Kant, Wittgenstein, and Transcendental Philosophy[1]

Robert Hanna
Univ. of Colorado at Boulder, USA

Philosophy … is in fact the science of the relation of all cognition and of all use of reason to the ultimate end of human reason, to which, as the highest, all other ends are subordinated, and in which they must all unite to form a unity. The field of philosophy in this cosmopolitan sense can be brought down to the following questions: 1. *What can I know?* 2. *What ought I to do?* 3. *What may I hope?* 4. *What is man? Metaphysics* answers the first question, *morals* the second, *religion* the third. Fundamentally, however, we could reckon all of this as anthropology, because the first three questions relate to the last one.

--I. Kant (*JL* 9: 24-25)

[In 1916] Wittgenstein read [Dostoevsky's *The Brothers Karamazov*] so often he knew whole passages of it by heart, particularly the speeches of the elder Zossima, who represented for him a powerful Christian ideal, a holy man who could "see directly into the souls of other people."

[In 1918, at a prisoner-of-war camp in Como] Wittgenstein got to know Hänsel after attending a class on logic that Hänsel was giving to prisoners who hoped on release to train as teachers. This led to regular discussions between them, during which Wittgenstein led Hänsel through the elements of symbolic logic and explained the ideas of the *Tractatus* to him. They also read Kant's *Critique of Pure Reason* together.

--R. Monk[2]

The limit of language is shown by its being impossible to describe the fact which corresponds to (is the translation of) a sentence, without simply repeating the sentence. (This has to do with the Kantian solution of the problem of philosophy.)

--L. Wittgenstein (*CV*: 10^e)

May God grant the philosopher insight into what lies in front of everyone's eyes.

--L. Wittgenstein (*CV*: 63^e)

I. Introduction

What is transcendental philosophy? The short-and-sweet answer is that transcendental philosophy is *rational anthropology*: i.e., *the philosophy of us and our manifest natural world*. A slightly longer answer is that it is the philosophy which tells us

(1) how the manifest natural world *must be*, in order to conform to the innately-specified structures of the basic cognitive and practical capacities of rational human animals,

(2) how rational human animals *must choose, act, and try to live*, in order to conform to the highest norms, rules, and standards they legislate for themselves, and also, tragically, almost inevitably fail to meet,

and

> (3) how philosophy *must not be*, because otherwise it will inevitably, and tragically, fall
> into logical antinomy, radical skepticism, and cognitive or practical self-alienation.

In view of (2) and (3), we can clearly see how the slightly longer answer is also

bittersweet. In this regard, transcendental philosophy is a foundational anticipation and/or

expression of Existential philosophy (e.g., Augustine's writings, Pascal's writings,

Spinoza's writings, Schopenhauer's writings, Nietzsche's writings, Kierkegaard's

writings, Heidegger's early writings, Camus's writings, or Sartre's early writings) and

literature (e.g., ancient Greek tragedy, Shakespeare's tragedies, Cervantes's *Don Quixote*,

Dostoevsky's novels, or Rilke's poetry), and thereby adequately captures a

metaphysically robust, morally-charged, and ultimately *tragic* sense of rational human

life.[3] Otherwise put, transcendental philosophy is *the philosophy of the rational human

condition*.[4]

Transcendental philosophy was created or discovered in the 18th century by

Immanuel Kant, and brilliantly worked out by him in the *Critique of Pure Reason*

(1781/1787), *Prolegomena to any Future Metaphysics* (1783), *Groundwork of the

Metaphysics of Morals* (1785), the *Critique of Practical Reason* (1788), the *Critique of

the Power of Judgment* (1790), and *Religion within the Boundaries of Mere Reason*

(1793). Transcendental philosophy was also re-created or re-discovered in the 20th

century by Ludwig Wittgenstein, and equally brilliantly worked out by him in the

Tractatus Logico-Philosophicus (1921) and the *Philosophical Investigations* (1953). The

purpose of this essay is to examine some central themes in the transcendental philosophy

jointly developed and practiced by Kant and Wittgenstein. More precisely, I will focus on

these four topics:

