objectivity *simpliciter*, I take it that Kant needs to show: first, that the relational categories are objectively valid for all phenomena; second, that the relational categories are objectively real for all phenomena, and; third, that the Analogies are constitutive principles. To do this, however, one must first understand some of Kant's basic terminology.[6]

By *object*, I mean some entity that is mind-independent in the sense that their *esse* is not *percipi*:[7] *per* Transcendental Idealism, all objects are either phenomena or noumena. By *concept*, I mean some *mark* (or collection of marks), which is itself just a part of the cognition of an object;[8] that is, a concept is simply a set of (non-contradictory) properties. Further, concepts can *contain* lower concepts under them as species of a genus and, ultimately, concepts may or may not contain objects; that is, concepts can be predicated of either objects or other concepts.[9] The relation between higher and lower concepts (and accordingly between concepts and objects instantiating the concepts) can be understood in terms of marks: the set of marks

[3] *Cf.* A146-7/B195-7, A180-1/B223, A238-9/B297-8.

[4] *E.g.*, he must establish that everything *must* be caused, not that everything just *happens* to be caused.

[5] There are several things for which a principle may be constitutive, but as I am mainly concerned in this work with whether a principle is constitutive *of experience*, I will drop the qualifier and speak simply of a principle being constitutive, unless, to avoid confusion, I explicitly do otherwise.

[6] Important disclaimer: while Kant often uses the terms as I explain them below, he is not strictly consistent across and even within works; therefore, while I take my accounts of these terms to be genuinely Kantian, I present them in an explicit attempt to enable a clearer and more precise usage than Kant's own.

[7] *Cf.* Bunch 72; Walker 76 (here I subscribe to his first sense of 'object', though not if "external world" is taken to exclusively mean the *phenomenal* world), 107-9.

[8] *Cf.* JL.9:58.

[9] One can think of objects and concepts thus differing not in *kind* but in *degree*, whereby a concept is an "underdetermined" object and an object is a "fully determined" concept.

constituting the genus is a strict subset of the set of marks constituting the species—but this has the non-trivial result that, given some concept, there is a *highest genus* but no *lowest species*.[10] One can distinguish between *essential* and *accidental* marks, such that a mark is essential to a concept insofar as the mark is an element of the set of marks making up the concept and, on the other hand, a mark is accidental to a concept insofar as the mark is essential for only some of the concept's species.[11] One can also distinguish between *pure* concepts, which can be formed *a priori*, and *empirical* concepts, which can only be formed *a posteriori* by abstracting from the particulars of experience; pure concepts can be further distinguished between *predicables*, which are derived from other pure concepts, and *categories*, which are not.[12] Finally, an *idea* is a concept which cannot, even in principle, be predicated of any objects of possible experience and is thus necessarily a pure concept.[13]

A concept's *schema* is the rule for subsuming objects under the concept;[14] as objects must be given in intuition to be subsumed under concepts, schemata can only apply to phenomena. Further, any given concept can have at most a single schema—though there may of course be additional "rules of thumb" derived from the original schema: for if, *per impossibile*, a concept has *several* schemata that could in principle conflict, there could be a contradiction in that one takes a single object to both have and not have some property—but if a concept has several schemata that cannot in principle conflict, there is no obvious sense in which there really are several schemata as opposed to equivalent formulations of a single schema.[15] *Pace* Kant, no concept, even an empirical one, is its own schema, so possession of a concept does not imply possession of its schema:[16] *i.e.*, one can *have* a concept but be unable to *apply* it.

Kant equates *objective validity* with *necessary universal validity*:[17] though some take Kant to *define* it as such,[18] I think he *argues* for this claim and so cannot be offering

[10] *Cf.* A657–9/B985–7; JL.9:59.
[11] *Cf.* JL.9.60–1.
[12] *Cf.* A79–82/B105–8; JL.9:92.
[13] *Cf.* A320/B377; JL.9:92.
[14] A138/B177, A140–2/B179–81.
[15] This argument is perfectly compatible with some concepts lacking schemata entirely—indeed, as ideas cannot be predicated of any possible phenomenon, ideas necessarily lack schemata.
[16] Chipman 43–6.
[17] P.4:298.
[18] Bunch certainly does (72), as, plausibly, does Allison (KTI 181).

a definition.[19] Although I find his argument unnecessarily vague if not outright invalid,[20] it nevertheless demonstrates he must define objective validity elsewhere.

I take it that a concept is objectively valid *for some object or class of objects* (*i.e.*, another concept) iff the object or class of objects can be thought of as possibly falling under the concept—*i.e.*, if thinking of the object as falling under the concept is compatible with thinking of the object as it actually is. As such, a concept is objectively valid for something iff the latter's marks are compatible with the former's. If a concept is objectively valid for something, it is the kind of thing that may or may not fall under the concept, while if a concept is not objectively valid for something, it cannot even be thought of as falling or not falling under the concept. The notion of objective validity is closely related to that of *category mistake*:[21] a category mistake occurs precisely when one attempts to predicate a concept of an object for which it is not objectively valid. As a concept can be objectively valid for something but fail to be objectively valid for something else,[22] Kant must establish not just that the relational categories are objectively valid for *something*, but that they are objectively valid *for all phenomena*; otherwise, any statement of his determinism might very well be a series of category mistakes.

Objective validity contrasts with objective *reality*: a concept is objectively real for some object or class of objects iff there is a rule by which one can determine whether the object or class of objects falls under the concept, at least in principle;[23] assuming a concept is objectively valid for some phenomena, it is objectively real only insofar as its schema has application,[24] so though concepts cannot be objectively real for something without also being objectively valid for it, the converse does not hold.[25] Thus the distinction between objective validity and objective reality aligns with that

[19]Kant's text suggests an argument for the biconditional claim that objective validity and necessary universal validity reciprocally imply each other: "for if…, then…, and hence…objective validity…signifies nothing other than its necessary universal validity. But also conversely, if we find cause to deem a judgment necessarily, universally valid…, we must then also deem it objective…; for…." (P.4:298).

[20]More on this later: since his argument specifically concerns the objective validity of *judgments*, I defer my discussion of the former until I have discussed the latter.

[21]*Cf.* Ryle 17-9.

[22]*E.g.*, *red* is objectively valid for *cars* but not *dreams*.

[23]*Cf.* Bunch 75, 77.

[24]*Cf.* OD.8:188-9.

[25]Ideas cannot be objectively real for any phenomenon—ideas necessarily lack any possible *real use* or application—but may nevertheless be objectively valid for some phenomena.

between *logical* and *real possibility*:[26] a concept is objectively valid for some object iff it is logically possible that the object falls under the concept; similarly, a concept is objectively real for some object iff it is really possible, *i.e.*, in accordance with the necessary conditions of experience, that the object falls under the concept.[27] As such, Kant must establish not just the relational categories' objective validity for all phenomena, but also their objective reality for all phenomena, lest the worry that the relational categories are mere ideas lacking any possible application remain.

A *judgment* is "the representation of the unity of the consciousness of various representations" by which one thinks of some object(s) or concept(s) as (not) falling under a concept,[28] *i.e.*, that through which one predicates (or does not predicate) a concept of a subject, which itself consists either of some object(s) or some concept(s).[29] A judgment is *problematic* insofar as one takes the subject to *possibly* fall under the predicate, *assertoric* or a *proposition*[30] insofar as one takes the subject to *actually* fall under the predicate, and *apodictic* insofar as one takes the subject to *necessarily* fall under the predicate;[31] as apodictic judgments can only be established *a priori*, their predicates are necessarily pure concepts.[32] Since *all* judgments require it to be possible that the predicate subsume the subject, the predicate is presupposed to be objectively valid for the subject; otherwise, the purported "judgment" would simply be a category mistake and thereby fail to qualify as a judgment *proper* for being, so to speak, "not even wrong". As such, all judgments are necessarily objectively valid and judgments are the only things that can have a truth-value.[33] Further, since propositions and apodictic judgments involve actually subsuming the subject under the predicate, they presuppose the objective reality of the predicate for the subject:

[26] *Cf.* BXXVIN, A220–1/B267–8, A244/B301–2, B308.

[27] Thus, on this distinction, Kant should discuss objective *reality* rather than *validity* at BXXVIN.

[28] JL.9:101; *cf.* CJ.5:179.

[29] Strictly speaking, this and the following are true only for *categorical* judgments, but since hypothetical and disjunctive judgments are unimportant for my argument, I set them aside here.

[30] JL.9:109.

[31] A74–5/B100; JL.9:108.

[32] A76/B101. *N.b.* This is only true for *synthetic* apodictic judgments—an *analytic* apodictic judgment can have any concept as a predicate, pure or empirical—but since the claims of the Analogies, my main concern for this Chapter, are quite plausibly *not* analytic but instead synthetic—at least, Kant surely thinks so; *cf.* A151/B190–1, A154–5/B193–4, A161–2/B201, B201–2n—the analytic/synthetic distinction is irrelevant for my purposes here.

[33] B141–2. *Cf.* JL.9:53; Allison, KTI 88, 138. Allison erroneously claims *all* judgments have truth value when only propositions and apodictic judgments do; a merely problematic judgment makes no claim at truth and so is neither true nor false.

without a rule by which the predicate can subsume the subject, there is no sense in which one can think of the subject as *actually*—and, *a fortiori*, *necessarily*—falling under the predicate in a judgment.

Occasionally, Kant distinguishes between *judgments of perception* and *experience* such that the latter predicate *objectively*, *i.e.*, of objects, while the former merely predicate *subjectively*:[34] *e.g.*, Kant claims that *sugar is sweet* really signifies *I feel a sensation of sweetness when I eat sugar*.[35] But if judgments of perception are, as Kant suggests, only subjectively valid, in what sense are they genuine judgments? The answer is that Kant draws the distinction incorrectly: the difference is not between "objective" and "subjective" predication but rather the relatively trivial one between judgments wherein *the judger* is the subject and those wherein something else is. Thus the judgment of perception *sugar is sweet* does not predicate something *of the sugar*, but rather *of myself*, meaning judgments of perception are true or false in the same way as judgments of experience: *e.g.*, I may think that I am in pain, but a knowledgeable doctor may correctly inform me otherwise.

As noted above, Kant argues that the objective validity of judgments is "interchangeable" with their *necessary universal validity* (NU-validity):[36]

> for if a judgment agrees with an object, then all judgments of the same object must also agree with one another, and hence the objective validity of a judgment of experience signifies nothing other than its necessary universal validity. But also conversely, if we find cause to deem a judgment necessarily, universally valid (which is never based on the perception, but on the pure concept of the understanding under which the perception is subsumed), we must then also deem it objective, *i.e.*, as expressing not merely a relation of a perception to a subject, but a property of an object; for there would be no reason why other judgments necessarily would have to agree with mine, if there were not the unity of the object an object to which they all refer, with which they all agree, and, for that reason, also must all harmonize among themselves.[37]

[34] P.4.298; JL.9:113.
[35] P.4:299.
[36] P.4:298.
[37] *Ibid.*

There are two possible senses of *agreement* Kant may have in mind which must be pried apart: judgments *weakly agree* about some object iff they are mutually *compatible*, *i.e.*, the truth of one does not entail the falsity of any other; separately, judgments *strongly agree* about some object iff they are *logically indistinguishable*, *i.e.*, the judgments are identical save for the identity of their judger. Accordingly, strong agreement is genuinely stronger in that it entails weak agreement, but not *vice versa*. One can then distinguish between two different senses of NU-validity: a judgment is *weakly* (*strongly*) NU-valid iff its truth entails the falsity of any judgments with which it does not weakly (strongly) agree, assuming all the judgments share a common subject.

The argument clearly holds for weak NU-validity—there is no sense in which incompatible judgments can both be true of the very same object. However, the argument for strong NU-validity requires an additional premise: *viz.*, for any object, all judgers necessarily apply concepts to it in exactly the same way regardless of any differences in relation between it and each judger. While this principle likely seemed eminently reasonable in 1783, it is not today: it rules out all forms of relativistic physics, meaning Einstein provides ample reason to reject it. *E.g.*, one of Einstein's thought-experiments demonstrates that applications of the concept of *simultaneity* depend on one's frame of reference,[38] something denied outright by the above anti-relativistic assumption. However, Einstein's example does not violate weak NU-validity: since there is a direct relationship between a judger's application of simultaneity and reference-frame, any judger can understand exactly why any other judger's judgment would differ—indeed, any judger can understand why *their own* judgment would differ were they in another reference-frame—so all judgments of simultaneity in the thought-experiment are mutually compatible and thus weakly agree.

Even if Kant *did* have strong NU-validity in mind, thereby presupposing the anti-relativistic principle, his overall system does not require it. Even in MF, the one work where it conceivably could be necessary, one could plausibly reconstruct a metaphysics of nature without it. Kant is clear MF constitutes a *special*, not a *transcendental* metaphysics, and so begins by presupposing an empirical concept, *viz.*, *matter*:[39] even he would willingly admit that, should the presupposed concept fail to accord with

[38] 29–37.

[39] MF.4:469–70, 472, 476–7.

the phenomenal world, the special metaphysics derived therefrom similarly fails, a mistake corrected by simply refining/replacing the original concept.[40]

A (theoretical) *principle*[41] is an *a priori* and thus apodictic judgment that can ground both propositions and other apodictic judgments but which is not itself derived from any judgment(s).[42] Kant distinguishes *constitutive* from *regulative principles*:[43] a principle is constitutive iff it predicates some pure concept of either *all* or *no* objects of possible experience, while a principle is regulative iff it is a subjectively necessary rule in accordance with an idea.[44] Constitutive principles establish necessary truths about the entire phenomenal world,[45] while regulative principles cannot establish *anything* about the phenomenal world.[46] Thus, beyond the objective validity and reality of the relational categories for all phenomena, Kant must establish that the Analogies are constitutive: should he fail, he cannot establish thoroughgoing determinism of the phenomenal world.

## 2. Key Presuppositions

Before turning to arguments in support of Kant's determinism, I need some claims Kant establishes throughout CR. Although these claims—and Kant's arguments for

[40]Indeed, even on the admittedly erroneous concept of matter Kant presupposes in MF, he nevertheless develops some basic relativistic machinery—*cf.* MF.4:481, 487, 555-6, 558-9—so there is no basis for taking the anti-relativistic assumption as necessary for his overall project.

[41]*N.b.* Though Kant sometimes talks of *empirical* principles (*cf.* A806/B834; MF.4:469-70), I am only concerned with what he calls *pure* or *a priori* ones. Further, one must distinguish *theoretical* principles from the *practical* ones discussed in Chapter II.

[42]A148-9/B188; JL.9:110.

[43]As noted above, my argument concerns principles constitutive or regulative *of experience* rather than something else, *e.g.*, *intuition*. Just because a principle is constitutive or regulative of experience does not imply it is constitutive or regulative of something else: as Kant explicitly states, though the Analogies are regulative of intuition, they are nevertheless constitutive of experience—A178-80/B221-3, A664/B692.

[44]*Cf.* JL.9:92. I take it that *any* apodictic judgment, not just principles, can be constitutive (regulative), and thus that apodictic judgments derived from constitutive (regulative) principles are themselves constitutive (regulative).

[45]Not all principles do so: some may not concern phenomena at all and even those that do may only require that *some* phenomena have some property.

[46]As subjectively necessary, regulative principles really concern one's own way of thinking—*i.e.*, a regulative principle states that one's thinking necessarily proceeds in such-and-such a manner—so the distinction between constitutive and regulative principles is analogous to that between judgments of experience and perception. Regulative principles, concerning ideas which, by definition, cannot apply to phenomena, necessarily involve a *transcendent* application of pure principles to noumena—*cf.* A295-6/B351-3.

them—deserve critical discussion, I must set them aside due to space constraints: I therefore simply presuppose that Kant has effectively demonstrated them here.

From the Metaphysical Deduction,[47] I presuppose Kant has established there cannot be a category not among the twelve listed candidates: *viz*.., the quantitative (unity, plurality, totality), the qualitative (reality, negation, limitation), the relational (substance, causality, community), and the modal (possibility, actuality, necessity) categories;[48] however, he has yet to establish there *are* any categories.[49]

From the Transcendental Deduction,[50] I presuppose Kant has established that at least *some* categories are required for experience to be possible;[51] further, that the quantitative and qualitative categories are objectively valid for *all* objects *tout court*,[52] because, unlike the relational categories, the quantitative and qualitative categories are constitutive of intuition.[53]

From the Schematism,[54] I presuppose Kant has established there are *schematized categories* falling under the true, *unschematized* ones;[55] since the schematized categories are species of the unschematized categories which, though pure, involve additional determinations from the *a priori* concept of time, the schematized categories are thus, strictly speaking, predicables.[56] As species of the unschematized categories, any schemata for the schematized categories also function as schemata

[47] A66–83/B91–116.

[48] A80/B106.

[49] *Cf*. Allison, KTI 133–56; Chipman 38.

[50] A84–130, B116–68.

[51] *Cf*. Chipman 38; Walker 77–8.

[52] *Cf*. Allison, KTI 159–201.

[53] Allison, KTI 199–200. Strictly speaking, this only establishes their objective validity for *phenomena*; however, since, *per* Transcendental Idealism, every phenomenon is grounded by some noumenon, this objective validity can "bridge" the transcendental divide. As the quantitative categories are objectively valid for phenomena—*i.e*., one can talk of *all* phenomena, *some* phenomena, or *this* phenomenon—one can also talk of all noumena grounding phenomena, some noumena grounding phenomena—*i.e*., all noumena grounding some phenomena—and the noumenon grounding this phenomenon; plausibly, one can then extend this from noumena grounding phenomena to noumena *simpliciter*, though this is admittedly a leap unnecessary for my argument. Further, since one can think of different noumena only insofar as they ground different phenomena, this seems to be a quality one attributes to noumena, and so one must accept the objective validity of the qualitative categories for noumena as well, at least for noumena which ground phenomena.

[54] A137–47/B176–87.

[55] *Cf*. Van Cleve 106; Allison, KTI 227.

[56] I nevertheless discuss them as if they were genuine categories.

for their unschematized counterparts.[57] Nevertheless, one must distinguish between the schematized categories and their schemata, as *all* concepts are distinct from their schemata. Further, I presuppose Kant has established that *number* and *degree* are the schemata of the schematized *quantitative* and *qualitative* categories, respectively, in general.[58]

From the Axioms of Intuition,[59] I presuppose Kant has established that the Axioms, *i.e.*, those principles subsuming all phenomena under the schematized quantitative categories, are constitutive: this in turn implies the schematized quantitative categories are objectively real for all phenomena, which itself implies the *un*schematized quantitative categories are as well.