(i) *The Subject and the World*: i.e., Kant's transcendental idealism/empirical realism in the first *Critique* and Wittgenstein's solipsism/pure realism in the *Tractatus*,

(ii) **If God's Existence is Unprovable, Then is Everything Permitted? Denying** *Wissen* **and Making Room for** *Glauben*: i.e., Kant's moral theology and Wittgenstein's ethical mysticism in the *Tractatus*,

(iii) *How and Why Logic is Transcendental*: i.e., Kant's doctrine of the transcendental ideality of logic in the first *Critique*, and early Wittgenstein's thesis that "logic is transcendental" in the *Tractatus*,

and finally

(iv) *The Critique of Self-Alienated Philosophy*: i.e., Kant's critical meta-philosophy in the first *Critique*, early Wittgenstein's view of logical analysis as "not a theory, but an activity" in the *Tractatus*, later Wittgenstein's thesis that "logic is grammar"in the *Investigations*, and, finally, later Wittgenstein's thesis that "the real discovery is the one that makes me capable of stopping doing philosophy when I want to" in the *Investigations*.

My conclusion will be that transcendental philosophy, as jointly done by Kant and Wittgenstein, not only constitutes a defensible, distinctive, and serious *alternative* to the other basic kinds of classical modern philosophy—i.e., Rationalism, Empiricism, and the more extreme forms of Idealism and/or Anti-Realism—but also provides a defensible, distinctive, and serious *successor-discipline* to Analytic philosophy. And this is precisely because transcendental philosophy, unlike the other kinds of modern or contemporary philosophy, alone provides "insight into what lies in front of everyone's eyes."

II. The Subject and the World

> Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them *a priori* through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an *a priori* cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus…. (*CPR* Bxvi)

> This is the way I have travelled: Idealism singles men out from the world as unique, solipsism singles me alone out, and at last I see that I too belong with the rest of the world, and so on the one side *nothing* is left over, and on the other side, as unique, *the world*. In this way idealism leads to realism if it is strictly thought out. (*NB*: 85ᵉ)

II.1 Transcendental Idealism and Empirical Realism

Famously or notoriously, Kant is a transcendental idealist. But what does this actually mean? According to Kant, a mental representation is *transcendental* when it is either part of, or derived from, our non-empirical (hence a priori) innately-specified spontaneous cognitive capacities (*CPR* A11/B25) (*P* 4: 373n.). Correspondingly, Kant's thesis of *representational transcendentalism* says that all the forms or structures of the representational contents of human cognitions are spontaneously generated a priori by our innately-specified cognitive capacities (i.e., cognitive *faculties*, or cognitive *powers*). And Kant's thesis of *cognitive idealism* says that all the proper objects of our specifically human sort of cognition—the notion of "being human" in this sense does not mean "belongs to the biological species *homo sapiens*," but rather means "rational but also embodied, and possessing our special sort of sensibility"—are nothing but appearances or phenomena, and never things-in-themselves.

Appearances (*Erscheinungen*) or *phenomena* are intersubjectively mind-dependent objects of actual or possible human sense perception. Things appear in this sense, precisely because they really *are* what they appear to, not because they are really *other* than what they appear to be. Only a *mere* appearance (*bloße Erscheinung*) or

illusion (*Schein*) is individually or egocentrically mind-dependent, and represents something to be other than what it really is. Furthermore, for Kant an appearance or phenomenon is token-identical with the intensional content of the objectively valid mental representation used to refer to it. Now in order to understand this token-identity thesis, we must also understand that for Kant all appearances or phenomena come in two different flavors:

(1) partially or wholly undetermined,

and

(2) fully determined.

Partially or wholly undetermined appearances are unconceptualized objects of empirical intuition, that is, the objects of "blind" intuitions. Fully determined appearances are fully conceptualized objects of empirical intuition, also known as *objects of experience*. So in other words, for Kant Transcendental Idealism says that *the experience of objects*—specifically as regards its well-formed intensional representational *content*, and specifically *not* as regards its psychological representational *vehicle*, which would entail psychologism—is identical to *the objects of experience*:

> You put the matter quite precisely when you say "The content (*Innbegriff*) of a representation is itself the object; and the activity of the mind whereby the content of a representation is represented is what is meant by 'referring to the object'." (*PC* 11: 314)

Noumena, by contrast, are *non-appearances* or *non-phenomena*. But, like appearances or phenomena, noumena also come in two flavors.

(1) *Things-in-themselves* or *noumena in the positive sense* are beings (whether objects or subjects) that exist independently of human minds, are non-sensory, and have a nature or real essence consisting of a set of intrinsic non-relational properties. Given Kant's theory of space and time in the Transcendental Aesthetic, it also follows directly

from the mind-independence and non-sensory character of things-in-themselves or positive noumena that they are *non-spatiotemporal*. But all rational human cognition is sensory and spatiotemporal. Therefore things-in-themselves or positive noumena are uncognizable and hence scientifically unknowable by creatures with minds like ours. They could be known *only* by a divine cognizer, or a being with a capacity for intellectual intuition. Examples of things-in-themselves or positive noumena, *were* they to exist, would be Platonic Forms or Ideas, Leibnizian monads, angelic spirits, God, unobservable physical microstructures (e.g., Locke's "real internal constitutions" of physical things), and so-on.

(2) *Noumena in the negative sense* comprise a class of things that is larger than the class of things-in-themselves or positive noumena. All positive noumena are negative noumena, but not all negative noumena are positive noumena. More precisely, negative noumena are any beings (whether objects or subjects) that have some non-sensory intrinsic properties: hence in that respect they transcend the bounds of human sensibility (e.g., non-Euclidean space, or reversible time). But in principle a negative noumenon can also *be* an empirical object, that is, an empirical object with a non-sensory intrinsic property. The fact that John, Paul, George, and Ringo were four includes a non-sensory intrinsic property, *their fourness*. The fact that Paul and Ringo still actually exist includes a non-sensory intrinsic property, *their actual existence*. The fact that a sunset is beautiful includes a non-sensory intrinsic property, *its beauty*. The fact that you are a human person includes a non-sensory intrinsic property, your non-denumerably absolutely intrinsically valuable human *dignity*. And so-on.

In any case, it is crucially important to note that for Kant *both the existence and also the non-existence of things-in-themselves or positive noumena are logically consistent with the thesis of Transcendental Idealism*. That is: Given the truth of Transcendental Idealism, it is logically possible that things-in-themselves or positive noumena exist, and also logically possible that they do not exist. This distinguishes Kant's radically weak sort of idealism from other more extreme types of Idealism and/or Anti-Realism, e.g., Cartesian skeptical idealism, Berkeleyan subjective idealism, Hegelian or neo-Hegelian objective idealism, Nietzschean "will to power" relativism, Carnapian constructive empiricism or phenomenalism, Quinean "web of belief" pragmatism, cultural relativism, and so-on. Moroever, and even more importantly, because things-in-themselves or positive noumena are both uncognizable and scientifically unknowable, then rational human animals *know with certainty that they cannot know* whether things-in-themselves or positive noumena exist or not, or what their nature is. *Kant is thus radically agnostic about the existence or non-existence of things-in-themselves or positive noumena, and also about their nature*.

Against the backdrop of the fundamental Kantian distinction between appearances or phenomena, and things-in-themselves or positive noumena, then transcendental idealism can be more carefully formulated as a two-part thesis:

*Transcendental idealism = (1) Representational Transcendentalism +
(2) Cognitive Idealism.*

(1) *Representational Transcendentalism* = Necessarily, all the forms or structures of rational human cognition are generated a priori by the empirically-triggered, yet stimulus-underdetermined, activities of our innately-specified spontaneous cognitive capacities (= cognitive faculties, cognitive powers).