From the Anticipations of Perception,[60] I presuppose Kant has established that the Anticipations, *i.e.*, those principles subsuming all perceptions under the schematized qualitative categories, are constitutive: this in turn implies the schematized qualitative categories are objectively real for all phenomena, which itself implies the *un*schematized qualitative categories are as well.

Finally, from the Mathematical Antinomies,[61] I presuppose Kant has established that Transcendental Realism inevitably leads to contradictions and so, as the only alternative, that Transcendental Idealism is constitutive: *i.e.*, phenomena are *not* things-in-themselves but, instead, only their appearances.[62] Thus, although Transcendental Idealism was merely a provisional hypothesis in the introduction, Kant establishes its objective validity, *i.e.*, that it is not simply a collection of category mistakes, only in the Mathematical Antinomies. This is because Transcendental Idealism requires that phenomena be grounded by noumena but Kant had not established unschematized causality's[63] objective validity for ordered pairs of noumena and phenomena:

[57]Hence 'schematized': for Kant, the unschematized categories cannot themselves have any schemata, so he introduces their schematized counterparts to provide some, albeit indirectly.

[58]A142–3/B182–3. Though the unschematized quantitative and qualitative categories are objectively valid for all objects *tout court*, as their schematized versions (supposedly) involve some temporal aspect which is itself objectively valid for all phenomena but *not* for noumena, the schematized quantitative and qualitative categories themselves are objectively valid for all phenomena but not for all noumena.

[59]A162–6/B202–7.

[60]A166–76/B207–18.

[61]A405–43/B432–71, A490–532/B518–60.

[62]*Cf.* Allison, KTF 25.

[63]*Cf.* Guyer 100–1. Guyer criticizes Kant for taking the unschematized relational category corresponding to the hypothetical form of judgment to be *causality* rather than *grounding*—but since the notion of causality Guyer discusses is only *schematized* causality and so, as he suggests, only "*one*

*i.e.*, it may be a category mistake to judge that some noumenon grounds some phenomenon. However, given that Transcendental Idealism is constitutive and it *presupposes* the objective validity of such a grounding relation, the Mathematical Antinomies therefore establish the latter by establishing the former.[64]

## 3. The Analogies of Experience

I take it Kant only establishes the schematized and unschematized relational categories' objective reality for all phenomena through demonstrating that the Analogies are constitutive. Further, since the relational categories' objective *validity*, unlike the quantitative and qualitative ones', has yet to be established for phenomena, this too is established through the Analogies: since the schematized and unschematized relational categories are shown to be objectively real for all phenomena and this presupposes that they are objectively valid for all phenomena, the latter follows from the former. But how can one prove that some principle is constitutive? Through a *transcendental argument*, *viz.*, an argument of the form: we have (some particular species of) experience; such-and-such is a necessary condition for the possibility of (this species of) experience; therefore, such-and-such is the case.[65] As Walker observes, transcendental arguments cannot establish by themselves that something is *constitutive of experience* in any meaningful sense: rather, they entail that *either* something is constitutive of experience *or else* the world just happens to be that way.[66] However, given Transcendental Idealism, the latter is rejected outright: since *all* experience is of a phenomenal world, which has structure only insofar as the mind *imposes* one upon it, structural features cannot *just happen* to hold of experience. Therefore, transcendental arguments, given Transcendental Idealism, *can* establish that principles are constitutive.

As such, the general argument structure for each triplet of an Analogy and its corresponding schematized and unschematized categories is as follows:

---

*form* of the relation of ground and consequence" (*ibid.*), *i.e.*, of unschematized causality, Kant already does as Guyer suggests: unschematized causality is distinct from schematized causality and can be understood as a relation between ground and consequent.

[64]The Mathematical Antinomies *do not*, however, establish unschematized causality's objective validity for noumena *tout court*: it may still be a category mistake to discuss one noumenon grounding another.

[65]*Cf.* Walker 10.

[66]11–2.

1. Through transcendental arguments, each Analogy is established as a necessary condition for the possibility of temporal experience;
2. Since, by Transcendental Idealism, *all* our experience is temporal, each Analogy is a necessary condition for the possibility of experience *tout court*;
3. As noted above, Transcendental Idealism entails that any necessary condition for the possibility of experience is constitutive, so each Analogy is constitutive;
4. As each Analogy subsumes all phenomena under its corresponding schematized relational category, each schematized relational category is objectively real for all phenomena;
5. As a concept's objective reality for something presupposes the concept's objective validity for it, each schematized relational category is objectively valid for all phenomena;
6. As subsuming objects under a species of some concept entails the objects also fall under the concept itself, each Analogy also subsumes all phenomena under its corresponding unschematized relational category;
7. As each Analogy subsumes all phenomena under its corresponding unschematized relational category, each unschematized relational category is objectively valid for all phenomena, and;
8. As a concept's objective reality for something presupposes the concept's objective validity for it, each unschematized relational category is objectively valid for all phenomena.

Since steps 2–8 follow straightforwardly from step 1, all one must do to accomplish Kant's objectives—prove the Analogies constitutive and the relational categories, both schematized and unschematized, objectively valid and real for all phenomena—is demonstrate that the Analogies are necessary conditions for the possibility of temporal experience; therefore, the following three subsections present transcendental arguments to establish this for each Analogy.[67]

[67] *N.b.* By no means do the following perfectly correspond to Kant's own arguments for the Analogies: though the arguments I present are certainly inspired by both Kant's and Allison's reconstruc-

## *i.* The First Analogy

The concept of *unschematized substance* is that of an object which can only exist as a subject and never as a predicate of a judgment,[68] while the concept of an *accident* is that of an object being some particular determination of another object.[69] While accidence is a relation between objects—*viz.*, that falling under the concept of accident and that in which the accident inheres—the concept of unschematized substance is a property. One can understand the distinction between these concepts as that between *adjectival* and *substantival* modes of existence:[70] for objects $x$ and $y$, $x$ exists adjectivally upon $y$, *i.e.*, $x$ falls under the concept of accident, iff $x$'s existence consists in $y$'s having some set of properties; something exists substantivally, *i.e.*, falls under the concept of unschematized substance, iff it does not exist adjectivally upon anything.[71] Thus an object's *state* is an accident inhering in the object.

An object *changes* insofar as it begins or ceases to exist, *i.e.*, *arises* or *perishes*, and *alters* insofar as it gains or loses properties;[72] if some accident changes, that in which it inheres itself either alters or changes. Since an object's state is an accident inhering in the object, any alteration therefore involves change, but, *prima facie*, there may be *absolute change*, *i.e.*, change of unschematized substance therefore ungrounded in alteration.

This, however, is precisely what the First Analogy denies: it claims that all phenomena are composed of something falling under the concept of *schematized substance*, *i.e.*, that which is *sempiternal*,[73] and that the amount of schematized substance neither increases nor decreases.[74] Since schematized substance is sempiternal, it cannot undergo absolute change, meaning the First Analogy entails that all change is

---

tions thereof, my aim is clarity, not accuracy in reconstruction—*cf.* A182–218/B224–65; KTI 229–74; see also Weizsäcker.

[68] *Cf.* B149, A147/B186, B288.

[69] A186–7/B229–30. Accidents *inhere* in something; substances merely *subsist*.

[70] Van Cleve 105.

[71] *N.b.* That $x$ is an accident inhering in $y$ does not entail $y$ is a substance; there may be some $z$ such that $y$ inheres in $z$, *etc.*

[72] A187/B230–1.

[73] A144/B183. *N.b.* By 'sempiternal', I do not mean that schematized substance has an *infinite* duration—this would contradict Kant's resolution to the First Antinomy (*cf.* A426–33/B454–61, A508–23/B536–51)—but rather that it persists throughout all time.

[74] *Cf.* A182/B224.

only *relative*, *i.e.*, mere alteration in unchanging, schematized substance. My transcendental argument for the First Analogy is as follows:[75]

1. Temporal experience is possible only if experience of objective change is possible.[76]
2. Experience of objective change is possible only if an experience of objective change is in principle distinguishable from a subjective change of experience of unchanging objects.
3. An experience of objective change is only distinguishable from a merely subjective change of experience if the object undergoing objective change is an accident in something—call it *pseudo-substance*—persisting throughout the change, *i.e.*, if the objective change is mere alteration in some pseudo-substance.[77,78]
4. From 2 and 3, one cannot experience an objective change in some object falling under pseudo-substance unless it is an accident inhering in some distinct object that itself falls under pseudo-substance, *etc.*; thus one can "step back", as it were, from some accident to underlying pseudo-substance repeatedly.[79]

[75]Although most of the obvious work towards establishing Kant's determinism is performed by the Second and Third Analogies, the Second Analogy (and, by extension, the Third Analogy as well) cannot even be stated without the results of the First Analogy, so one must establish that it is constitutive—*cf.* B232-3; PM.20:329-30.

[76]*I.e.*, there is no sense in which one's experience is temporal if it is impossible for one to experience a change among objects.

[77]If, from one moment to the next, everything I experience was replaced with something else, I could not tell whether I experienced objective or merely subjective change, *i.e.*, whether the objects *themselves* underwent change, and nothing about my epistemic situation improves if some object unrelated to the change I am investigating persists throughout the change. My epistemic situation improves only if some object *related* to the change persists throughout the change: if $x$ was an accident inhering in $y$ and, though I still experience $y$, I no longer experience $x$, I know $x$ perished—*mutatis mutandis* for arising.

[78]*N.b.* Though I take this to be true and, further, that Kant does as well, some might question it and thereby place all the Analogies in doubt; this, therefore, is one of the key premises of Kant's determinism. Since this work merely explores Kant's thought, I simply assume its truth here.

[79]This "stepping back" is not unfamiliar in ordinary life: take, *e.g.*, an apple. One can step back from the apple to the cells composing the apple, organelles composing the cells, molecules composing the organelles, atoms composing the molecules, sub-atomic particles composing the atoms, and even mass-energy composing the sub-atomic particles—indeed, future breakthroughs in physics may enable even further steps.

5. This process of stepping back is finite; the ultimate, underlying object has no essential properties and thus falls under the concept of *unschematized substance*.[80]

6. Since objects falling under unschematized substance ground all possible experience of objective change, they persist throughout all possible change; thus objects falling under unschematized substance are sempiternal, *i.e.*, fall under both *unschematized* and *schematized* substance.[81]

7. Since objects falling under substance are sempiternal, they neither arise nor perish, *i.e.*, the amount of substance neither increases nor decreases.[82]

8. Any object of possible experience either can or cannot change: if it can change, it must be an accident inhering in substance; if it cannot change, it *is* substance. Therefore, all objects of possible experience are composed of substance.[83]

9. Thus temporal experience is possible only if all objects of possible experience are composed of substance such that the quantity of substance neither increases nor decreases, *i.e.*, only if all change is mere alteration in substance.

[80]In each step back from accident to underlying pseudo-substance, one necessarily abstracts away from some of the essential marks of the accident's concept and renders them accidental marks in the pseudo-substance's concept—otherwise, there is no sense in which the accident inheres in that particular pseudo-substance. (For Kant's views on abstraction, see A.7:131.) But as there can only be *finitely* many essential marks in any accident's concept—otherwise one would be unable to cognize the accident as falling under its concept—one must reach a pseudo-substance with *no* essential properties, all having been abstracted away, within some finite number of steps back: since an object with no essential properties falls under unschematized substance, the process of stepping back necessarily ends with an object falling under unschematized substance.

[81]I drop the qualifiers 'schematized' and 'unschematized' for the remainder of the argument: something falls under substance iff it falls under *both* schematized *and* unschematized substance.

[82]Since the quantitative categories are objectively valid for all phenomena, this is not a category mistake.

[83]The first conditional follows straightforwardly from the argument thus far; the second conditional is slightly more complex. If an object of possible experience cannot change, it is sempiternal and thus falls under schematized substance—but does it fall under unschematized substance as well? Yes: if it cannot even in principle change, there is no sense in which its existence consists in some distinct object's having some property—as such, it does not exist adjectivally upon anything and so is unschematized substance.

Thus the transcendental argument establishes that the First Analogy is a necessary condition for the possibility of temporal experience; by the rest of the general argument structure above, schematized and unschematized substance are both objectively valid and real for all phenomena and, further, the First Analogy is constitutive.

### *ii.* The Second Analogy

*Unschematized causality* is the relation between ground and consequent such that the consequent necessarily follows from the ground, *i.e.*, the existence of the ground entails the existence of the consequent; *schematized causality* is the relation between cause and effect such that the effect necessarily succeeds the cause *temporally*.[84] The Second Analogy states that all alteration in the phenomenal world takes place in accordance with laws of schematized causality and since, by the First Analogy, all change is mere alteration in substance, the Second Analogy therefore entails that *everything that happens* in the phenomenal world takes place in accordance with causal laws.[85] My transcendental argument for the Second Analogy is as follows:

1. Temporal experience is possible only if experience of the objective succession of states in an alteration is possible.[86]
2. Experience of the objective succession of states in an alteration is possible only if one can distinguish an experience of the objective succession of one state after the other from an experience of the reversed objective succession of the other state after the one; *i.e.*, only if one can distinguish an experience of *A*-*B* from one of *B*-*A*, where *A* and *B* are states of some substance.
3. One cannot perceive time and so cannot tell the objective order of succession of the states in an alteration through mere perception.
4. There is an objective fact concerning the order of succession of the states.

[84] *Cf.* A112, B162–3, B168, A144/B183, A243–4/B301–2.

[85] *Cf.* A189/B232.

[86] *I.e.*, since an alteration is a change of states, an experience of the objective succession of states in an alteration is the experience that one state succeeds the other, not just *subjectively*, but *objectively*, or in the object.

5. For one to distinguish between an experience of objective succession and an experience of the reversed objective succession, there must be some rule through which one orders the states.[87]
6. Thus temporal experience is only possible insofar as one applies these rules.
7. Since these rules are conditions of temporal experience, they cannot be acquired *a posteriori* by observing the objective succession of states and abstracting rules from observations.
8. Therefore, such rules must already be held and applied to phenomena *a priori*—meaning they are among the conditions imposed by the mind on phenomena—and thus apply necessarily: in any alteration, there is a rule by which one state necessarily succeeds the other in accordance with the rule.
9. Such rules are simply rules of schematized causality, *i.e.*, *schematized causal laws* in accordance with which alterations take place.
10. Since the existence of the effect necessarily follows from the existence of the cause, such schematized causal laws are also *un*schematized causal laws.[88]
11. The cause of some particular alteration must have arisen at some time: had it always existed, the effect would always have existed, too, and there would have been no alteration whatsoever.
12. Since the cause must have arisen and, *per* the First Analogy, all change is mere alteration, the cause of any alteration must itself be the effect of some earlier alteration.

[87]One can distinguish between strong and weak senses of ordering states such that, according to the strong sense, the rule is the *reason* why the states have their order, while, according to the weak sense, the rule is merely a *tool* by which one cognizes their order. Since, from Transcendental Idealism, phenomena are temporal while noumena are not and, further, phenomena simply are noumena when one's mind imposes certain conditions upon them, the rules must order objective states in the strong sense: they are simply the mind's own rules by which objects of experience are placed into a temporal order. These rules need not be applied *consciously*: they will shortly be shown to be merely a subset of the conditions one's mind imposes on phenomena, and Transcendental Idealism is not committed to the claim that these conditions are imposed *knowingly* on phenomena, only that they are imposed on phenomena *at all*.

[88]I therefore drop all qualification and simply talk of causal laws *simpliciter*.

13. Accordingly, temporal experience is possible only if *every* state of *every* phenomenon changes in accordance with causal laws, *i.e.*, everything that happens in the phenomenal world does so in accordance with causal laws.

Thus the transcendental argument establishes that the Second Analogy is a necessary condition for the possibility of temporal experience; by the rest of the general argument structure above, schematized and unschematized causality are both objectively valid and real for all phenomena and the Second Analogy is constitutive. Moreover, insofar as causal laws are thus constitutive of the phenomenal world, they can be considered *laws of nature*.[89]

Before continuing to the Third Analogy, one must note that the above argument concerns *succession* and not specifically *strict succession*: effects are only required by causal laws to *weakly succeed* their causes, *i.e.*, effects can arise no earlier than their causes,[90] even if some effects do strictly succeed their causes.

### *iii.* The Third Analogy

*Unschematized community* is the reciprocal determination, *i.e.*, reciprocal unschematized causality, of several objects; accordingly, *schematized community* is reciprocal schematized causality of several objects.[91] The Third Analogy states that phenomena are simultaneous only insofar as they are in schematized community. My transcendental argument for the Third Analogy is as follows:

1. Temporal experience is possible only if experience of the objective simultaneity of distinct objects is possible.
2. Experience of the objective simultaneity of distinct objects is possible only if one can distinguish between merely subjective alteration of experience among distinct objects and experience of objective alteration of a single object.[92]

[89] *Cf.* A114, A125–8, B164–5, A216/B263, A418–9/B446–7; MF.4:467.
[90] *Cf.* A202–4/B247–9.
[91] *Cf.* B112–3, A144/B183–4, A244/B302.
[92] *I.e.*, one must be able to distinguish between (i) looking at one object and then another and (ii) a single object changing every time one *takes* oneself to look at something else.

3. *Per* the Second Analogy, all objective alteration takes place in accordance with causal laws such that the effect weakly succeeds the cause.
4. One can distinguish between *direct* and *indirect causality*: a causal relationship is direct iff there are no intervening "links" between cause and effect, while a causal relationship is indirect iff there is some intervening chain of direct causal relations which connect cause and effect.[93]
5. Weak succession is transitive.[94]
6. Therefore, an effect weakly succeeds all its direct and indirect causes.
7. Simultaneity is equivalent to reciprocal weak succession.[95]
8. One can distinguish merely subjective alteration of experience among several objects from experience of a single object's objective alteration only if each object determines the states of the others through either direct or indirect causal relations.[96]
9. Thus the states of the objects reciprocally weakly succeed each other and are thereby simultaneous.
10. This reciprocal causal relation is simply schematized community: the community is *immediate* or *mediate* insofar as the reciprocal causal relation is direct or indirect, respectively.
11. Therefore, temporal experience is only possible if all simultaneous objects are in community, either immediate or mediate.

[93] *I.e.*, the Second Analogy concerns direct causal relations as the existence of the effect follows from the existence of the cause. An indirect causal relationship, however, is one in which the effect's existence does not follow from the cause's *immediately*, but through a chain of direct causal relationships.

[94] *Cf.* Allison, KTI 273.

[95] *I.e.*, if some $x$ arises no later than $y$ and $y$ arises no later than $x$, $x$ and $y$ are simultaneous.