(2) *Cognitive Idealism* = Necessarily, all the proper objects of rational human cognition are nothing but appearances or phenomena (i.e., mind-dependent, sensory, spatiotemporal, directly perceivable objects) and never things-in-themselves or positive

noumena (i.e., mind-independent, non-sensible, non-spatiotemporal, real essences constituted by intrinsic non-relational properties) (*CPR* A369 and *Prol* 4: 293-294, 375).

Now (1) + (2) = Kant's "Copernican revolution" in metaphysics, as famously or notoriously formulated in the first epigraph of this section:

> Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them *a priori* through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an *a priori* cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus…. (CPR Bxvi),

which I will rationally reconstruct as *The Conformity Thesis*:

> It is *not* the case that rational human minds passively conform to the objects they cognize, as in classical Rationalism and classical Empiricism. On the contrary, necessarily, all the proper objects of rational human cognition conform to—i.e., they have the *same* form or structure as, or are *isomorphic* to—the forms or structures that are non-empirically generated by our innately-specified spontaneous cognitive capacities. So necessarily the form or structure of the manifestly real natural world we cognize is *mind-dependent*.

In this way, all versions of Transcendental Idealism hold that the manifest natural world we directly perceive must *in some sense* conform to the non-empirical structures of our innate cognitive capacities. Many Kantians are also committed to *Strong Transcendental Idealism*, or STI for short, which says:

> (1) Things-in-themselves (a.k.a . "positive noumena," or Really Real things, i.e., things as they could exist in a "lonely" way, altogether independently of rational human minds or anything else, by virtue of their intrinsic non-relational properties) really exist and cause our perceptions, although rational human cognizers only ever perceive mere appearances or subjective phenomena.

> (2) Rational human cognizers actually impose the non-empirical structures of their innate cognitive capacities onto the manifest natural world they cognize, i.e., necessarily, all the immanent forms or structures of the proper objects of human cognition are literally *type-identical to* the forms or structures that are non-empirically generated by our innately-specified spontaneous cognitive capacities.

and

> (3) Necessarily, if all rational human cognizers went out of existence, then so would the manifest natural world they cognize.

But some Kantians—or at least *one* contemporary Kantian, i.e., R.H.—think(s) that STI

is objectively false and are committed instead only to the objective truth of *Weak* or

Counterfactual Transcendental Idealism, or WCTI for short, which says:

> (i) Things-in-themselves are logically possible, but at the same time it is necessarily unknowable and unprovable whether things-in-themselves exist or not, hence for the purposes of legitimate metaphysics, epistemology, and ethics, they can be ignored (= *methodological eliminativism about things-in-themselves*).

> (ii) Necessarily, all the proper objects of rational human cognition have the *same* forms or structures as—i.e., they are *isomorphic* to—the forms or structures that are non-empirically generated by our innately-specified spontaneous cognitive capacities, but at the same time those worldly forms or structures are *not* literally type-identical to those mental forms or structures (= *the necessary-conformity-without-literal-identity thesis*).

> (iii) It is a necessary condition of the existence of the manifest natural world that if some rational human animals *were* to exist in that world, then they *would* directly cognize that world via essentially non-conceptual (i.e., intuitional) representational content, at least to some extent (= *the counterfactual cognizability thesis*).

and

> (iv) The manifest natural world has at some earlier times existed without rational human animals to cognize it directly, and could exist even if no rational human animals existed to cognize it directly, even though some rational human animals now actually exist in that world—e.g., I (R.H.) now actually exist in the manifest natural world—who do in fact cognize it directly, at least to some extent (= *the actual existence thesis*).

Here is a slightly more precise formulation of WCTI's crucial thesis (iii), the

counterfactual cognizability thesis:

$$\text{Syn Ap} \sim (\Box x)\,(\Box y)\,[\text{MNWx} \rightarrow \{(\text{RHAy \& MNWy}\} \sim\rightarrow \text{DCyx}\}]$$

<u>Definitions</u>:

Syn Ap ~ = synthetically a priori necessarily
P ~→ Q = If P were the case, then Q would be the case.
MNWx = x belongs to the manifest (i.e., apparent) natural world
MNWy = y belongs to the manifest (i.e., apparent) natural world
RHAy = y is a rational human animal
DCyx = y directly cognizes (i.e., essentially non-conceptually cognizes) x, at least to some extent

<u>Natural Language Translation</u>:

Synthetically a priori necessarily, anything that belongs to the manifest natural world is such that if some rational human animals *were* to exist in that world, then they *would* directly cognize that thing, at least to some extent.

2 Crucial Implications:

(1) The counterfactual cognizability thesis holds even if no rational human animals actually exist.

(2) If anything is such that rational human animals are unable to cognize it directly, at least to some extent—e.g., things-in-themselves—then that thing does *not* belong to the manifest natural world.

Having stated WCTI as carefully as I can, there are at least two significant philosophical questions that can still be raised about it.

The first question is the *historical* philosophical question of whether Kant's own Transcendental Idealism should be understood as STI or instead as WCTI. My own view on this question, for what it is worth, is that Kant himself simply *oscillated between* STI on the one hand and WCTI on the other hand. Some Kant-texts support one reading, and other Kant-texts support the other reading. The Transcendental Aesthetic and the Analytic of Concepts in the first *Critique* mostly support the STI reading. But Kant's remarks about "empirical realism," the Refutation of Idealism, and the Analytic of Principles more generally (especially the Postulates of Empirical Thought), mostly support the WCTI reading.

The second question—and for me, the massively more important of the two questions—is the *objective* philosophical question of whether either STI or WCTI is in fact true, or whether both are in fact false. My own view on this question, again for what it is worth, is that STI is plausibly arguably *false*, whereas WCTI is plausibly arguably *true*. And here are my basic reasons for holding that STI is false, and that WCTI is true.

On the one hand, I think that it is clearly false that if all actual human minds including mine went out of existence, then the manifest natural world would necessarily go out of existence too. I think that it is clearly false that, e.g., the actual existence of Pike's Peak (a 14,000 foot mountain near Colorado Springs, CO, USA, with a cog

railway that runs right to the summit[5]) strictly depends on the actual existence of human minds including mine. Clearly, I think, Pike's Peak *can* exist even if we, including myself, do not, and in fact I also think that Pike's Peak *actually existed millions of years before any conscious minds of any kind existed*, including of course the conscious minds of all rational human animals, including mine. In this way a great many things, including mountains like Pike's Peak, exist *objectively*—as do shoes, ships, sealing wax, cabbages, kings, seas that do not boil, and pigs without wings. They are, all of them, *neither* subjective (strictly dependent on individual minds) *nor* relative (strictly dependent on cultures or societies). They are all *moderately mind-independent*. So STI is clearly false.

But on the *other* hand, I do also think that it is clearly true that necessarily, if the manifest natural world were not directly cognizable by conscious rational animals like us, at least to some extent, then the manifest natural world would not exist. The manifest natural world, insofar as it now actually exists in its moderately mind-independent way, could not be such that *it is inherently impossible to cognize it directly*; and the manifest natural world, insofar as it now actually exists in its moderately mind-independent way, could not be such that *its actual existence renders our conscious rational animal actual existence impossible*. How could that be the case, given the actual fact that the manifest natural world actually exists now in its moderately mind-independent state, given the other actual fact that we ourselves do actually exist now as rational human animals in the manifest natural world, and given the further actual fact that we do now directly perceive, and thus essentially non-conceptually cognize, some parts of the actual manifest natural world, e.g., our own living animal bodies in actual space and actual time?[6] Therefore, necessarily, the actual existence of the manifest natural world does not render our

conscious rational human animal actual existence in that world impossible. On the contrary, the actual existence of the manifest natural world *renders our conscious rational human animal actual existence in that world necessarily possible*. Here, and now more explicitly, I am arguing in the following way, by using one empirical premise and two modal principles, in addition to the familiar classical logical principle of Existential Generalization:

> Empirical premise: I, a rational human animal, actually exist in the manifest natural world.
> Modal principle 1: Actually P → Possibly P
> Modal principle 2: Possibly P → Necessarily Possibly P (i.e., the characteristic modal axiom of S5).