[96] Were there *no* causal relations, direct or indirect, among the objects, one could make no claims of objective weak succession among their states and so one could have no experience of their objective simultaneity; were there *unidirectional* causal relations, direct or indirect, the effects would objectively weakly succeed the causes, but this is consistent with the alteration being objective rather than merely subjective: therefore, subjective alteration of experience can be distinguished from experience of objective alteration only if objects are in reciprocal causal relations.

Thus the transcendental argument establishes that the Third Analogy is a necessary condition for the possibility of temporal experience; by the rest of the general argument structure above, schematized and unschematized community are both objectively valid and real for all phenomena and the Third Analogy is constitutive.

❧ ❧ ❧

With all three Analogies shown to be constitutive, one can understand Kant's determinism as consisting of the resulting *Law of Natural Causality* (LNC), which requires that *all* phenomena be causally connected to each other, either directly or indirectly; since the Analogies are constitutive and they entail LNC, LNC is constitutive as well.[97]

There are two key points to note about LNC: first, LNC neither addresses nor is intended to address the problem of induction; LNC only guarantees that phenomena *are* connected by causal laws, not that these laws can be *identified*.[98] Kant's response to the problem of induction only comes in CJ: *per* a merely regulative principle, the power of judgment necessarily makes a transcendental presupposition that the phenomenal world is *purposive* to our cognitive faculties and, *a fortiori*, we can identify natural laws—while we can have no guarantee of the *truth* of this transcendental presupposition, we have no choice but to act *as if* it were true as it follows from a regulative principle.[99] Second, LNC describes a complete, thoroughgoing determinism: since the Analogies are constitutive, if some phenomenon were, *per impossibile*, underdetermined by natural laws or not causally related to all other phenomena, it would be impossible, even in principle, to determine such an object's objective duration and it would therefore not be an object of possible experience, *i.e.*, part of the phenomenal world.

## 4. Resolution of the Third Antinomy

The Third Antinomy is the contradiction that arises on the supposition of Transcendental Realism between, at base, a world with freedom and a world without it;

[97]However, that means LNC is not itself a principle—hence the *law* of natural causality, not the *principle* of the same.

[98]*Cf.* Allison, KTI 259.

[99]*Cf.* CJ.5:184; FI.20:208–9.

although, for my purposes, Kant's arguments for this contradiction are irrelevant, his *resolution* of the contradiction is vital for establishing Transcendental Compatibilism. Kant starts with the notion of *transcendental freedom*, or the capacity for something to begin a state from itself, *i.e.*, independently of being so determined by natural laws.[100] As transcendental freedom necessarily entails the ability to do otherwise—indeed, this seems to be one of Kant's motivations for introducing it at all[101]—it can be considered a libertarian-style notion of freedom.

If, *per impossibile*, Transcendental Realism were true and therefore there were no distinction between phenomena and noumena but somehow LNC were nevertheless still constitutive, standard incompatibilist arguments establish there could be no transcendental freedom;[102] however, since Transcendental Idealism is constitutive of experience, transcendental freedom can be shown to be compatible with LNC through the distinction between *empirical* and *intelligible character*.[103] *Per* Transcendental Idealism, all phenomena must be grounded by noumena, so there is no contradiction implicit in noumena, themselves not subject to LNC, grounding phenomena which are.[104] Therefore, while an object's *empirical character* describes its phenomenal actions in accordance with LNC, its *intelligible character*, should it have one, is the noumenal ground of its empirical character. Since such an intelligible character, as noumenal, is not subject to LNC but, as an intelligible character, has particular phenomenal effects, *viz.*, the formation of some or other empirical character and thereby the performance of some or other phenomenal actions, an object with such an intelligible character is therefore transcendentally free.[105]

Not just *any* object can have an intelligible character: otherwise, Kant must accept the absurd conclusion that *all* things, as grounded by noumena, are transcendentally free. Since intelligible characters can only be determined through intelligible grounds, *viz.*, *reasons*,[106] only *rational beings* can possibly have intelligible char-

[100] *Cf.* A444-6/B4724, A448/B476, A533/B561. *N.b.* This is not a *temporal* beginning but a *causal* beginning, *i.e.*, an uncaused cause—*cf.* A450/B478.
[101] A334/B362.
[102] *Cf.* A534/B562, A536/B564.
[103] *Cf.* Allison, "Antinomy" 23.
[104] A536-7/B564-5; *cf.* PM.20:291-2.
[105] A539-41/B567-9.
[106] *Cf.* A545-6/B573-4.

acters.[107] But if a rational being has an intelligible character, its empirical character is grounded, independently of its determination in accordance with natural laws, by its intelligible character; even though its empirical character is determined in accordance with LNC, it is nevertheless *also* transcendentally free. By extension, since all of the rational being's actions are grounded by its transcendentally free empirical character, all of its actions are therefore transcendentally freely performed; such a being, considered phenomenally, would be *both* completely determined in accordance with LNC *and* transcendentally free in all its actions.[108]

Thus there are three key conclusions to draw: first, the concept of transcendental freedom is objectively valid only for rational beings, though it is yet to be established whether it is objectively real for them. Second, transcendental freedom is completely compatible with LNC and therefore Kant's determinism: accordingly, the first claim of Transcendental Compatibilism is established. Third, there is not nor can there be any theoretical warrant concerning whether rational beings are transcendentally free. Such warrant would either be an *a priori* guarantee that all or no rational beings are transcendentally free or else *a posteriori* evidence that adds or removes subjective grounds for taking a particular rational being to be transcendentally free, but neither kind of warrant is available: since transcendental freedom is (i) *compatible* with LNC but (ii) *not constitutive*, it is possible but not guaranteed that there be transcendentally free rational beings; further, since LNC is constitutive for both transcendentally free and any *unfree* rational beings, the acts of transcendentally free rational beings will be in accordance with natural laws just as those of unfree rational beings are, so nothing in the observable causal structure of a rational being's action can give any evidence either for or against that rational being's transcendental freedom.

[107] This is not to say that *all* rational beings have intelligible characters, but that rationality is a prerequisite.

[108] *Cf.* A536/B564; Allison, KTF 74.

— II —

# Transcendental Compatibilism

As the second and third propositions of Transcendental Compatibilism concern morality, §1 and §2 develop the core of Kant's moral theory and demonstrate the Singularity of Freedom—a claim which entails that transcendental freedom is both necessary and sufficient for moral responsibility—thereby establishing the second proposition of Transcendental Compatibilism; §3 then examines Kant's doctrine of the Fact of Reason and concludes that rational beings under normal circumstances are both free and subject to the moral law, thereby establishing the third proposition of Transcendental Compatibilism. By responding to three potentially fatal objections, §4 shows that, given Transcendental Compatibilism, Kant can have an account of ordinary goodness for ordinary people.

## 1. The Basic Components of Kant's Moral Theory

A *practical principle*, to be distinguished from the *theoretical* ones of the previous Chapter, is a practical judgment that can ground an *action*;[1] they can be moral, but also merely technical or pragmatic.[2] Practical principles do not denote causal relations, but provide *reasons* to perform actions.[3] As rational beings alone can act from reasons,[4] only rational beings can be subject to practical principles—though it should be noted that some actions performed by rational beings, *e.g.*, tripping, are not grounded by practical principles.

[1] *Cf.* G.4:393, 399–400, 407, 412–3; PP.8:381.
[2] *Cf.* G.4:416–7; CPR.5:20; MM.6:409; A.7:322; LP.9:449, 455.
[3] *Cf.* Korsgaard 57.
[4] §I.4.

A practical principle can be either *subjective* or *objective*: it is subjective, or a *maxim*, insofar as a rational being takes it to apply to itself, and objective, or a *law*,[5] insofar as it actually applies to some class of rational beings.[6] Analogously to the distinctions between objective and subjective theoretical judgments and constitutive and regulative theoretical principles, maxims and laws have the same general *form* but different *subjects* in the practical judgment: the subject of a maxim is 'I' while the subject of a law is 'every member of class $K$', where $K$ is some (possibly non-strict) sub-class of rational beings. Therefore, though I only specifically discuss the form of maxims in what follows, the discussion applies equally to laws so long as one substitutes 'I' for 'every member of class $K$'.

The general form of a maxim is: *I may, in circumstance $C$, perform deed $D$ to advance end $E$*—so one can denote a maxim $M$ as the ordered triplet $M = \langle C, D, E \rangle$. One of Kant's example maxims can therefore be reformulated as "I may, when a lying promise is likely or expected to be believed, give a lying promise for my own financial gain".[7] To better explain the form of maxims, I must discuss each of the three components—circumstance, deed, and end—as well as the meaning of 'may'.[8]

The notion of circumstance as it can function in a maxim is not a fully determinate state of affairs but rather a *constraint* or *requirement* on the set of possible circumstances to which the maxim can apply; as such, the role of circumstance in a maxim is really a rule which determines in which states of affairs one takes the maxim to apply to oneself. Though a maxim may be stated informally without specifying any such rule, the rule can nevertheless be implicitly defined, at the very least as those states of affairs in which the deed is possible—and if the circumstance of a maxim is genuinely "left blank" or "left open", this implies the maxim is held to apply to oneself in *all* circumstances.[9] Finally, since the circumstance of a maxim is really a rule for the application of a maxim, one can only take the maxim to apply insofar as one *takes* one's present state of affairs to fall under the rule; thus one cannot act on a

[5] *N.b.* To avoid any confusion with the deterministic laws discussed in Chapter I, I refer to objective practical principles as *laws* and the causal relations of LNC as *natural laws* or *laws of nature*.

[6] *Cf.* G.4:401n1, 421n1; CPR.5:19; MM.6.225.

[7] *Cf.* G.4:403, 422.

[8] I refer to maxims as *belonging* to a rational being insofar as it takes them to apply to itself, *i.e.*, it is tautologous to say that one takes one's maxims to apply to oneself—*cf.* O'Neill, AP 13, 16, 62, 97.

[9] Similarly, if the deed or end components of a maxim are left open, the maxim is taken to apply to *any* deed or *any* end.

maxim whose circumstance is such that one cannot even in principle know that one is in such a circumstance, *e.g.*, "when I am being successfully misled".

A *deed* is an action which can be *imputed* to the rational being that performs it;[10] a *mere action*, though perhaps performed by a rational being, cannot be imputed to it. Note that a deed need not be *consciously* performed: though *unconscious* actions, *e.g.*, sleepwalking, are clearly not imputable to any rational being, *subconscious* ones, *e.g.*, acting on implicit bias, plausibly can be.[11] I will shortly show, but stipulate for now, that an action is a deed iff it is performed from a maxim, and so the "action" component of maxims is not just an *action* but specifically a *deed*.

An *end* is some aim or goal that a rational being may have, either *positively* insofar as it is something to be achieved or attained, or *negatively* insofar as it is something to be preserved and not subverted or undermined, so far as one can help it;[12] overall, one can say that a rational being aims to *advance* its ends. This is not to say that a rational being aims to advance some or all of its ends *at all costs* and therefore *never* subverts *any* end: there may be—indeed, frequently are—situations in which some ends must be subverted for the sake of advancing other ends that one values more highly, but this does not imply that ends subverted in these cases are not genuine ends. A *means to an end* is itself an end set for the sake of advancing some further end, while a *mere means* is a means set *only* for the sake of advancing some further end and, as such, is not in itself taken to be an end.[13] The deed of a maxim is, by stipulation, a means to the maxim's end and thus the deed is *itself* an end that must be set; eventually this will "bottom out", so to speak, into ends so basic that one need not (or perhaps cannot) specify any means to them.[14] Thus, since an action, as stipulated above, is a deed iff it is performed from a maxim and, further, all maxims have ends, an action is a deed insofar as it has an end.[15]

The 'may' in a maxim expresses *permissibility*: this can be moral permissibility, but it can also be a notion of permissibility in some non-moral practical sphere.[16]

[10] *Cf.* MM.6:223-4, 227-8, 437-8; R.6:21, 31.

[11] *Cf.* O'Neill, CoR 84.

[12] Korsgaard 16.

[13] *Cf.* G.4:427-8; MM.6:381.

[14] *E.g.*, the means to typing out this sentence is moving my fingers in a particular pattern, but I cannot specify any means that I take to moving my fingers.

[15] *Cf.* MM.6:384-5.

[16] *E.g.*, cooking: a recipe may call for some ingredient but permit its substitution with something else.

Most commentators seem to take it that the form of maxims includes 'will' rather than 'may': call this the standard view.[17] I take it that my view is better than the standard view for several reasons. First, Kant's own examples of maxims frequently use 'may' or related terms.[18] Second, maxims on my view have more obviously *moral* (or, more generally, *practical*) content than on the standard view, on which maxims seem more like predictions than moral (practical) principles: just as, by Hume's Guillotine, one cannot derive an *ought* from an *is*, one cannot derive an *ought* from a *will be*.[19] Third, given that nobody questions that, for Kant, an imperative is a formulation of a law, *i.e.*, an *objective* practical principle, that can always be expressed through an *ought*,[20] it makes much more sense for a maxim, *i.e.*, a *subjective* practical principle, to also be expressed through deontic notions—otherwise, in what sense are maxims and laws both practical principles? Fourth, though it seems *prima facie* possible to take some maxim to apply to oneself and yet never act from it, the standard view has difficulty accounting for this possibility and must make maxims highly complicated counterfactuals—to account both for the possibility that the maxim's circumstances never arise and for the possibility that, even if its circumstances do arise, one acts instead on some *other* maxim—whereas it presents no trouble for my view. Fifth, the standard view cannot clearly distinguish between two principles—the Formulae of Universal Law and of Laws of Nature—Kant takes to be distinct, while this is easily done on my view: the Formula of Universal Law concerns the *law* corresponding to a maxim—*i.e.*, the objective practical principle applying to all morally responsible beings with the same circumstance, deed, and end as the maxim—whereas the Formula of Laws of Nature concerns a hypothetical *law of nature* in accordance with which all morally responsible beings would universally adopt the maxim, *i.e.*, the *type* of the maxim.[21] Sixth, Kant thinks that the moral worth of deeds is derivative upon the moral worth of the maxims from which they are performed, but if, as is

[17] *Cf.* Rawls, *Lectures* 168; O'Neill, CoR 97. Plausibly, O'Neill suggests that 'will' might best be interpreted as 'may' in many cases—*cf.* AP 55-6, 99—so it is unclear whether the standard view actually differs from my own, but this has little consequence for my argument: if there is no disagreement, my proposed mode of presentation is clearer and more precise, so the standard view ought to be abandoned; if there *is* disagreement, the standard view is incorrect and so ought to be abandoned.

[18] *Cf.* G.4:402, 403, 422 (here 'could' is most naturally taken as equivalent to 'may'); CPR.5:69; TP.8:304-5.

[19] *Cf.* A547/B575, A550/B578.

[20] G.4:413—more on this shortly.

[21] *Cf.* G.4:402, 421; CPR.5:69.

the case on the standard view, maxims are ultimately logical constructions of deeds, this raises the worry, which does not arise on my view, that Kant's view might be circular.

Although all maxims have the components and overall form as described, there are four logical *qualities* of maxims which can be distinguished analogously to Kant's distinction from the Table of Judgments.[22] (For a more precise notation, I use '$x \preceq y$' to denote the relation '$x$ may $y$', '$x \not\preceq y$' to denote 'it is not the case that $x$ may $y$, and '$\neg$' to denote 'not'.) An *affirmative* maxim has the exact form described above—*I may, in C, do D for E*, or, alternatively, $I \preceq \langle C, D, E\rangle$—and expresses that one takes one's doing $D$ for $E$ to be permissible in $C$. A *negative* maxim is the contradictory of its corresponding affirmative maxim—*it is not the case that I may, in C, do D for E*, or, alternatively, $I \not\preceq \langle C, D, E\rangle$—and expresses that one takes one's doing $D$ for $E$ to be impermissible, *i.e.*, *forbidden* or *prohibited*, in $C$. An *infinite* maxim is the subcontrary of its corresponding affirmative maxim—*I may, in C, refrain from doing D for E*, or, alternatively, $I \preceq \neg\langle C, D, E\rangle$—and expresses that one takes one's not advancing $E$ through $D$ to be permissible in $C$.[23] Finally, a *negative infinite* maxim is the contradictory of its corresponding infinite maxim—*it is not the case that I may, in C, refrain from doing D for E*, or, alternatively, $I \not\preceq \neg\langle C, D, E\rangle$—and expresses that one takes one's not advancing $E$ through $D$ to be impermissible in $C$, *i.e.*, that doing $D$ for $E$ is *required* or *obligatory* in $C$. As such, one can derive straightforward relationships between 'may' and 'ought', where the latter expresses an obligation or duty that can be either positive, *i.e.*, a requirement, or negative, *i.e.*, a prohibition.[24] Thus a lack of obligation is equivalent to the conjunction of an affirmative maxim and its corresponding infinite maxim,[25] a positive obligation is equivalent to the conjunction of an affirmative maxim and its corresponding negative infinite maxim,[26] and a negative obligation is equivalent to the conjunction of a negative maxim and

[22] *Cf.* A70/B95; JL.9:103-4.

[23] *N.b.* An infinite maxim leaves open six possibilities: whether $I \preceq \langle C' \neq C, D, E\rangle$; whether $I \preceq \langle C, D' \neq D, E\rangle$; whether $I \preceq \langle C, D, E' \neq E\rangle$; whether $I \preceq \langle C' \neq C, D' \neq D, E\rangle$; whether $I \preceq \langle C' \neq C, D, E' \neq E\rangle$; and whether $I \preceq \langle C, D' \neq D, E' \neq E\rangle$. (I ignore the seventh "possibility" of whether $I \preceq \langle C' \neq C, D' \neq D, E' \neq E\rangle$ since this is always true.) However, since, *per* OIC, one may do something for some end regardless of one's present circumstances, it follows that at least one of $I \preceq \langle C, D' \neq D, E\rangle$, $I \preceq \langle C, D, E' \neq E\rangle$, and $I \preceq \langle C, D' \neq D, E' \neq E\rangle$ is true.

[24] To expand the notation, I use '$x \prec y$' to denote the relation '$x$ ought to $y$'.

[25] *I.e.*, $I \preceq \langle C, D, E\rangle \wedge I \preceq \neg\langle C, D, E\rangle \leftrightarrow I \not\prec \langle C, D, E\rangle$.