> (1) I, R.H., a rational human animal, actually exist in the actual manifest natural world. (Empirical premise.)

> (2) Some rational human animals actually exist in the actual manifest natural world. (From (1), by Existential Generalization.)

> (3) Therefore, given the actual existence of the manifest natural world, some rational human animals actually exist in that world. (From (2).)

> (4) Whatever is is actual is also possible. (Premise, from Modal principle 1.)

> (5) Therefore, given the actual existence of the manifest natural world, it is possible that some rational human animals actually exist in that world. (From (3) and (4).)

> (6) If anything is possible, then it is necessarily possible. (Premise, from Modal principle 2.)

> (7) Therefore, given the actual existence of the manifest natural world, it is necessarily possible that some rational human animals actually exist in that world. (From (5) and (6).) QED

This argument is sound whether, on the one hand, the modalities are analytic, conceptual, or logical (a.k.a. weakly metaphysically a priori necessary or possible), or on the other hand, they are synthetic a priori, essentially non-conceptual, a priori essentialist, or non-logical (a.k.a. strongly metaphysically a priori necessary or possible). For these reasons, then, it is plausibly arguable that STI is false and that WCTI is true.

In any case, whether one accepts STI or WCTI, Transcendental Idealism is sharply distinct from Berkeley's *Subjective Idealism*, which says:

(i) ~ ◊ (□x) (x is material and external), i.e., the material external world is impossible.

and

(ii) ~ (□x) (x is either an idea in a conscious mind or x is a conscious mind), i.e., necessarily, anything which exists is either an idea contained in a conscious mind or a conscious mind.

and also from Cartesian *Skeptical Idealism*, which says:

(iii) ◊ ~ (□x) (x exists outside my own conscious states), i.e., possibly nothing exists outside my own conscious states.

In sharp contrast to Berkeleyan Subjective Idealism,

(i) As a mind-dependence thesis, Transcendental Idealism does *not* apply to *all* objects whatsoever—on the contrary, as a mind-dependence thesis, Transcendental Idealism applies *only* to appearances or phenomena, and *never* to things-in-themselves or noumena.

(ii) Transcendental Idealism does *not* say that the material external world is impossible—on the contrary, it holds that necessarily, if I am conscious of my own mental states in time, *then the material external world also exists in space*.

and

(iii) Transcendental Idealism does *not* say that all the proper objects of all human cognition are *nothing but ideas* (i.e., objects existing merely in "inner sense")—on the contrary, it holds that *all* material external objects in space are *also* proper objects of human cognition (i.e., objects existing also in "outer sense").

And in equally sharp contrast to Cartesian Skeptical Idealism, Transcendental Idealism does *not* say that it is possible that nothing exists outside my own conscious states (i.e., inner sense): on the contrary, Transcendental Idealism synthetically a priori necessarily entails *that necessarily some directly knowable material things actually exists outside my conscious states (i.e., inner sense) in space*, i.e., it synthetically a priori necessarily entails the *falsity* of both Berkeleyan Subjective Idealism and Cartesian Skeptical Idealism alike, and also the *truth* of Empirical Realism:

[The] empirical realist grants to matter, as appearance, a reality which need not be inferred, but is immediately perceived (*unmittelbar wahrgenommen*). (*CPR* A371)

Every outer perception … immediately proves (*beweiset unmittelbar*) something real in space, or rather [what is represented through outer perception] is itself the real; to that extent, empirical realism is beyond doubt, i.e., to our outer intuitions there corresponds something real in space. (*CPR* A375)

And this Empirical Realism is in fact the explicit two-part conclusion of Kant's well-known Refutation of Idealism:

[I]t always remains a scandal of philosophy and universal human reason that the existence of things outside us (from which after all get the whole matter for our cognitions, even for our inner sense) should have to be assumed on [the basis of] faith (*auf Glauben*), and that if it occurs to anyone to doubt it, we should be unable to answer him with a satisfactory proof. (*CPR* Bxxxix n.)