[26] *I.e.*, $I \preceq \langle C, D, E\rangle \wedge I \not\preceq \neg\langle C, D, E\rangle \leftrightarrow I \prec \langle C, D, E\rangle$.

its corresponding infinite maxim;[27] since the conjunction of a negative maxim and its corresponding negative infinite maxim entails conflicting obligations—*viz.*, one ought to do $D$ for $E$ in $C$ and ought *not* to do $D$ for $E$ in $C$—OIC rules out this possibility.[28]

A rational being's *will* is its capacity to act from its maxims;[29] all rational beings have wills.[30] The ability to *choose* presupposes that a being has a will: a being can only be said to choose an option if the being has some reason for choosing it, although not necessarily one of which it was consciously aware in so choosing. This is not to say that choice is possible only if reasons favor a *unique* option but merely that there must be *some* reason for choosing the option that is ultimately chosen, regardless of whether there are also reasons—perhaps even the very same reasons—for choosing another option; even in a case of arbitrary choice, one can only be said to choose one of the equally attractive options insofar as there is a reason why one so chooses.[31] Since, as noted above, one acts on some reason insofar as one acts from some practical principle—and, insofar as one takes the reason to apply to oneself, this practical principle must be a maxim—choice is only possible in accordance with maxims; thus the will is a rational being's power of choice.[32] This thus vindicates my earlier stipulation that an action is a deed iff it is performed from a maxim: since choice is only possible in accordance with maxims and an action can only be imputed to an agent insofar as the agent chooses to do it, an action can only be imputed to an agent, and thus be a deed, insofar as it is done from a maxim.

A practical principle *determines* one's will iff it is sufficient to ground one's choice—thus a maxim or law determines one's will iff one cannot choose except from the maxim or law—while a law *necessitates* one's will insofar as it applies to oneself but is insufficient to determine one's will.[33] Since a law that applies to a rational being is nothing other than an objective reason valid for that being, one can only violate

[27] *I.e.*, $I \not\preceq \langle C, D, E\rangle \wedge I \preceq \neg\langle C, D, E\rangle \leftrightarrow I \prec \neg\langle C, D, E\rangle$

[28] *I.e.*, $(I \not\preceq \langle C, D, E\rangle \wedge I \not\preceq \neg\langle C, D, E\rangle) \leftrightarrow (I \prec \langle C, D, E\rangle \wedge I \prec \neg\langle C, D, E\rangle) \rightarrow \bot$

[29] *Cf.* G.4:412–3, 446.

[30] Since one's maxim is simply the judgment that some practical principle applies to oneself, no rational being can act *directly* from consciousness of some law except insofar as it takes the law to *apply* to it, *i.e.*, a rational being can only act from a law when it is mediated by a maxim.

[31] *E.g.*, "I need to choose *something* and they are all equally good options, so I might as well choose $x$".

[32] This is setting aside the *Wille*/*Willkür* distinction: though it is important, it arises naturally through the combination of my basic concepts, so I do not use it throughout.

[33] G.4:412–3.

a law that necessitates one's will by violating a *rational requirement* and so this is only possible for rational beings for whom choice is not determined by reason alone: Kant takes it that *all finite rational* beings are like this and, as such, a law can only necessitate the will of a finite rational being, while laws are necessarily determining for any infinite rational beings.[34] Insofar as a law applies to a being, *i.e.*, either determines or necessitates its will, the law is a *command* for it and, insofar as a command only necessitates a being's will, it is formulated through an *imperative*: since the command only necessitates and does not determine a being's will, the corresponding imperative expresses that, though it is possible that the being not comply, it *ought* to comply nevertheless.[35] Therefore, commands necessarily apply to finite rational beings through imperatives—and, further, can only be complied with insofar as the being takes the command to apply to itself, *i.e.*, insofar as the being has the command's corresponding maxim—while it applies "directly" to any infinite rational beings.

Kant takes it that genuine *moral responsibility* requires more than mere libertarian-style independence from natural laws, so though transcendental freedom is *necessary* for moral responsibility, it is not of itself sufficient:[36] it is merely a *negative* determination of the kind of freedom necessary and sufficient for moral responsibility, which Kant calls *practical freedom*. The *positive* determination of practical freedom is *autonomy*, *i.e.*, a being is morally responsible iff it is both transcendentally free and only commanded by laws that it gives itself.[37] Since transcendental freedom is conceptually distinct from autonomy—there can plausibly be a transcendentally free being that is *heteronomous*, *i.e.*, commanded only by laws given by an external authority—negative and positive freedom are conceptually distinct. While it is sufficient for my purposes here to simply assume Kant's identification of moral responsibility and practical freedom, *i.e.*, transcendental freedom and autonomy, his view, though unorthodox, has its merits. Essentially Kant's claim boils down to the

[34] *Cf.* G.4:414, 439; CPR.5:25, 32–3, 82, 123–4.

[35] G.4:413.

[36] *E.g.*, it is *prima facie* plausible that a transcendentally free being's actions are determinable independently from natural laws only because they are completely determined by the arbitrary commands of another transcendentally free being: though, ultimately, its actions are determined independently from natural laws and so are transcendentally free, such a being is little more than a puppet which, accordingly, cannot be considered morally responsible.

[37] *Cf.* A533–4/B561–2; G.4:446–7, 452–3; CPR.5:29–30, 31, 33, 42–3, 47, 48, 87, 95–7, 132, 133–4; MM.6:213–4; CF.7:27; WE.8:35; OT.8:145–6, 146n. Since practical freedom presupposes transcendental freedom and, as demonstrated in §I.4, transcendental freedom is only objectively valid for rational beings, practical freedom can only be objectively valid for rational beings as well.

Rousseauian claim that the sense of freedom relevant for moral responsibility is that by which one is subject only to those authorities to which one freely submits: if a being is "automatically" subject to any authority *other than its own*—*e.g.*, some person or group, biological facts, a deity—then it ultimately answers not to *itself* but to this foreign authority and thereby fails to be practically free.

An imperative can be either *categorical* or *hypothetical*: a hypothetical imperative requires that a (finite) rational being act from its corresponding law conditional upon being in particular circumstances or having particular ends, while a categorical imperative requires that a (finite) rational being act from its corresponding law *tout court*, *i.e.*, without reference to circumstances or ends.[38] Thus only categorical imperatives express laws which leave both circumstance and end completely open. Following O'Neill,[39] I take it that, for Kant, the *Principle of Hypothetical Imperatives* (PHI)—that willing some end rationally involves willing all its necessary means,[40] willing some sufficient means, and setting all these means as ends themselves[41]—is an analytic truth about willing.[42] Indeed, it suffices for a definition which differentiates willing from, *e.g.*, *wishing*:[43] wishing is not bound by the rational requirements of PHI, while, by definition, willing is.[44] PHI thus provides a clear sense of *practical contradiction* as regards willing,[45] such that if one takes oneself to will some end but violates PHI, one *cannot* will the end one takes oneself to will and thereby contradicts oneself.

Therefore something is a *moral law* iff it applies to all morally responsible beings; further, a *Moral Law* is a moral law which grounds other moral laws but is itself not

[38] *Cf.* G.4:414–7, 441; CPR.5:20.

[39] *Cf.* AP 154n21, 219; CA 41, 43–4.

[40] *Cf.* G.4:417; CPR.5:26n.

[41] I thus combine several of what O'Neill takes to be distinct *Principles of Rational Intending* into the single PHI; *cf.* CoR 91–2.

[42] One might object to PHI because it seems to require a kind of rational omniscience in that, in setting an end, one already knows what is necessary and what is sufficient for the end. But this is a misunderstanding: if one is unaware of what means are necessary or sufficient for some end, all PHI requires is that one set oneself the end of *learning* what the end's means are, since this is itself a necessary means to setting the end.

[43] *Cf.* MM.6:213–4; A.7:251; O'Neill CoR 98; CA 43–4.

[44] Indeed, of all the practical attitudes one can take towards an end, *only* willing can be subject to rational requirements as Kant views rationality as nothing more or less than instrumental reasoning and, as such, nothing more or less than the capacity to will in general.

[45] *Cf.* Korsgaard 93–5.

grounded by anything else.[46] Kant is clear that, to play the role of a Moral Law, a law must command all morally responsible beings *categorically*, *not hypothetically*, *i.e.*, with respect to some possible circumstances or ends, because, if morality exists, it cannot be grounded by any particular set of circumstances or ends.[47] Thus, if a Moral Law exists, it can only apply to finite beings through a categorical imperative:[48] *i.e.*, a Moral Law can only be expressed through a categorical imperative, while all other moral laws can only be expressed through hypothetical imperatives since, as they are grounded by a Moral Law, they require some rule specifying antecedent conditions for their application.[49]

Further, there can be at most a *single* categorical imperative for any class of rational beings and thus at most a single *Moral Law*.[50] For if there were multiple purported categorical imperatives for any class of rational beings, either they could or could not in principle conflict: if the former, it would be possible to have conflicting obligations, so, *per* OIC, it cannot be that all the purported categorical imperatives are really categorical imperatives; if the latter, however, then there is no sense in which the "multiple" imperatives are genuinely distinct rather than simply equivalent formulations of the same categorical command, so one can meaningfully treat them as different *formulations*, but not different *imperatives*. Therefore, there can only be at most a single categorical imperative for any class of rational beings; *a fortiori*, there can only be at most a single categorical imperative for morally responsible, *i.e.*, practically free beings: call this *the Categorical Imperative* (CI). But since a Moral Law can only apply to morally responsible finite beings through a categorical imperative, there can be at most a single Moral Law, namely, that expressed by CI. Thus, when one can identify a categorical imperative that applies to practically free rational beings, one can immediately infer that it is *the* CI and that it expresses the Moral Law.

[46] Thus a Moral Law is to moral laws as theoretical principles are to theoretical judgments.

[47] *Cf.* G.4:416, 420, 441; CPR.5:32–3.

[48] Though some, *e.g.*, Foot, deny this, I set aside its justification in this work and simply accept it as a premise.

[49] *N.b.* As imperatives are distinct from the laws they expresses, a Moral Law is distinct from a categorical imperative.

[50] *Cf.* G.4:420–1.

Assuming there *is* a Moral Law, a finite rational being's maxim is *good* (or *moral*) iff it does not violate the Moral Law and *bad* (or *immoral* or *evil*) otherwise;[51] further, an action is *legal* iff there is a good maxim from which the action could have been performed and *illegal* otherwise; finally, a deed is good (or moral) iff it is done from a good maxim and bad (or immoral or evil) otherwise.[52] As such, a mere action is inherently *amoral*, *i.e.*, morality is not objectively valid for mere actions.

## 2. The Moral Law and the Singularity of Freedom

Kant distinguishes between three aspects of a finite,[53] morally responsible being, *viz.*, *animality*, *humanity*, and *personality*.[54] Animality, or *living nature*, is the set of natural drives, instincts, reflexes, processes, *etc.*, that constitute a being's capacity for mere action:[55] as such, animality is inherently amoral.[56]

Humanity, or *rational nature*, is the capacity to act from maxims and so the ability to *adopt* and *abandon* maxims: since one can only have an end insofar as it appears in one's maxim, this is simply the ability to *set* or *abandon* ends.[57] Humanity is a kind of *means-ends* or *instrumental rationality*: since, *per* PHI, setting an end already presupposes the capacity to understand what it is for something to be a means to the end, the ability to set ends already requires instrumental rationality. Further, since choice can only proceed in accordance with maxims and humanity is the ability to adopt or abandon maxims, a being's will, *viz.*, its power of choice, is therefore interchangeable with its humanity; this is unsurprising, given that humanity is simply rational nature and, as previously discussed, rational beings have wills simply in virtue of being rational. Finally, since adopting or abandoning some maxim is itself an action, insofar as a rational being adopts or abandons a maxim *in accordance with some other maxim*, the adoption or abandonment is a deed.

[51] Since infinite rational beings' wills are determined by commands, *all* maxims of infinite rational beings would be good.

[52] *Cf.* CPR.5:71, 151; R.6:14, 30, 46–7; MM.6:214, 219, 225.

[53] Unless specified, all beings I discuss hereafter are assumed to be finite.

[54] R.6:26.

[55] *Cf.* G.4:459; CPR.5:9n, 33; R.6:26; MM.6:211, 386–7; A.7:265, 267; LP.9:441, 449.

[56] *Cf.* R.6:57–8; LP.9:448.

[57] *Cf.* G.4:408, 428–9, 433, 436, 437–8; MM.6:381, 386–7, 391–2, 434–5; R.6:26, 26n, 60.

Since transcendental freedom is objectively valid for rational beings,[58] it is not a category mistake to consider humanity as being transcendentally free: thus *mere* humanity is the capacity to set ends only insofar as this capacity is determined by natural laws, while transcendentally free humanity is the capacity to set ends (possibly) independently from being so determined by natural laws.[59] While a merely rational being has some ultimate end set *for it* by the laws of nature and can only set ends for itself insofar as they are ultimately means to its ultimate end,[60] a transcendentally free rational being must be thought of as setting *all* its own ends: since a transcendentally free rational being can adopt or abandon an end of its own accord, even if some end were originally set for it through a natural law, such an end would only remain its end insofar as it chooses not to abandon it, *i.e.*, insofar as it approves of the end.[61] This is what Allison calls the *Incorporation Thesis*:[62] a transcendentally free rational being only has a maxim insofar as it freely adopts it or, what amounts to the same thing, insofar as it freely abstains from abandoning it—*i.e.*, transcendentally free humanity only has ends it sets *for itself*.[63]

An important lemma for the argument that follows is that transcendentally free humanity necessary has at least *some* end at all times. If, at some time, a rational being had *no* ends, no action could be imputed to it since all deeds require ends; such a being has no ability to adopt or abandon maxims except in accordance with laws of nature, meaning it lacks transcendentally free humanity.[64] I take it this is what Kant means when he repeatedly claims that happiness is a "necessary" end: since he simply defines happiness as having one's ends fulfilled, this amounts to the claim that one necessarily has ends and therefore necessarily aims to satisfy them.[65]

---

[58] §I.4.

[59] *Cf.* G.4:446.

[60] There cannot be multiple ultimate ends for the same reason there cannot be multiple categorical imperatives: indeed, such an ultimate end is *precisely* the categorical imperative that applies to the merely rational being.

[61] This shows why autonomy presupposes transcendental freedom: since merely rational beings can only have ultimate ends set *for* them and, further, such an ultimate end comes with a corresponding categorical imperative—*viz.*, to always acts so to advance the end—merely rational beings are necessarily *heteronomous*, as their categorical imperatives are not self-given but rather given by whatever natural processes set their ultimate ends.

[62] KTF 39–40.

[63] *Cf.* R.6:23–4; MM.6:239, 381.

[64] O'Neill recognizes that this is a key point, but simply assumes it as a premise—AP 182–3.

[65] *Cf.* TP.8:278–80, 284–5; G.4:393, 395, 396, 399, 415, 416–7, 418–9; CPR.5:22, 25, 73, 110–1, 124; MM.6:377–8; R.6:58; A.7:325.

Personality, or *moral nature*, is simply that in virtue of which a rational being is morally responsible; thus it is a being's practical freedom or, alternatively, autonomy. Personality is not included in humanity but instead builds upon it:[66] at the end of this section, I ultimately conclude that personality is nothing other than transcendentally free humanity and therefore that transcendental freedom and practical freedom, though conceptually distinct, are merely different ways of understanding the very same kind of freedom.

❧ ❧ ❧

An *end-in-itself* is an end which cannot be a mere means by its very nature: *i.e.*, treating an end-in-itself as a mere means yields a practical contradiction and so violates the requirements of rationality.[67] Korsgaard argues that humanity *confers value* on something by setting it as an end and, as such, it is a *source of value*:[68] I agree, but only for *transcendentally free* humanity, not, as Korsgaard suggests, humanity *tout court*. Indeed, I claim further that transcendentally free humanity is the *only* source of value of which finite beings can be aware;[69] while I do not have in mind any thick notion of value—for my purposes, something is *valuable* insofar as one rationally ought to set it as one's end in virtue of its having value[70]—this is nevertheless an important, non-trivial point.

It is easiest to explain why this must be so by responding in turn to five objections Langton makes to Korsgaard:[71]

1. *Objection*: "I could regard my actions as rational without regarding my ends as [valuable]".

[66]Rawls sometimes claims that humanity includes personality and sometimes that the latter is not reducible to the former: I reject the former claim and endorse the latter, as, in a quite remarkable passage, Kant explicitly states that humanity is, of itself, insufficient for personality—*cf.* *Lectures* 188, 292–3; R.6:26n.

[67]*Cf.* G.4:427–8; MM.6:434–5.

[68]*Cf.* ix–x, 16, 17, 112, 123, 128, 239–40, 241, 260–1, 284, 301, 345.

[69]*Cf.* MM.6:422–3.

[70]This plausibly covers all kinds of values—pragmatic, moral, and even aesthetic. *N.b.* Insofar as one aims to advance one's ends either by positively achieving them or merely negatively preserving them, this preservation need not be *temporal*: *e.g.*, one may seek to preserve the aesthetic value of *ephemerality* precisely by *not* temporally preserving the ephemeral object.

[71]169. To avoid potential confusion between moral and non-moral notions of goodness, I rephrase Langton's objections, which I take to be about non-moral goodness, to discuss *value*; since I have a relatively thin notion of value in mind, I take it that Langton's objections expressly concern value.

*Reply*: I concede this for rational beings *in general*, but not for transcendentally free rational beings: if one can set one's ends independently of the laws of nature, one cannot take oneself to be acting rationally if one does not regard the ends upon which one actually acts as valuable at the time at which one acts. It seems one *must* take something to be valuable if one is to have transcendentally free humanity at all: if one were a universal nihilist who denies that anything has value, one would lack reasons to do anything, meaning one cannot act on a maxim and, *a fortiori*, cannot set anything as an end. Thus a transcendentally free rational being must regard its ends as valuable insofar as its deeds are rational.[72]

2. *Objection*: "I could regard my ends as [valuable] without according to myself a power of conferring [value] on the objects of my choice".

   *Reply*: From the previous reply, if one has transcendentally free humanity, one must regard *all* one's ends as being valuable, but, further, one must also regard *only* one's ends as valuable: as value is simply that in virtue of which one rationally ought to set something as one's end, one cannot, on pain of irrationality, regard something as valuable without also setting it as one's end.[73] Therefore, one regards something as valuable iff it is one's end: but how is this possible? Either (i) one *confers* value upon something by setting it as one's end or (ii) some things are valuable *independently* of one's setting them as ends and, through an "axiological sense", one is aware of this independent value and sets one's ends accordingly. Note that these options, though not mutually exclusive, are exhaustive: if one takes something to have value, either one is unaware of any independent value and, as such, one only takes it to be valuable

[72]For merely rational beings, this does not hold as the question of value need not arise: a merely rational being simply reasons its way to its ultimate end.

[73]Of course, this is not, *e.g.*, a *physical* impossibility, but only a *rational* one—however, since Langton's objections concern an agent which considers itself to be acting rationally, my response is sufficient.

because one confers value upon it, or one is aware of its independent value, though one can perhaps confer additional or distinct value upon it by setting it as one's end. However, Kant must reject (ii): even assuming the concept of independent value is objectively valid for phenomena,[74] if reasonable disagreements about whether something is intrinsically valuable are to be possible, any independent value cannot be reducible to a thing's non-relational empirical properties.[75] Therefore, the concept of independent value cannot have any schema by which it can be applied to phenomena and so, though it may very well be that certain things really have independent value, no finite being, as a discursive intellect, can ever know of any phenomenon whether it has independent value. Since transcendentally free rational beings necessary value something iff they set it as an end but cannot be aware of any possible independent value, it must be that transcendentally free humanity confers value upon its ends.

3. *Objection*: "I could accord to myself a power of conferring [value] on the objects of my choice without according the same power to others".

   *Reply*: Since the reply to the previous objection makes no reference to any particular being but, instead, only to its transcendentally free humanity, it must be that *all* transcendentally free rational beings confer value on their ends. It may be that only *some* rational beings are transcendentally free and, as such, that only those beings confer value on their ends, but insofar as one takes a being to be transcendentally free, one must take them to confer value on their ends. The only way to deny a rational being the power to confer value on its ends is to deny that it is transcendentally free.

[74] This is already unclear: what would it be, *e.g.*, for something to have independent extrinsic value?

[75] If this disagreement is to be *genuine*, independent value cannot be reducible to relational empirical properties either.

4. *Objection*: "I could accord to others the power of conferring [value] on the objects of their choice without according intrinsic value to them".

   *Reply*: I take it that this objection claims that one can take another being to confer value on its ends without oneself taking those ends to be valuable. From the previous reply, one accords a being the power to confer value on its ends iff it is transcendentally free: however, this is not the power to confer value *only for the being setting the end*, but the power to confer *value*, for if it were only the former, this would amount to the claim that, objectively speaking, there is *no* value whatsoever of which finite beings can be aware. One must accord all transcendentally free rational beings the power to confer genuine value and, as such, each transcendentally free rational being is required to set all ends set by every other transcendentally free rational being.[76]

5. *Objection*: "I could accord intrinsic value to them without their having, or acquiring, intrinsic value".

   *Reply*: I take it this is the skeptical claim that one might take oneself or others to have the power to confer value and yet be mistaken, and nothing I will say in this section can overcome this doubt; however, assuming that there *is* a transcendentally free rational being, one cannot doubt that *it* has the power to confer value on its ends—the only thing that can be doubted here is whether there is transcendentally free humanity at all.

Thus transcendentally free humanity confers value on its ends and, so far as finite beings can be aware, is the only possible source of value.

But if some arbitrary end is only valuable insofar as it is set by transcendentally free humanity, it must be that transcendentally free humanity has unconditional value and, as such, is an end-in-itself. From the lemma proved above, transcendentally free humanity necessarily has at least some end, but, as this end's value can only

[76]This is why Kant claims that the happiness of others is an obligatory end—*cf.* MM.6:385, 388, 393–4.

be, so far as a finite being can tell, conditioned upon being set as an end by transcendentally free humanity, transcendentally free humanity is therefore a necessary means to enjoying the value of the end. Thus, *per* PHI, a transcendentally free rational being setting an end is rationally required to set transcendentally free humanity as an end as well, meaning that transcendentally free humanity is always rationally required to set itself as an end. Since transcendentally free humanity confers value on its ends, it therefore is rationally required to confer value *on itself* so long as it exists, but, since existence is, as Kant famously claims, not a real predicate,[77] transcendentally free humanity therefore has unconditioned value. And since transcendentally free humanity, by its very nature, cannot be rationally set as a mere means, it is therefore an end-in-itself.[78]

The *Formula of Humanity* (FH) is: *always treat transcendentally free humanity as an end-in-itself and never as a mere means*.[79] Since FH expresses rational requirements for transcendentally free rational beings, it is an imperative expressing a law; further, since these requirements are not restricted by reference to any possible circumstances or ends, it is a categorical imperative, namely, that for transcendentally free rational beings. Since, as noted above, the *practical* freedom necessary and sufficient for moral responsibility presupposes transcendental freedom and CI is the categor-

[77] *Cf.* A598/B626.

[78] This argument rests on the lemma that transcendentally free humanity necessarily has at least *some* end at all times and, further, since transcendentally free humanity necessarily values all and only its ends, that transcendentally free humanity necessarily *values something*. Langton flatly rejects this last claim, even going so far as to say that doing so "needs little defense" (181n35), but this is mistaken. Though it is certainly plausible that one fail to find anything *pleasurable*, fail to find value in what one *used* to value, or even, in the depths of one's despair, *think* that one fails to value anything at all, this is not at all equivalent to one *actually failing* to value anything: if a transcendentally free rational being genuinely did not take anything to have value, it could have no reason to do anything whatsoever, and so could perform *no deeds*—though such a miserable wreck may be possible, it *cannot* be transcendentally free as, by definition, a transcendentally free rational being is capable of setting ends, and thus performing deeds, independently of being determined by natural laws, something of which such a being is manifestly incapable.

[79] *Cf.* G.4:427–9, 431, 436, 437–8; CPR.5:87, 110; MM.6:386–7, 391–2, 418, 422–3, 430, 434–5; LP.9:443, 44. *N.b.* In the canonical statement at G.4:429, Kant qualifies humanity with the phrase *whether in your own person or in the person of any other*, though, to my knowledge, few if any commentators take this as an important part of the formula. As stated above, I aim to show that personality is equivalent to transcendentally free humanity and so, on my account of personality, FH states exactly what I claim, rather than the standard view that it concerns humanity *tout court*. Though it is plausible that Kant had yet to develop his full account of personality by the publication of G, this is irrelevant for my argument: the view that humanity *tout court* is an end-in-itself is, as Langton rightly claims, incorrect, but, given Kant's (possibly later) account of personality, this is not what FH claims (or ought to have claimed) in the first place.

ical imperative for practically free rational beings, CI cannot conflict with FH: FH is therefore *entailed* by CI, meaning that, should there be a Moral Law, FH expresses at least *some* of its requirements for morally responsible finite beings. As such, personality presupposes transcendentally free humanity and so a *person* is, at the very least, a transcendentally free rational being:[80] personality, though *distinct* from humanity, is nevertheless "inseparable" from it.[81] While this may have been implicit from the fact that autonomy presupposes transcendentally free humanity, it is important to emphasize: it is not merely the case that transcendental freedom is a *prerequisite* of practical freedom, but rather that practical freedom *contains*, in a sense, transcendental freedom and therefore FH functions as a positive characterization of CI. This is *close* to establishing the second claim of Transcendental Compatibilism, but it is not yet enough: what remains to be seen is that transcendental freedom is equivalent to practical freedom, and, as such, CI is nothing other than FH.

First, however, there is a possible objection: Kant claims the *good will*, understood as a will that only adopts moral maxims, *i.e.*, wills in accordance with the Moral Law,[82] is the only thing that is good without limitation and, further, that it is an end-in-itself:[83] does this contradict my claim that transcendentally free humanity is an end-in-itself? O'Neill suggests that humanity is an end-in-itself *because* it is a prerequisite for a good will,[84] but this cannot work: first, this would apply to mere humanity as well as transcendentally free humanity; second, it requires, but fails to provide, any reason why a good will is an end-in-itself; and, third, it would make humanity only a *necessary* end, not an end-in-itself—*i.e.*, even though humanity would necessarily be set as an end, it would be rationally permissible to take it as a mere means to a good will. Korsgaard suggests that humanity simply *is* a good will, or at least is when in its "perfected state";[85] this seems to be little more than a rephrasing of O'Neill's position and thus faces the same issues.

---

[80] *Cf.* G.4:428; CPR.5:78, 87, 110; MM.6:223–4, 418, 422–3; R.6:26n, 27; LP.9:455.

[81] CF.7:59. This mirrors Rawls's account of the relation between rationality and reasonability, where 'rationality' is (roughly) Rawls's term for humanity and 'reasonability' that for personality, such that, though the reasonable presupposes and builds upon the rational, there is no "logical route…from the rational to the reasonable"—*Lectures* 293, *cf.* 164–5, 184–5, 223; *Political Liberalism* 48–9n1.

[82] *Cf.* G.4:414, 437–9.

[83] G.4:393–4.

[84] CoR 137–8.

[85] 16, 114.

I think that O'Neill and Korsgaard address this problem in the wrong direction: transcendentally free humanity is not an end-in-itself because of the good will, but the good will is an end-in-itself because of transcendentally free humanity. As noted above, a rational being's will is interchangeable with its humanity: therefore, since, at the very least, the Moral Law requires one to abide by FH, a good will necessarily treats itself as an end-in-itself or, alternatively, if a rational being has a good will, it has transcendentally free humanity that always sets itself as an end. Thus there is no tension in saying that both transcendentally free humanity and a good will are ends-in-themselves as the truth of the former entails the truth of the latter. As regards the *goodness* of a good will, this can be understood as either moral or non-moral goodness, *viz*., value, but, either way, there is no tension: since, by definition, a good will is the only thing that completely abides by the Moral Law, it must be morally good without limitation; and since a good will necessarily is such that it confers value on itself without any limiting condition, it therefore has unlimited value as well.[86]

This explanation makes better sense than O'Neill's and Korsgaard's of Kant's claim that *all* persons, even evil ones, have *dignity*, *viz*., unlimited value:[87] it makes little sense for a person to *actually have* dignity if that which grounds dignity is a (perhaps unactualized) *possibility*—why wouldn't those who are actually good have more value than those who merely *can* be good? All persons have equal dignity because all persons have transcendentally free humanity and, as such, are ends-in-themselves; nevertheless, there remains a clear sense in which good persons are better than evil ones—*i.e*., *morally*.

❧ ❧ ❧

Just as a rational being is autonomous iff laws only apply to it that it gives itself—*i.e*., iff it cannot be subject to any requirements whose origin are alien to itself—and heteronomous otherwise, one can think, by extension, of practical principles themselves as being autonomous or heteronomous, if only loosely speaking: a law or maxim is autonomous iff it can apply to autonomous beings, *i.e*., can be given by the being to itself, and heteronomous otherwise. The *Formula of Autonomy* (FA) simply states that one ought to act only from principles consistent with one's autonomy, *i.e*.,

[86]This is not to say that this value must be *infinite*, rather that nothing can possibly be valued more.
[87]*Cf*. G.4:434-6; MM.6:434-5; A.7:295.

autonomous laws and maxims:[88] as is quite clear, FA is therefore a categorical imperative that applies to all autonomous, *i.e.*, morally responsible, beings and therefore is a formulation of CI expressing the Moral Law.

One can see that FH is an autonomous law: since transcendentally free humanity, by its very nature, is necessitated by FH, there is no sense in which FH has a source alien to transcendentally free humanity. Further, it must also be that FH requires transcendentally free humanity to reject any putative heteronomous laws. A putative heteronomous law either can or cannot in principle conflict with FH: if the former, *per* OIC, transcendentally free humanity cannot be necessitated by it and so FH requires transcendentally free humanity to reject it; if the latter, the law is either entailed by FH or there is no sense in which it is really distinct from FH so, either way, the law is really autonomous. FH is therefore equivalent to FA, meaning *both* are equivalent formulations of CI and, as such, both express the Moral Law:[89] a tran-

[88] *Cf.* G.4:431.

[89] This is one part of what can be called Kant's *Equivalence Thesis*, the claim that all the categorical imperatives he discusses in G are really equivalent formulations of CI—*cf.* G.4:437–8. Though commentators squabble about how many distinct formulations Kant actually presents, I take it there are five: in addition to FH and FA, these are the *Formula of Universal Law* (FUL), the *Formula of Laws of Nature* (FLN), and the *Formula of the Kingdom of Ends* (FKE)—*cf.* G.4:402, 421, 439. Though the equivalence of FH and FA with the other three formulations is not necessary for my main argument, it is of overarching importance if my general account of Kant's moral theory is to succeed—indeed, FUL and FLN are often considered to be the paradigmatic formulations of CI—and so I present a few quick argument sketches here which, when properly fleshed out, would fully establish the Equivalence Thesis.

From FH/FA, something is a universal moral law iff it is consistent with treating transcendentally free humanity as an end-in-itself and, as such, one ought only to act on maxims which correspond to such moral laws, *i.e.*, maxims which themselves are consistent with treating transcendentally free humanity as an end in itself; however, since FUL simply states that one ought only to act on maxims that one can consistently will to be universal (moral) laws, FH/FA permit exactly the same maxims as does FUL, meaning FUL is equivalent to FH and FA. (The common distinction between maxims that are ruled impermissible by FUL by contradictions in *conception* as opposed to contradictions in the *will* seems quite plausibly to be the precise analogue of treating transcendentally free humanity as a mere means by either failing to *positively* advance it or refraining from *negatively* preserving it.)

For Kant, a maxim's *type* is a hypothetical law of nature in accordance with which *all* persons would universally adopt the maxim—*cf.* G.4:402, 421; CPR.5:69. FLN states that one ought only to act on maxims whose *types* can be consistently willed: however, since willing the type of a maxim ruled impermissible by FUL/FH/FA violates the necessary means for treating transcendentally free humanity as an end-in-itself, by PHI, one cannot consistently will the type of such a maxim. As such, FLN also permits exactly the same maxims as FUL, FH, and FA, and so all four are equivalent.

FKE states that one ought only to act in accordance with the maxims legislated by a sovereign member of a possible kingdom of ends, where a *kingdom of ends* is simply a systematic union of a totality of some set of persons' ends, a *member* of a kingdom of ends is a person who both gives and is subject to the universal laws applying to all persons whose ends comprise the kingdom of ends,

scendentally free rational being is such that laws can only necessitate its will insofar as it gives them to itself—*i.e.*, a transcendentally free rational being is necessarily *autonomous*—and so transcendental freedom is not merely *necessary* but also, perhaps unexpectedly, *sufficient* for practical freedom. As practical freedom presupposes transcendental freedom and transcendental freedom entails practical freedom, they are really just two ways of understanding a single property of rational beings, *viz.*, *freedom simpliciter*:[90] this, then, is the claim of the *Singularity of Freedom*. Further, since moral responsibility or personality is equivalent to practically free humanity, it follows that personality does not simply *presuppose* free humanity, as was shown above—*personality is nothing other than free humanity*; similarly, CI does not simply presuppose FH—CI *is nothing other than* FH. As such, I have now established the second claim of Transcendental Compatibilism: the very same freedom compatible with LNC is both necessary and sufficient for moral responsibility.[91]

## 3. The Fact of Reason

Though my discussion of the *Fact of Reason* (FR) will be brief, FR is of such importance as to merit its own section. First, however, one must understand exactly what Kant takes FR to be, as he himself is not altogether clear. Crucially, FR is a *factum* as opposed to a *datum*, *i.e.*, something one *does*, not something one is *given*.[92] Thus one is only conscious of FR *through* one's undergoing moral reasoning and judgment.[93]

One must take FR to be *one's awareness from one's moral reasoning and judgment of one's own moral responsibility*:[94] *i.e.*, since one undergoes moral deliberation, one has a practical warrant to take oneself to be morally responsible. Since, from the Singularity of Freedom, one can only be morally responsible, *i.e.*, a person, if one

and a *sovereign* of a kingdom of ends is a member whose lawgiving is independent of that of all other members—*cf.* G.4:433. As such, a sovereign of a possible kingdom of ends is only subject to laws it gives itself, meaning that a maxim is ruled permissible by FKE iff it is ruled permissible by FA, and so FKE is equivalent to FUL, FLN, FH, and FA; therefore, all five formulations of CI Kant gives in G are equivalent.

[90] As such, I feel free to drop the qualifiers 'transcendental' and 'practical' and simply discuss them both as 'freedom'.

[91] In terms of the *Wille/Willkür* distinction, freedom of *Willkür* is transcendentally free humanity while freedom of Wille is personality: as such, while the freedom of *Wille* is conceptually distinct from freedom of *Willkür*, the one nevertheless reciprocally implies the other.

[92] *Cf.* O'Neill, CoR 64–5; Łuków 211.

[93] *Cf.* Rawls, *Lectures* 260.

[94] *Cf.* CPR.5:31–2.

is free, this is therefore a warrant for taking oneself to be *free*. As there can be no theoretical grounds for or against the claim that some rational being is free,[95] this practical warrant from FR is therefore *indefeasible*, meaning one can be *certain* one is free, precisely because one is certain one is morally responsible.[96] Given that I know I am free and, further, that freedom is necessary and sufficient for moral responsibility, I thereby know that *the Moral Law is objectively real*.

As Łuków notes, FR cannot *of itself* prove anything about either freedom or the Moral Law and can instead only "testif[y]" to their "legitimacy";[97] however, it *can* prove their objective reality alongside the claim that nothing else can provide any grounds for accepting or rejecting the reality of freedom or the Moral Law. This therefore explains Kant's odd locutions about freedom and the Moral Law being real *from a practical point of view*:[98] all this means is that one's certainty of their objective reality derives from practical and not from theoretical sources, not that it is any less certain, and it is in this sense that freedom is a *postulate of practical reason*.[99]

As such, FR—or, more precisely, *consciousness* of FR—provides freedom with an empirical schema: by recognizing that some rational being is conscious of FR, one can judge it to be free and thus a person.[100] So while the resolution to the Third Antinomy establishes the objective *validity* of freedom for rational beings, FR establishes its objective *reality* for rational beings.[101] This therefore establishes the third and final claim of Transcendental Compatibilism: since, under normal circumstances, human beings *are* conscious of FR, it follows that, despite being completely determined by natural laws, they are genuinely free and therefore morally responsible.

## 4. Three Objections to Transcendental Compatibilism

It seems that Transcendental Compatibilism describes something like this: through a *noumenal choice*, one freely chooses one's intelligible character, which, in turn,

[95] §I.4.

[96] *Cf.* CPR.29–30, 31, 42, 47, 121; R.6:49–50n, 50n; OT.8:141n, 141–2.

[97] 215.

[98] *Cf.* CPR.5:49–50, 55, 56, 121, 133–5; MM.6:438n.

[99] *Cf.* CPR.5:132. Though, to be sure, this is not to say that Kant is correct in claiming that either immortality or God are also postulates.

[100] Of course, this presents the problem of *how* one can recognize whether another being is conscious of FR, but I view this as tractable, for it is obvious in everyday life that one *does* recognize when others engage in moral deliberation and judgment.

[101] *Cf.* CPR.5:47, 49.