[T]he consciousness of my existence is at the same time (*zugleich*) an immediate consciousness of the existence of other things outside me. (*CPR* B276)

II.2 Transcendental Idealism/Empirical Realism and Solipsism/Pure Realism

Wittgenstein's *Tractatus Logico-Philosophicus* was so-named in latinate English by G.E. Moore, who was clearly punning on the title of Spinoza's *Tractatus Theologico-Politicus*,[7] and also equally clearly alluding to the surprisingly Spinozistic proposition 6.45 of the *Tractatus*:

The intuition (*Anschauung*) of the world sub specie aeterni is its intuition as a limited whole. The feeling of the world as a limited whole is the mystical. (*TLP* 6.45, 187)

But the actual German title of the *Tractatus* is "*Logisch-philosophische Abhandlung*," i.e., a "logico-philosophical treatise," i.e., a *treatise in philosophical logic*.

Despite its intentionally and ironically anodyne title, the *Tractatus* offers a radically new conception of philosophical logic, according to which:

(1) Not only mathematics but also metaphysics reduces to the propositions of logic (including both the truth-functional tautologies and the logico-philosophical truths of the *Tractatus* itself) together with factual propositions.

(2) Factual propositions and facts alike reduce to logically-structured complexes of ontologically neutral "objects," which can variously play the structural roles of both particulars and universals (including both properties and relations).

(3) Factual propositions are nothing but linguistic facts that "picture" other facts according to one-to-one isomorphic correspondence relations.

(4) All non-factual propositions are either (a) "senseless" (*sinnlos*) truth-functional tautologies expressing nothing but the formal meanings and deductive implications of the logical constants, (b) the logico-philosophical propositions of the *Tractatus* itself, or (c) "nonsensical" (*unsinnig*) pseudo-propositions that violate logico-syntactic rules and logico-semantic categories, especially including all the synthetic a priori claims of traditional metaphysics.

(5) The logical constants do not represent facts or refer to objects of any sort (*TLP* 4.0312, 69) but instead merely "display" (*darstellen*) the a priori logical "scaffolding of the world" (*TLP* 6.124, 165), which is also "the limits of my language" (*TLP* 5.6, 149), and can only be "shown" (*zeigen*) or non-propositionally indicated, *not* "said" (*sagen*) or propositionally described.

(6) The logical form of the world is therefore "transcendental":

6.13 Logic is not a theory but a reflexion of the world. Logic is transcendental. (*TLP* 169)

and finally

(7) The logical form of the world reduces to the language-using metaphysical subject or ego, who is not in any way part of the world but in fact solipsistically identical to the form of the world itself.

Looking at theses (5), (6), and (7), we can clearly see that Wittgenstein's radically new conception of philosophical logic is correspondingly radically ontologically ascetic, since everything logically reduces to one simple thing: *the language-using metaphysical subject or ego*. Indeed, it is by means of theses (5) and (6) that Wittgenstein directly expresses the surprising and often-overlooked but quite indisputable fact that the *Tractatus* is every bit as much a *neo-Kantian* idealistic metaphysical treatise directly inspired by Schopenhauer's *World as Will and Representation* (1819/1844/1859),[8] and thereby indirectly inspired by Kant's *Critique of Pure Reason*—which, in fact, Wittgenstein carefully read along with Hänsel in 1918, three years before the publication of the *Tractatus*—as it is a logico-philosophical treatise inspired by Frege's *Begriffsschrift*, Frege's attack on psychologism, Moore's and Russell's attacks on

Idealism, and Russell's and Whitehead's *Principia Mathematica*. Whereas Moore and Russell explicitly *abandoned* and *rejected* Kant's epistemology and metaphysics, Wittgenstein in fact *reformulated* and *sublimated* them. And from this standpoint, we can see that the *Tractatus* is fundamentally an essay in *transcendental logic in the Kantian sense*. I will have much more to say about transcendental logic, and the thesis that logic is transcendental, in section IV.