*grounds* one's empirical character, which, finally, *determines* one's deeds; since the noumenal choice of intelligible character is free and it ultimately determines one's deeds, one can be thought to freely choose one's deeds and thus be morally responsible for them. Not only is this a decidedly odd view, it faces three potent objections: the *Split-Personality Objection*, the *Misplaced Freedom Objection*, and the *Explosive Responsibility Objection*.

The Split-Personality Objection claims that, if the above picture is right, one's empirical character must, in an important respect, be distinct from and therefore *split* from one's intelligible character: but, as one's empirical character is a quality of one's phenomenal self and one's intelligible character is a quality of one's noumenal self, this entails that one's phenomenal and noumenal selves are split. Indeed, Kant himself notes that a "human being…must think of a *dual personality* in himself, a *doubled self*…".[102] Though such a split entails no contradiction, this is an unpalatable view far removed from everyday experience.

The Misplaced Freedom Objection claims that, since one's noumenal choice sufficiently determines all one's empirical choices, strictly speaking, the only free choice one can make is noumenal. But if this is true, there is no sense in which one's phenomenal self is genuinely free: since there are no genuine choices possible in the phenomenal world, *no* phenomena can be free, meaning all freedom is strictly speaking *noumenal*.[103] As such, though human beings can be morally responsible for their deeds, it is only noumenal human beings that actually are morally responsible, thus precluding any account of OGOP: ordinary, *i.e.*, phenomenal, people cannot be morally responsible, even if their noumenal counterparts can be, so they cannot strictly speaking be said to be good.

Finally, the Explosive Responsibility Objection maintains that the above picture generates absurd results for moral responsibility. Assume, as the picture suggests, that one is morally responsible for one's deeds only because one's free choice grounds one's empirical character which ultimately determines that one performs just those deeds. Since, as something that arises in the phenomenal world, one's empirical character must form in accordance with LNC and, *per* PHI, one can only choose some end insofar as one also chooses all its necessary means, it follows that, in choosing one's empirical character, one must also choose the course of history that ultimately leads

[102] MM.6:438n, emphasis mine.
[103] *Cf.* Allison, KTF 139.

to one's empirical character forming in the way it does; thus, if one freely chooses and is thereby responsible for one's empirical character, one must also freely choose and thus be responsible for the *entire course of history* leading up to one's lifetime and, since everything in the phenomenal world necessarily happens in accordance with causal laws, *everything that will ever happen*. Thus if one is ever morally responsible for *anything*, one is morally responsible for *everything*.[104] Since this is a ridiculous moral view, one therefore ought to reject Transcendental Compatibilism and, with it, any hope of a Kantian account of OGOP.

❧ ❧ ❧

These are certainly potent objections, but I believe each can be overcome. The first step is to clarify exactly what the empirical and intelligible characters must be for Transcendental Compatibilism.

Kant takes one's intelligible character to be the *unity* of one's empirical character throughout one's *moral lifetime*,[105] where one's moral lifetime is understood as the union of all the intervals of one's lifetime in which one can be conscious of FR: as such, it excludes, at the very least, one's childhood and any period of senility, as well as periods such as when one is asleep. I take it that one's empirical character, *at some time*, is nothing other than the totality of one's maxims as that time.[106] One's empirical character is therefore *dynamic* and not *static*: however, since the adoption or abandonment of a maxim, insofar as it is a deed, can only be done in accordance with a maxim, it follows that one's empirical character at any time contains the laws dictating its own alteration. Further, since this discussion only concerns people insofar as they are morally responsible, one's only concern is one's empirical character insofar as one is a person, *i.e.*, insofar as one can freely abandon any maxim already in one's empirical character and freely adopt any that is not.[107]

[104] *Cf.* Walker 148–9; Allison, KTF 138–9.

[105] *Cf.* CPR.5:99, 123–4; R.6:47–8, 70n, 74–5n.

[106] *Cf.* PM.20:336; Allison may have this or a similar view in mind as well, though it is not entirely clear—*cf.* "Antinomy" 26.

[107] *N.b.* This is a different sense of 'character' from that by which one might call something 'characteristic' or 'uncharacteristic' of someone: a person can do something that is "out of character", but, insofar as it is a deed and thus imputable to the person, it is necessarily done from a maxim that is part of their empirical character. However, empirical character is still very closely related to this ordinary conception of character in that one's judgments of characteristic-ness are based on what one *thinks* someone's empirical character is.

Several points support this understanding of empirical character. First, this fits the requirement of empirical character that it determine everything one does, at least insofar as one is morally responsible, as everything one does for which one is morally responsible is necessarily done in accordance with a maxim; further, since, on this view, an empirical character contains the rules through which it evolves over time, it is consistent with Kant's claim that, if one could know a person's empirical character, one could perfectly predict everything they would ever do.[108] Second, the idea that one's empirical character is dynamic is attractive, both because it is plausibly closer to pre-theoretical notions of character and because Kant seems to presuppose this in his doctrine of the *change of heart*, as will be discussed in Chapter III. Third, as one's intelligible character is to be understood as the unity of one's empirical character and, as will also be discussed in Chapter III, Kant takes one's intelligible character to contain a kind of maxim, it makes sense for one's empirical character to be composed of maxims: otherwise, it would be puzzling how the unity of something not composed of maxims is itself maxim-like.

With this understanding of empirical character and its relation to intelligible character, one can provide satisfying responses to the first two objections above. In response to the Split-Personality Objection, the distinction between one's empirical and intelligible characters is the same transcendental distinction as between phenomena and the noumena that ground them: one only has a *single* character which can be understood in *two* ways. The Split-Personality Objection can therefore be rejected outright for taking a Transcendental Realist interpretation of Transcendental Compatibilism.[109]

One can give a similarly quick response to the Misplaced Freedom Objection: while it is true that, *noumenally*, one only has a single free choice—*viz.*, one's choice of intelligible character—this is simply considering the plurality of free choices one makes throughout one's moral lifetime as a unity, choices which, *phenomenally*, genuinely form a plurality. Under normal circumstances—*i.e.*, when one is conscious of FR—one is *genuinely free* in one's choices in everyday life, even if one's everyday choices are also completely determined by causal laws; since one's noumenal self is numerically identical with one's phenomenal self and, as the resolution to the Third

[108] *Cf.* CPR.5:99.

[109] *Cf.* PM.20:270. This is precisely what Kant goes on to say in the passage cited above from MM.6:438n.

Antinomy demonstrates, nothing about the phenomenal world is inherently inconsistent with freedom, one cannot consistently maintain that one's noumenal self is free but that one's phenomenal self is not.

The third objection, however, requires more than just a fuller account of empirical character—indeed, so far as I can see, it can only be addressed by bringing in a principle I have proven elsewhere:[110] assuming the actual world is completely deterministic and is such that some event $x$ does not happen at $t$, for any arbitrary relation of similarity (other than complete identity) and for any arbitrarily long period from $t_0$ to $t - \varepsilon$ where $\varepsilon$ is some (possibly null) "ramp-up" period, there is a possible world $w$ with the exact same deterministic laws as the actual world which is relevantly similar to the actual world from $t_0$ to $t - \varepsilon$ such that $x$ happens at $t$ in $w$. In slightly plainer terms, this is the claim that, so long as $w$ is not identical to the actual world before $t$, there is a $w$ which shares the exact same deterministic laws as the actual world and which *approximates* the actual world for any arbitrarily long finite period until shortly before $t$, when $w$ stops approximating the actual world and something that does not happen in the actual world at $t$ happens in $w$.

For the purposes of responding to the Explosive Responsibility Objection, let the relevant similarity relation be the *phenomenological indiscernibility of experience for all rational beings*: as such, a possible world is similar to the actual world in the relevant respect at some time iff every rational being in the actual world not only exists in the possible world but is such that they cannot distinguish the experience they have in the actual world from that which they have in the possible world. Further, let the event that occurs (or does not occur) in the actual world be one's adoption (or abandonment) of some particular maxim at $t$; since the "ramp-up" period $\varepsilon$ is simply the period of time it takes for a world similar to the actual world at $t - \varepsilon$ to change in accordance with natural laws so that one adopts (or abandons) the maxim at $t$ and, quite plausibly, the minimum time it takes for one to adopt or abandon a maxim, even if one is not already considering it, can be fairly close to zero, the required period of similarity can stretch from as far in the past as one would like—the formation of the Solar System, say—until just before $t$.

One can call a possible course of history in which every rational beings' experience is phenomenologically indiscernible from that which they have in the actual

[110] Shapiro 11-3; this reference was omitted in the original submitted version of this thesis for marking anonymity reasons.

world *approximately identical* to the actual course of history; by extension, one can call the set of all approximately identical histories an *approximate course of history*. One plausibly only ever cares about an approximate course of history, since, by construction, no rational being can distinguish one course of history from another approximately identical to it. It therefore follows from the above principle that, for *any* time in my moral lifetime when I adopt (abandon) a maxim, it is possible, given the approximate course of history until shortly before that time and holding the laws of nature completely fixed, that I refrain from doing so.

Further, while the possibility of performing *external* actions can be constrained by one's circumstance—*e.g.*, I cannot bite into a pear when I do not have one—adopting or abandoning maxims, as merely *internal* acts, are plausibly under no such constraint: thus it seems plausible that, even if one is deliberating whether to adopt (abandon) a maxim concerning some particular set of circumstances, one instead adopt (abandon) a maxim concerning a completely different set of circumstances. Indeed, this seems a familiar event in everyday life: *e.g.*, I may be thinking about what I intend to eat for lunch when I suddenly remember that my brother's birthday is next week and decide that I will try to organize a birthday dinner in the next few days. Further, it seems that one can also adopt (abandon) maxims even when, just a moment before, one is not consciously deliberating anything at all, as would be the case in the previous example if, rather than choosing lunch, I was instead reading a book. As such, one can quite plausibly extend the previous conclusion: not only is it possible that one *refrain* from adopting (abandoning) some maxim one actually adopts (abandons), it is also possible that one instead adopt (abandon) some *other* maxim; *i.e.*, at any time in one's moral lifetime, it is possible, holding both the laws of nature and the approximate course of history until shortly before that time fixed, that one (i) could have refrained from adopting (abandoning) a maxim one actually adopts (abandons), (ii) could have adopted (abandoned) a maxim one actually refrained from adopting (abandoning), (iii) could have adopted (abandoned) a *different* maxim from that which one actually adopts (abandons), or (iv) could have adopted (abandoned) a maxim when one actually abandoned (adopted) some other maxim. As such, it follows that, *at any point in one's moral lifetime*, one's empirical character is *independent*

of the approximate course of history until shortly before that point, even while being completely determined by fixed causal laws.[111]

Indeed, even if the course of history were *not* approximately identical to the actual course of history, one could still have the same empirical character at all times as one would have had in the actual world:[112] *i.e.*, since, regardless of what the approximate course of history is, one's empirical character is independent from it, one can keep "correcting", so to speak, one's empirical character to what it actually would have been, even if the course of history were significantly different from that of the actual world. It follows that one's empirical character is *completely independent* of the course of history, not merely the *approximate* course of history.[113]

Since the principle requires that the laws of nature be held completely fixed across similar worlds, this does not, as one might suspect, rest on any presumed inability to know all the laws of nature in complete detail and with certainty, but rather on the Kantian premise that it is *impossible* to know the "state of the phenomenal world" in complete detail and with certainty. Indeed, even the concept of the "state of the phenomenal world" involves a transcendental error, as Kant's resolution to the First Antinomy entails that the phenomenal world can be neither finite nor infinite and, as such, *is not a totality*,[114] meaning that one cannot coherently refer to the set of all phenomena, as one must if one is to flesh out the concept of the "state of the phenomenal world":[115] no matter how many phenomena's states one takes

[111]There is a potential objection that maxims are constrained by the concepts one possesses: if maxims are in some way conceptual, one cannot adopt a maxim that involves concepts one does not possess. Even granting this objection, this would only constrain the maxims one can *adopt*, *not* those one can abandon—and since, regardless of which concepts one actually possesses, there would still be a huge range of maxims that one can adopt, this "constraint" is not very constraining. The objection makes a valid point, but does nothing to undermine the overall claim that one's empirical character is independent of the approximate course of history.

[112]This assumes both that it makes sense to talk of something as happening at "the same time" in such a world and that there is a sense in which one is the same person in the two courses of history: both of these are straightforward while holding the approximate course of history fixed, but they are more contentious here.

[113]This assumes that, as noted above in n111, one possesses the same concepts in both worlds; while this possibly rules out the very strong claim that *no matter* the course of history, one's empirical character could nevertheless be the same as it actually would have been, the result remains quite strong—so long as one can possess the same concepts, one's empirical character is *completely* independent of the course of history.

[114]*Cf.* A426–33/B454–61, A504–5/B532–3, A517–23/B545–51.

[115]One must distinguish between talking about all phenomena and the *set*, *i.e.*, *totality*, thereof, such that while talking about the set of all phenomena is a transcendental error, it is permissible to discuss all phenomena, as long as this is understood as shorthand: *e.g.*, "all phenomena are such-and-such" is

oneself to know in complete detail and with certainty, there would always be more phenomena which, by LNC, are causally related to those phenomena about whose states one knows, and so, even if one did somehow know all natural laws in complete detail and with certainty, one still could not calculate, like a Laplacian demon, future states even of the phenomena whose current states one knows, as they may well be directly causally influenced by some phenomena about whose state one is currently ignorant.[116] Further, this problem cannot be overcome by appealing to a "God's-eye perspective" from which an *infinite* being—Laplace's demon itself—can know the entire state of the phenomenal world and thereby calculate all its future states, for such a perspective can only be of the *noumenal* world, not the phenomenal one, and since laws of nature only apply to phenomena, such a calculation of the entire universe is impossible for both finite and infinite beings.

With this result, one can finally respond to the Explosive Responsibility Objection: even though one's empirical character must arise in accordance with natural laws, it is nevertheless independent of the course of history such that one can be responsible for one's empirical character—and, as such, for one's deeds—but *not* for the entire course of history. Indeed, on this view of Transcendental Compatibilism, even assuming the laws of nature are somehow preselected, the choice of any one person's intelligible character would be insufficient to determine the course of history—for that, independent choices of *all* persons would be required at the very least—so it could not possibly be that any individual person is morally responsible for the entire course of history; and since the choices of different persons are *independent*, it cannot be that any group of persons is collectively responsible.

As such, Transcendental Compatibilism, with a little assistance from my principle, has been shown to have the resources to claim that human beings, under normal circumstances, are genuinely free in their everyday lives—as such, ordinary people are *persons*, *i.e.*, morally responsible. Therefore no reason remains to think that Kant cannot have an account of OGOP: since ordinary people *are* morally responsible, his moral theory ought to be able to accommodate a sense of their moral worth.

shorthand for something like, "since so-and-so is a constitutive principle, it follows that some object can only be a phenomenon insofar as it has such-and-such property".

[116] Additionally, one cannot know even one's empirical character in complete detail with certainty, as will be discussed in the following Chapter.

— III —

# The Goodness of Persons

After first establishing, in §1, Kant's account of *noumenal moral worth*, *i.e.*, goodness for noumenal persons, I then derive, in §2 and §3, his account of *phenomenal moral worth*, *i.e.*, goodness for phenomenal persons. Though this account establishes that goodness is both objectively valid and objectively real for ordinary human beings, it is insufficient for OGOP because it apparently lacks a useful schema through which one can straightforwardly make judgments concerning the goodness of other persons: once the account is further characterized in §4 to show it does have such a schema, I conclude that this Kantian account of goodness can play the role of OGOP.

## 1. Noumenal Goodness

A maxim $M = \langle C, D, E \rangle$ can be said to *contain* another maxim $M' = \langle C', D', E' \rangle \neq M$ iff $C$ entails $C'$, $D$ entails $D'$, $E$ entails $E'$, and $M$ and $M'$ are both in a rational being's set of maxims (at some time);[1] since $M \neq M'$, it follows that at least one of $C'$, $D'$, and $E'$ fails to entail $C$, $D$, or $E$, respectively.[2] As such, one can think of $M'$ adding some additional requirement(s) beyond $M$'s circumstances, deed, or end.[3] A *supreme maxim* is a maxim which contains all other (non-tautological) maxims

[1] That is, either $M$ and $M'$ are both in a rational being's empirical character at some time or they are both in a rational being's intelligible character; as such, *time* is not crucial for containment, only that maxims be in the same set.

[2] *N.b.* Since both are in the same set of maxims, they are necessarily consistent.

[3] This definition ignores differences in the quality of maxims and, as such, one may instead want to say that, rather than defining a relation between *maxims*, it should rather be a relation between the conjunctions of all maxims in the set sharing the same circumstance, deed, and end; however, it is sufficient for my purposes here that I speak of containment holding between maxims *per se* rather than conjunctions of maxims.

in its set and *cannot* be contained by any (non-tautological) maxim in the set. By definition, a supreme maxim must have its circumstance, deed, and end left open, *i.e.*, a supreme maxim necessarily concerns *all* circumstances and *any* deed for *any* end. There are four putative supreme maxims: the *positive* supreme maxim (*I may always do anything for any end*), the *negative* supreme maxim (*it is not the case that I may always do anything for any end*), the *infinite* supreme maxim (*I may always refrain from doing anything for any end*), and the *negative infinite* supreme maxim (*it is not the case that I may always refrain from doing anything for any end*).

Since it is logically impossible to refrain from doing anything for any end—this refraining would itself be a deed and, as such, constitute doing *something* for *some* end—the putative infinite supreme maxim is self-contradictory and thus cannot appear in any rational being's set of maxims: as such, it cannot be a supreme maxim. Similarly, the negative infinite supreme maxim, as involving the negation of a contradiction, is *tautologous* and, as such, is also ruled out as a possible supreme maxim. Therefore, the only possible supreme maxims are the positive and the negative supreme maxims and, as the two are genuine contradictories, any set of maxims can have at most one of the two;[4] I will also stipulate that any set of maxims has *exactly* one of the two.[5]

These two possible supreme maxims are exactly those Kant takes to be constitutive of noumenal moral worth: the positive supreme maxim is what Kant calls the *evil* supreme maxim while the negative supreme maxim is what Kant calls the *good* supreme maxim.[6] According to the positive supreme maxim, one takes oneself to be permitted to adopt any maxim, regardless of its morality, meaning that one will only adopt moral maxims insofar as they advance one's happiness, *viz.*, whatever ends one happens to have[7]—but this is nothing other than subordinating the Moral Law to the condition of one's happiness and thus is equivalent to the evil supreme maxim. Conversely, according to the negative supreme maxim, one takes oneself to be constrained in which maxims one can permissibly adopt but, as such constraint must derive from laws and, insofar as one is a *person*, such constraint can only come from the Moral Law, one takes oneself to be permitted to adopt only good maxims—

[4]This is directly related to Kant's self-professed rigorism: *cf.* R.6:22-5, 36.

[5]This stipulation should not be controversial—in Chapter II, I was willing to admit that one may not be conscious of all maxims in one's character—but it is important for my argument.

[6]*Cf.* R.6:30-1, 36.

[7]§II.2.

but this is nothing other than subordinating one's happiness to the condition of the Moral Law and thus is equivalent to the good supreme maxim.[8] Thus every person's character, both empirical at every time in the person's moral lifetime and intelligible, is such that it has *either* the good supreme maxim *or* the evil supreme maxim.

For Kant, one is noumenally good iff the supreme maxim of one's intelligible character is good and, conversely, one is noumenally evil iff the supreme maxim of one's intelligible character is evil:[9] this is unsurprising given that having the good supreme maxim simply amounts to adopting maxims only insofar as they are moral and, as such, possessing a good will. Note that one has exactly one supreme maxim and, insofar as one's *intelligible*, *i.e.*, *noumenal*, character is concerned, there can be no change from one supreme maxim to the other.

One last notable feature of noumenal moral worth is that, as one's supreme maxim must be *imputable* to oneself, it must somehow be freely adopted and, as such, *chosen*: this is precisely the doctrine of noumenal choice such that, in noumenally choosing one's intelligible character, one thereby chooses one of the two supreme maxims.[10] However, as Kant rightly notes, this choice—both of one's supreme maxim in particular and one's entire intelligible character more generally—is *inscrutable*.[11] This can be clarified by comparison to phenomenal choice: for a phenomenal choice to be free, it must be made in accordance with some maxim one has already freely adopted (or freely refrained from abandoning), but this model of choice cannot apply noumenally. Not only does this model inherently rely on temporal succession, but, even if this were generalized to mere logical grounding, it could still not be applied to noumenal choice: for since there are multiple possible intelligible characters even holding the supreme maxim constant, there could be no fundamental explanation (at least of which finite beings can be aware) for why some being chooses one possible intelligible character over another. Thus a person's choice of intelligible character must be absolutely and completely *spontaneous*.

Though this inscrutability thesis might appear to be in itself problematic for Kant, insofar as it only applies noumenally, it is actually a merit: it fits well with both

[8]Strictly speaking, this equivalence holds only for *persons*, but is sufficient for my purposes as goodness is only objectively valid for persons.

[9]*Cf.* R.6:47–8, 74–5n.

[10]It does not follow that the noumenal choice is *binary*, only that, so far as *goodness* is concerned, one can take it to be so.

[11]*Cf.* R.6:21n, 21–2, 31–2, 40, 51.

Transcendental Idealism's general conclusion that very little can be known about things-in-themselves and also with plausible views on freedom—*viz*., if a being is genuinely free, ultimately its choices must bottom out into a spontaneous or even somewhat arbitrary choice.[12]

## 2. Phenomenal Goodness I: Radical Evil and the Change of Heart

Similarly to the above account of noumenal goodness, phenomenal goodness at a time simply consists in having the good supreme maxim in one's empirical character at that time; *mutatis mutandis* for evil. However, one can say much more to further characterize phenomenal goodness so as to eventually find a useful schema for it, by which it can function in everyday moral judgments as an account of OGOP.

*Radical evil* is that which can be a *source* of evil, and, as such, it must not be confused with evil itself:[13] there is no contradiction in a person being good despite being radically evil—indeed, Kant claims (unhelpfully without argument) that *all finite persons* are radically evil.[14] More specifically, radical evil must be understood as a being's capacity to be *tempted*,[15] and, as such, one can construct an argument to support Kant's claim.

All finite persons are capable of self-deception because they cannot intuit themselves as they are *in themselves* but only insofar as they *appear* to themselves.[16] *I.e*., one cannot directly observe one's empirical character but only infer what it might be from one's deeds and the path one takes one's actual and hypothetical deliberations[17] to undergo or have undergone. Quite clearly one's knowledge of one's deeds alone, even if one could be sure that one knows about all of them—itself doubtful since one's deeds need not all be consciously performed and, further, one may forget or misremember even those that were performed consciously—is insufficient to determine one's empirical character: not only is it possible that one has maxims

[12] It cannot be that, at any time, one's choice of *empirical* character is ultimately inscrutable as this would contradict LNC: as such, while one's *noumenal* choice of character is inscrutable, one must always be able to provide some causal explanation for why one's empirical character is as it is.

[13] *Cf*. R.6:32; A.7:324.

[14] R.6:32–3. He specifically claims this for all *human beings*, but, given that he uses 'humanity' to mean 'rational nature', I take it that his claim can be extended.

[15] *Cf*. R.6:37–8; MM.6:397.

[16] *Cf*. G.4:407; R.6:38, 57, 63, 75; MM.6:392–3, 441, 447; TP.8:284–5; ET.8:329–30; PM.20:269.

[17] *I.e*., both deliberations in which one is deciding what to do in one's present circumstances and those in which one is deciding what one *would* do if one were in some possible set of circumstances.

one has not yet acted on, the deed of a maxim is by itself insufficient to specify its circumstance and end. Similarly, one's experience of one's deliberations is also insufficient: not only might some deliberation be subconscious or one might forget or misremember important details even of conscious deliberations, but, since there might always be some circumstances, deeds, or ends that have yet to play a role in one's deliberations, maxims might be in one's empirical character of which one has yet had no experience. By virtue of their finitude, all finite persons can have *no* guarantees either that those maxims they take themselves to hold are ones that they actually hold, or even that, were they somehow accurate concerning every maxim they take themselves to hold, there are not still further maxims they hold about which they have made no judgment whatsoever: in short, all finite persons are capable of self-deceit.[18]

Further, all finite persons, as finite and therefore not self-sufficient, have needs—*e.g.*, biological or social needs—the satisfaction of which can sometimes conflict with the Moral Law:[19] finite persons—indeed, all finite beings in general—need certain resources to survive and, since there is no guarantee that they can acquire these resources without violating the Moral Law and all finite persons are capable of self-deceit, it is possible that a finite person would adopt some or all of their needs as unqualified ends and, accordingly, adopt evil maxims. Therefore, having needs and the capacity for self-deceit are jointly sufficient for radical evil;[20] since all finite persons have both, all finite persons are therefore radically evil, even if they are not all evil *per se*.

Indeed, this is consistent with the doctrine from Chapter II that the Moral Law only *necessitates* finite persons' wills but *determines* those of infinite persons: since an infinite person intuits itself *qua* noumenon, it is incapable of self-deception and, further, as infinite persons are self-sufficient,[21] infinite persons cannot possibly adopt

[18]This result holds regardless of possible scientific progress concerning the ability to predict behavior given, *e.g.*, a person's brain-scan: even if, *per impossibile*, scientists were somehow *completely certain* that they possessed a complete Theory of Everything that perfectly encapsulates *all* natural laws, a person's brain-scan could only allow neuroscientists/psychologists to infer how the person would act in various circumstances and, importantly, *not* any ends the person has. Therefore, this predictive ability is insufficient to determine the person's empirical character: presumably the only way to go beyond "mere" prediction would be to ask the person what their ends are, but this would face the same problems highlighted above.

[19]*Cf.* G.4:399, 428; CPR.5:25; MM.6:393; A.7:267; LP.9:492.

[20]*Cf.* G.4:399; MM.6:380–1, 388, 394; CF.8:329–30.

[21]Strictly speaking, I only need one of these claims, but Kant himself would be committed to both.

evil maxims, meaning they are never radically evil, and, as such, have *holy wills* which necessarily choose according to the Moral Law.[22] A finite person, strictly speaking, *cannot* have a holy will even if it can have a *good* will: its finitude is inconsistent with holiness.

That all finite persons are radically evil allows one to make some sense of Kant's odd claim that all persons are evil, *i.e.*, have the evil supreme maxim in their empirical characters, *from birth*:[23] what Kant should rather have said is that, in an important sense, the evil supreme maxim must be considered the *default* for finite persons. Any finite person must have had some kind of *moral childhood* during which they develop their rational capacity but nonetheless adopt maxims:[24] since, on the one hand, a child's rational capacities are insufficiently developed, children cannot be conscious of FR and, as such, are not strictly speaking *persons*, while, on the other hand, children must have at least *some* maxims by the end of childhood, there can be no guarantee that, at the end of childhood, all the child's maxims will be good.[25] Since children are not persons, this is of no immediate moral concern—insofar as they are children, they cannot be morally responsible for the maxims they adopt—but it becomes a potential problem when children are sufficiently developed to be conscious of FR and, as such, become persons. Upon becoming a *person*, *i.e.*, a moral adult, those maxims one adopted during one's childhood "carry over" into one's empirical character, and while one cannot be responsible for one's empirical character during one's childhood, one is, as an adult, responsible for one's entire empirical character, including maxims carried over from one's childhood; however, due to radical evil, one cannot, even as a moral adult, know precisely which maxims constitute one's empirical character at a time with certainty, meaning that, even at the start of one's moral lifetime,

[22] *Cf.* G.4:414, 439; CPR.5:32–3; R.6:46–7; MM.6:396–7, 405.

[23] *Cf.* R.6:22, 25, 31, 39n, 58n, 66–7, 72, 76–7.

[24] *N.b.* This does not assume that, as is the case for humans, this developmental period takes place (partly) after birth, hatching, *etc.* and that all finite persons are therefore what one could term rationally *altricial*: even if this developmental period takes place exclusively *before* birth and that some finite persons are what one could term rationally *precocial*, such persons would nevertheless already have to have maxims upon being born, hatching, *etc.*—or else they would have no ends at all and, by the lemma established in §II.2, could not be persons.

[25] One can consider *all* human beings who, for whatever reason, fail to be persons for some period to be in a kind of *moral childhood*—even if using this language smacks of paternalism—including those with mental deficiencies or illnesses; *cf.* A.7:202–20. Since such moral children are human beings and yet not persons, I would expect Kant to hold that the duties one might take to be owed to moral children are, in fact, really duties to oneself to cultivate, *e.g.*, sympathy, consistent with how he addresses the problem of apparent duties to non-rational animals—*cf.* MM.6:443.

one may very well have bad maxims within one's empirical character dating back to one's childhood. Unless, therefore, one takes relevant precautions to abandon evil maxims one may retain from childhood, one's empirical character may very well be evil. Indeed, since one has a duty to be good, one ought to take such precautions and failure to do so implies that one's empirical character actually *is* evil: since a good will is an obligatory end, one must take whatever available means are necessary to uncover potentially evil maxims one may have.[26] Empirical goodness must require what Kant calls a *change of heart* from the evil supreme maxim to the good one.[27] Though Kant himself is unclear on this, such a change of heart can only be phenomenal: not only would it make no sense for there to be an atemporal change as change itself presupposes temporal succession, but Kant in his anthropological writings often refers to changes of heart occurring during people's lives.[28] *I.e.*, it is only by undergoing a positive change to the good supreme maxim—either from the evil supreme maxim or perhaps "directly" if one adopts the good supreme maxim immediately upon becoming conscious of FR[29]—that one can be phenomenally good.[30]

[26]Strictly speaking, the precautions are a necessary means to identifying any evil maxims one may have, which is itself a necessary means to abandoning all evil maxims one may have, itself a necessary means to having a good will.

[27]*Cf.* R.6:46–8, 51, 66–7.

[28]*Cf.* A.7:186, 236, 294–5, 321; LP.9:450, 455, 480–1, 487.

[29]Kant does not think the change of heart is at all common before middle age—*cf.* A.7:201, 294–5—but this must be understood as an empirical generalization rather than a modal claim concerning the possibility of undergoing a change of heart: Kant surely admits that healthy young adults are conscious of FR—indeed, he thinks the typical age at which one becomes a moral person is around sixteen (*cf.* LP.9:453), though this too seems to be an empirical generalization—and, as such, it must be possible for them to undergo a change of heart by OIC.

[30]It is not obvious how such a change of heart, necessary if one is to be phenomenally good, is possible given that *noumenal* goodness consists in having the good supreme maxim: that is, since one's intelligible character is simply one's empirical character considered as a unity throughout one's moral lifetime, how is it possible that one's intelligible character be *completely good* while one's empirical character be, at least at some times, *completely evil*? This, however, is a question that cannot be answered, at least not by finite beings such as ourselves, as it requires insight into the relationship between phenomena and their underlying noumena which is completely unavailable to us. But though we cannot understand its *possibility*, we can nevertheless be completely certain that it *must be the case*. First, one's intelligible character either has the supreme evil maxim or the supreme good maxim and, quite clearly, having an evil intelligible character precludes one from undergoing a change of heart—see the below discussion of ingenuine changes of heart. Second, due to radical evil, one must undergo a change of heart if one is to be phenomenally good; further, if one is actually a person, as one must be if one has any intelligible character at all, one *ought* to be phenomenally good and so, by OIC, one *can* be phenomenally good, meaning undergoing a change of heart is possible. If undergoing a change of heart is possible but incompatible with having an evil intelligible character, it must therefore be compatible with having a *good* intelligible character.

Allison—and perhaps even Kant himself—suggests that the change of heart must be from evil to *holiness*:[31] since such a change of heart cannot be a mere series of gradual changes but must instead be a *complete* and *utter* departure from the evil maxim,[32] it would seem that nothing short of a change to holiness would suffice. However, this is misguided: since finitude is incompatible with holiness and, quite plausibly, finitude is an essential property of finite persons, no finite person can rationally set holiness as its end by OIC.[33] Allison's claim notwithstanding, a change of heart to the good supreme maxim *is* a complete turn away from evil—it entails that one refuses to freely adopt bad maxims and that one is committed to abandoning any one might already have—though it surely is not a turn away from *radical* evil, as that is impossible for *any* kind of person: infinite persons are not radically evil and, as such, cannot turn away from it, while finite persons are radically evil as a consequence of their finitude and, as such, cannot eliminate radical evil. But, since radical evil is *not* evil but only its possible *source*, there is no need to eliminate radical evil provided that, through a change of heart, a finite person turns away from evil itself.

There can be no *ingenuine* changes of heart—either one's supreme maxim has changed or it has not—and, since the good supreme maxim is the contradictory of the evil one, a change of heart necessarily entails a complete turn away from evil. Further, if one has had a change of heart, one cannot choose to renege. Though it is rationally possible on the evil supreme maxim to abandon it and adopt the good supreme maxim—after all, the evil supreme maxim places no restrictions on what maxims one may abandon or adopt—it is not rationally possible on the good supreme maxim to abandon it and instead adopt the evil one: the evil supreme maxim is, naturally, an evil maxim ruled out by the good supreme maxim.[34] This is not to

[31] *Cf.* KTF 170–1, 175, 178–9; R.6:46–7, 66–7, 74–5n; MM.6:446. It is not obvious that Kant's claims must be understood as Allison suggests—indeed, I think they ought not to be so taken, precisely for the reasons I state shortly.

[32] *Cf.* MM.6:477.

[33] This is why Kant feels the need to argue for the postulate of *immortality*: essentially, he argues that if the Moral Law really does require that we become holy, we must therefore be infinite beings, if only temporally. I find this argument unpersuasive precisely because there is no need to think the Moral Law demands holiness. Indeed, I take it that, if FR is, as Kant claims, the *sole* fact of reason (CPR.5:31), there can only be a single postulate—*i.e.*, without an additional fact of reason, any purported postulate would either be extravagant and ungrounded, meaning it ought to be rejected as transcendent, or would be no more than a restatement of freedom—though an argument for this claim is beyond the scope of the present work.

[34] Since, however, radical evil entails that one can never know with certainty whether one *has* undergone a genuine change of heart, it is entirely possible that, though one has not, one deceives

say that a person who has undergone a change of heart cannot possibly have the evil supreme maxim thereafter, but only that they cannot have it through their own choice: *e.g.*, recuperation after a stroke may be sufficiently similar to one's childhood that one may have the evil supreme maxim at its conclusion without having freely chosen it and thereby being morally responsible for it at the time—though one surely *is* morally responsible for it once recuperation is complete and one ought then to have a *second* change of heart. Further, it is entirely possible—even fairly likely—that one who has undergone a change of heart will ever adopt or act on a bad maxim: having a good will does not entail moral omniscience, and, from radical evil, one may very well be mistaken about what maxim one takes oneself to be deliberating and, further, one can also be mistaken about whether the maxim one takes oneself to be deliberating is moral.[35] If, however, one takes one's proposed maxim to be bad, one cannot rationally act on it if one's supreme maxim is good, meaning all knowing violations of the Moral Law entail that any previous putative change of heart was only apparent; as such, a change of heart entails a subsequent period in which one *strives* to be good.[36]

## 3. Phenomenal Goodness II: Virtue and Progress

*Virtue* is the moral strength by which one does one's duty in the face of temptation.[37] However, since persons are, by definition, free and something can only be a person's end insofar as the person freely adopts it (or freely refrains from abandoning it), all temptation a person may face is, strictly speaking, self-imposed.[38] One can distinguish between two "grades" of virtue: the lower grade is the strength to *overcome* temptation that one has already adopted, while the higher grade is the strength to *abandon* temptation one might have adopted or else to *refrain* from adopting it. This is not to say that the higher grade is in itself morally superior to the lower grade, but only that the higher grade requires greater strength of will, *i.e.*, more virtue.

---

oneself into believing one has and one's later evil deeds *appear* to be a relapse into the evil supreme maxim; Kant discusses this very scenario—*cf.* R.6:68–9, 94.

[35]That is, even if one has correctly identified the maxim one is deliberating, one may mistakenly take it to be moral when it is really immoral or *vice versa*.

[36]*Cf.* CPR.5:32–3, 123–4; R.6:46–8, 51, 66–7, 74–5n; MM.6:446.

[37]*Cf.* R.6:23n, 57; MM.6:380, 390, 392–3, 394, 396–7, 405, 407–8, 477; WE.8:35–6; TP.8:278–9; LP.9:492.

[38]*Cf.* Allison, KTF 164; R.6:57, 93; MM.6:409; TP.8:283n.

The methods by which even someone who has had a change of heart may nevertheless be tempted to transgress duty are *self-deceit* and *rationalization*, where 'rationalization' relies on self-deceit because it consists of convincing oneself that a bad maxim is in fact moral. Therefore the lower grade of virtue consists in one's ability to both recognize that one is possibly rationalizing and to introspect in an attempt to combat self-deceit: since one has, by stipulation, already incorporated the temptation in question, this kind of investigation is contrary to one's happiness and requires genuine strength of will to undertake. The higher grade can be further subdivided into *preemptive* and *reactionary* components, such that, preemptively, one introspects to consider what ends *could* tempt oneself and attempts to refrain from adopting them, while, reactively, one goes beyond merely overcoming the temptation through the lower grade of virtue by attempting to abandon it; even if, as a free being, one knows one *can* always abandon or refrain from adopting some tempting end, as noted above, this requires even greater strength of will than the lower grade.

The higher grade of virtue clarifies Kant's repeated claim that a sufficiently virtuous person has an *ever-cheerful heart*:[39] since one's happiness is simply the satisfaction of one's ends and, so long as one has sufficiently strong virtue of the higher grade, one can abandon and refrain from adopting all ends that conflict with duty, one can therefore always be happy in doing one's duty—though such a feat requires remarkable will-power. Indeed, although the lower grade of virtue alone is sufficient for one to be *good*, it is consistent with one's being completely miserable while being good: only through the higher grade of virtue can one have something of a guarantee that one can be happy in doing one's duty. Therefore, one is rationally required to develop one's virtue to include the higher grade: since happiness is a natural end, one must, by PHI, will its necessary means; however, insofar as one is good, the necessary means to happiness include developing the higher grade of virtue, so one must set developing the higher grade of virtue as one's end. A good person cannot rest content with merely the lower grade of virtue, even though it is sufficient for being good, and is rationally required to develop their virtue into its higher grade. Kant goes further, claiming that virtue necessarily atrophies unless it is actively developed:[40] since one can never *know* that one has sufficient virtue to combat whatever temptation one may face—precisely because of the possibility of

[39] *Cf.* Allison, KTF 163–4; R.6:23n; MM.6:377–8, 484–5; A.7:235; LP.9:485, 499.
[40] *Cf.* CPR.5:32–3; R.6:93; MM.6:409.

self-deceit and rationalization that virtue is meant to combat—one can never secure oneself permissible means to one's happiness by contenting oneself with the degree of virtue one already has developed. Any move to stop developing one's virtue further is itself a rationalization that must be combated through virtue.

As such, a change of heart is necessarily followed by a period in which one ceaselessly develops one's virtue, and this may explain why Allison and Kant have mistaken the duty to be *good* for a duty to be *holy*. In particular, there are two possible misunderstandings here. First, one might mistake a permanent duty to develop one's virtue for a duty to have *perfectly developed* virtue, *i.e.*, the perfected ability to combat all temptation, even before it is adopted, which is therefore equivalent to holiness. Second, one might mistake the change of heart as described above for a resolution to develop one's virtue rather than a resolution to be *good* and, as a necessary means to that end, therefore to improving one's virtue: the former would indeed be insufficient for a change of heart as it is nothing other than an attempt to break away from evil gradually; the latter, however, *is* a complete break from evil. This potential misunderstanding is analogous to mistaking a *moral law*, *e.g.*, the duty not to sadistically inflict pain on unsuspecting passersby, for the *Moral Law*: while the moral law in question is morally binding, it is only binding *as a result* of the Moral Law—to claim otherwise is heteronomy.

This account of phenomenal goodness—that one first undergoes a change of heart and subsequently enters a period of ceaselessly developing one's virtue—is clearly objectively valid for people in their everyday lives: indeed, it is also objectively real for people in their everyday lives, as it provides a schema by which one can judge whether someone—oneself or another—is phenomenally good. However, it cannot yet function as an account of OGOP as its schema has fairly limited use in normal life: first, one cannot obviously tell whether someone, even oneself, has genuinely had a change of heart or is only rationalizing, and, second, it is difficult to discern whether another person's virtue has developed. This can be possible if one observes them in a scenario one knows is tempting for them and later observes them in a similarly tempting scenario in which they more easily do their duty, but not only can it be difficult to evaluate how tempting something is for another, this requires multiple observations and, as such, seems at odds with our apparent ability in everyday life to find grounds to take someone one has just met to be good or bad.

This account of goodness therefore requires further characterization regarding its schema to fully play the role of OGOP.

## 4. A Schema for Goodness: Conscientiousness and Reflectiveness

One's *conscience* is one's judgment as to whether one has done all one can to determine whether the maxim one takes oneself to be deliberating is moral.[41] As such, one's conscience cannot err: one may mistakenly take oneself to be deliberating a maxim one in fact is *not* deliberating, or even mistakenly judge that a moral maxim is immoral and *vice versa*, but one cannot mistakenly take oneself to have put in a good faith effort to uncover any possible self-deceit or rationalization. If one *has* done one's due diligence, it follows that errors one might have made concerning the maxim one is deliberating and its morality can only be uncovered through additional experience that is not currently available to oneself at the time of deliberation: thus, by OIC, one is only obligated to screen one's maxims for their morality in accordance with the requirements of one's conscience. As such, *conscientiousness* is the maxim by which one always heeds one's conscience such that one never adopts or acts on a maxim unless one conscientiously feels it permissible to do so; on the other hand, *unconscientiousness* is the maxim by which one does not (always) heed one's conscience.

*Reflection* is the activity of either reexamining some deliberation one faced in the past or hypothesizing some unexperienced scenario and considering how one might deliberate in it—these can be called *past-oriented* and *future-oriented* reflection, respectively—so that one can better prepare for future deliberations.[42] In past-oriented reflection, one considers: (i) whether the maxim one *took* oneself to be considering was in fact the maxim one *was* considering and, if not, *why* one might have been deceived; (ii) whether the maxim one took oneself to have acted upon is moral and, if not, why one was tempted to rationalize; (iii) whether the maxim one *currently takes* oneself to have acted upon is moral and, if not, what one ought to have done instead; (iv) whether one was at all tempted even if one would not have acted differently, and; (v) how one might overcome or even abandon whatever temptation one faced. Similarly, in future-oriented reflection, one considers: (i) what one

[41] *Cf.* R.6:70n, 185–6, 189; MM.6:394, 399–401, 430, 437–8, 438n; ET.8:330; LP.9:495.
[42] *Cf.* MM.6:485.

ought to do; (ii) what one might be tempted to do, and; (iii) how one might preemptively counteract this temptation. Finally, *reflectiveness* is the maxim by which one reflects, both in past- and future-orientations, when one can spare the time and, further, makes time for such reflection.

I take it that one must be committed to both conscientiousness and reflectiveness to develop one's virtue: more specifically, one's commitment to conscientiousness amounts to one's having the lower grade of virtue—one cannot hope to overcome temptation unless one makes a good faith effort to identify self-deceit and rationalization—while reflectiveness is both necessary and sufficient for further developing both the lower and higher grades. As with the requirement of developing one's virtue, it is not the case that a good person is conscientious and reflective for their own sake, but rather because they are rationally required by the good supreme maxim.[43]

Further, both conscientiousness and reflectiveness have their own schemata by which one can judge someone to be conscientious or reflective, even from only a single encounter,[44] and, since conscientiousness and reflectiveness are both necessary and sufficient for empirical goodness, their schemata can therefore function as a schema for empirical goodness itself. As such, I have developed a Kantian account of goodness that can play the role OGOP: only by judging someone to be both conscientious and reflective can one judge them to be *good*.

[43]This clarifies why education—specifically *moral* education—cannot be a prerequisite for goodness, so that even those who were never "taught any better" can nevertheless be good: while a moral education is indispensable for improving one's understanding of what the Moral Law actually *is* and, as such, what one's duties *are*, no amount of moral education can *make* one conscientious or reflective and, even if it is morally preferable not to err in one's judgments concerning what one's duty is, one cannot be morally obligated to not so err, but only to be conscientious and reflective. One therefore has a duty, as a moral adult, to undergo moral education as much as one can if one is to remain conscientious—one cannot take oneself to have done all one can do to ascertain one's duty if one refrains from moral education which will help one in ascertaining one's duty—and, further, one also has a duty to help provide moral education for children so that they may themselves become good upon becoming persons.

[44]I cannot give more detail on these schemata without entering a domain—psychology—in which I lack the knowledge to do anything beyond merely restating facts about ordinary experience: still, I do find one can "just tell", so to speak, that someone is conscientious, and the same for reflectiveness.

# Conclusion
## Ordinary Goodness for Ordinary People

Considered as a whole, Kant's system not only *can* but *must* have a place for a conception of goodness that applies to people in their everyday lives—that is, a conception of ordinary goodness for ordinary people—as can be seen from the three claims which I have shown to comprise Kant's doctrine of Transcendental Compatibilism, his unique response to the free will problem. First, he demonstrates, *contra* incompatibilists, that, though the phenomenal world necessarily is completely causally determined by laws of nature, this determinism is nevertheless completely compatible with transcendental freedom, *i.e.*, the libertarian-style freedom to act independently of being so determined by natural laws. Second, he establishes, *contra* compatibilists, that transcendental freedom alone is both necessary and sufficient for moral responsibility: though the kind of *practical* freedom necessary and sufficient for moral responsibility—namely, *autonomy*—is conceptually distinct from transcendental freedom, they are actually equivalent and are simply different ways of understanding the same underlying property of freedom. Third, he shows that ordinary people in their everyday lives are genuinely morally responsible and, as such, *free* because of the Fact of Reason: that is, normal human beings at typical decision-points can be conscious of their own moral deliberation and, from this, have a practical warrant for taking themselves to be both morally responsible and free—further, since, there can be no theoretical warrant either way, this practical warrant is indefeasible and, as such, constitutes genuine knowledge and certainty. Thus while the first, purely *metaphysical* claim has no immediate bearing for moral responsibility and, even alongside the second, *moral* claim, has no immediate bear-

ing for the moral responsibility of ordinary people in their everyday lives, taking all three claims together—metaphysical, moral, and anthropological—establishes that ordinary people are genuinely both morally responsible as well as completely determined by natural laws; as such, not only *can* Kant's system have room for an account of ordinary goodness for ordinary people, it *must* if it is not to fall into complete absurdity. Although all three claims are certainly part of Kant's system, few commentators have recognized all three—fewer still have drawn out the implications gathered here.

Moreover, using the insights of Transcendental Compatibilism, one can see that Kant actually provides an account of ordinary goodness for ordinary people, albeit implicitly: it is nothing more than conscientiousness and reflectiveness. That is, someone is good insofar as they never intentionally act unless they take themselves to have done everything possible at the decision-point to ensure their action is moral and, further, they both reconsider previous decisions they have faced and consider future decisions they may face in an attempt to improve their ability to overcome any temptation to be evil. Thus while one may ultimately disagree with Kant's particular account of goodness—as well as those of related concepts such as character and virtue—one cannot deny its rightful place in his overall system.

My general approach supports what one might call the *Holistic Thesis* for Kant's system: to properly understand *any* component of Kant's system, one must understand its place in the system as a whole and, as such, no parts of Kant's system can properly be separated from the rest.[1] By no means does this work constitute a conclusive *proof* of the Holistic Thesis, but it does make the latter seem more plausible than it might otherwise appear since Transcendental Compatibilism is quite clearly a doctrine concerning the relation between various domains of Kant's system—metaphysical, moral, and anthropological—and, as I have shown, it is only by examining these domains insofar as they *interrelate* that one can fully address objections and problems which, on the face of it, only concern each domain in *isolation*.

Contrary to the relatively common criticisms that Kant's moral theory is either an empty formalism that is unable to guide action or else exceedingly rigid and thereby inapplicable to ordinary life,[2] it provides an account of goodness that

[1] As such, one could also call this the *General Inseparability Thesis* as it entails, among other things, a rejection of what Allison calls the *Separability Thesis*—*cf.* KTI xiii–iv, 6–11.

[2] *Cf.* O'Neill, AP 1–4.

can not only function in evaluative judgment but also guide one's deliberations concerning how one ought to act. Furthermore, not only is Kant's account of goodness quite plausible in its own right, but it—particularly its core component of conscientiousness—seems to be almost ubiquitous in everyday moral experience, not to mention art and literature: indeed, even in the midst of an obvious farce, Gilbert and Sullivan's Pirate King hits upon something central when he proclaims one must "[a]lways act in accordance with the dictates of your conscience…, and chance the consequences".

# Bibliography


Allison, Henry E. "The Antinomy of Pure Reason Section Nine (A51567/B54395)" in *Essays on Kant*. Oxford University Press, 2012.

—. *Kant's Theory of Freedom*. Cambridge University Press (CUP), 1990.

—. *Kant's Transcendental Idealism: an Interpretation and Defense*. Yale University Press, 2004.

—. "The Non-Spatiality of Things in Themselves for Kant." *Journal of the History of Philosophy*, vol. 14, no. 3, 1976, 313–21.

Bunch, Aaron. "'Objective Validity' and 'Objective Reality' in Kant's B-Deduction of the Categories." *Kantian Review*, vol. 14, no. 2, 2010, 67–92.

Chipman, Lauchlan. "Kant's Categories and Their Schematism." *Kant-Studien*, vol. 63, no. 1, 1972, 36–50.

Einstein, Albert. *Relativity: the Special and the General Theory*. Translated by Robert W. Lawson. Pi Press, 2005.

Falkenstein, Lorne. "Kant's Argument for the Non-Spatiotemporality of Things in Themselves." *Kant-Studien*, vol. 80, no. 1, 1989, 265–83.

Foot, Philippa. "Morality as a System of Hypothetical Imperatives" in *Virtues and Vices*. Oxford University Press, 2002.

Gilbert, William Schwenck and Arthur Sullivan. *The Pirates of Penzance, or The Slave of Duty*. Wikisource, 2017.

Guyer, Paul. *Knowledge, Reason, and Taste: Kant's Response to Hume*. Princeton University Press, 2008.

Herman, Barbara. "Making Room for Character" in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*. Edited by Stephen Engstrom and Jennifer Whiting. CUP, 1998.

Kant, Immanuel. *An Answer to the Question: What is Enlightenment?* in *Practical Philosophy*.
—. *Anthropology from a Pragmatic Point of View* in *Anthropology, History, and Education*. Translated by Robert B. Louden.
—. *Anthropology, History, and Education*. Edited by Günter Zöller and Robert B. Louden. CUP, 2014.
—. *The Conflict of the Faculties* in *Religion and Rational Theology*. Translated by Mary J. Gregor and Robert Anchor.
—. *Critique of Practical Reason* in *Practical Philosophy*.
—. *Critique of Pure Reason*. Translated and edited by Paul Guyer and Allen W. Wood. CUP, 2017.
—. *Critique of the Power of Judgment*. Edited by Paul Guyer, translated by Paul Guyer and Eric Matthews. CUP, 2007.
—. *The End of All Things* in *Religion and Rational Theology*. Translated by Allen W. Wood.
—. *Groundwork of the Metaphysics of Morals* in *Practical Philosophy*.
—. *The Jäsche Logic* in *Lectures on Logic*. Translated and edited by J. Michael Young. CUP, 1992.
—. *Lectures on Pedagogy* in *Anthropology, History, and Education*. Translated by Robert B. Louden.
—. *Metaphysical Foundations of Natural Science* in *Theoretical Philosophy after 1781*. Translated by Michael Friedman.
—. *The Metaphysics of Morals* in *Practical Philosophy*.
—. *On a Discovery Whereby Any New Critique of Pure Reason is to be Made Superfluous by an Older One* in *Theoretical Philosophy after 1781*. Translated by Henry Allison.
—. *On the Common Saying: That May be Correct in Theory, but It is of No Use in Practice* in *Practical Philosophy*.
—. *Practical Philosophy*. Translated and edited by Mary J. Gregor. CUP, 1999.
—. *Prolegomena to Any Future Metaphysics that Will be Able to Come Forward as Science* in *Theoretical Philosophy after 1781*. Translated by Gary Hatfield.
—. *Religion and Rational Theology*. Edited by Allen W. Wood and George di Giovanni. CUP, 2005.
—. *Religion Within the Boundaries of Mere Reason* in *Religion and Rational Theology*. Translated by George di Giovanni.

—. *Theoretical Philosophy after 1781*. Edited by Henry Allison and Peter Heath. CUP, 2010.

—. *Toward Perpetual Peace* in *Practical Philosophy*.

—. *What Does It Mean to Orient Oneself in Thinking?* in *Religion and Rational Theology*. Translated by Allen W. Wood.

—. *What Real Progress has Metaphysics Made in Germany Since the Time of Leibniz and Wolff?* in *Theoretical Philosophy after 1781*. Translated by Peter Heath.

Korsgaard, Christine M. *Creating the Kingdom of Ends*. CUP, 1999.

Langton, Rae. "Objective and Unconditioned Value." *Philosophical Review*, vol. 116, no. 2, 2007, 157–85.

Łuków, Paweł. "The Fact of Reason: Kant's Passage to Ordinary Moral Knowledge." *Kant-Studien*, vol. 84, no. 2, 1993, 204–21.

O'Neill, Onora. *Acting on Principle: an Essay on Kantian Ethics*. CUP, 2013.

—. *Constructing Authorities: Reason, Politics and Interpretation in Kant's Philosophy*. CUP, 2015.

—. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. CUP, 1990.

Rawls, John. *Lectures on the History of Moral Philosophy*. Edited by Barbara Herman. Harvard University Press, 2003.

—. *Political Liberalism*. Columbia University Press, 2005.

Ryle, Gilbert. *The Concept of Mind*. Penguin Books, 1978.

Shapiro, Jonathan D. "All's Different that Ends Differently: Counterfactuals, Determinism, and Antifinitism." 2020.*

Van Cleve, James. *Problems from Kant*. Oxford University Press, 1999.

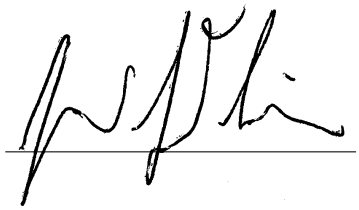
Walker, Ralph C. S. *Kant*. Routledge & Kegan Paul, 1978.

Weizsäcker, Carl Friedrich V. "Kant's 'First Analogy of Experience' and Conservation Principles of Physics." *Synthese*, vol. 23, no. 1, 1971, 75–95.

*This work was assessed as one of six submitted essays for the BPhil; it was therefore omitted in the original submitted version of this thesis for marking anonymity reasons.

# Declaration of Originality of Authorship of a Thesis Submitted for Examination in the BPhil in Philosophy in the Academic Year 2019–2020

I, Jonathan D. Shapiro, affirm that my submitted thesis entitled 'Ordinary Goodness for Ordinary People: an Essay on the Kantian System' is my own original work, except where otherwise stated, and that I have only shown drafts of it to my supervisor, Prof. Ralph C. S. Walker, in approximately eight hours of supervision, and that I have received help in its preparation from no other Oxford Philosophy Faculty members.

Signed:  Date: June 30, 2020