As Wittgenstein stresses in the Preface of the *Tractatus*, he "makes no claim to novelty in points of detail" and does not care whether he is borrowing ideas from other philosophers, especially Frege and Russell. It is also very clear from the *Notebooks 1914-1916*, however, that Wittgenstein was heavily influenced by Schopenhauer. Indeed Wittgenstein told von Wright that "he had read Schopenhauer's *Die Welt als Wille und Vorstellung* in his youth and that his first philosophy was a Schopenhauerian epistemological idealism." And in 1931 Wittgenstein wrote that "Boltzmann, Hertz, Schopenhauer, Frege, Russell, Kraus, Loos, Weininger, Spengler, [and] Sraffa have influenced me" (*CV*: 19[e], underlining added).

Now von Wright also says that

> I know nothing about how this interest [in Schopenhauerian epistemological idealism] was related to [Wittgenstein's] interest in logic and the philosophy of mathematics, except that I remember his saying that it was Frege's conceptual realism which made him abandon his earlier idealistic views.[9]

Given that von Wright by, his own admission, knows "nothing about" how Wittgenstein's idealism is related to his interest in logic and the philosophy of mathematics, then is von Wright likely to be correct, in remembering Wittgenstein saying something about Frege's conceptual realism, that in the *Tractatus* Wittgenstein actually *abandons* his earlier "Schopenhauerian epistemological idealism" by reading and

thinking about Frege's conceptual realism? Or is it rather the case that von Wright simply knows nothing about that too? My proposal is that on this particular point von Wright is, well, simply *von Wrong*; that Frege's conceptual realism motivated early Wittgenstein's anti-*psychologism*, not his supposed anti-*idealism*; and that in fact in the *Tractatus* early Wittgenstein *reformulates* and *sublimates* his earlier neo-Kantian, Schopenhauerian epistemological idealism by, *first*, reading and then thinking long and hard about Frege's, Moore's, and Russell's ideas *prior to 1914*, and also, *second*, by reading and then thinking long and hard about Kant's ideas *after 1918*.

More precisely, if I am correct, then:

(1) In the *Tractatus* Wittgenstein *accepts* the basic framework of Kant's Transcendental Idealism/Empirical Realism and theory of cognition, and in particular Wittgenstein *accepts* a version of Strong Transcendental Idealism or STI, but *rejects* Kant's "modal dualism" of analytic synthetic necessary truths and opts for a "modal monism" of logically necessary truths.

(2) In the *Tractatus* Wittgenstein *accepts* Schopenhauer's reduction of both the metaphysical subject and the metaphysical object (or "thing in itself") of Kant's Transcendental Idealism, to the will.

(3) In the *Tractatus* Wittgenstein *accepts* the basic project of logical analysis as implicit in Fregean logicism, but *rejects* Frege's fundamental appeal to set theory.

(4) In the *Tractatus* Wittgenstein *accepts* the Frege-Russell idea that logic is first philosophy, but *rejects* both of their conceptions of logic: for Wittgenstein, logic is neither the science of laws of truth nor the absolutely general science of deduction; instead, for Wittgenstein, *logic is transcendental* in the Kantian sense.

(5) In the *Tractatus* Wittgenstein *accepts* Frege's semantics of sense (*Sinn*) and Meaning or reference (*Bedeutung*)—that is, he accepts what von Wright calls "Frege's conceptual realism," and thereby rejects *logical psychologism*—but also *rejects* Frege's platonist ontology of the "third realm" and also *rejects* Russell's one-factor or "'Fido'-Fido" theory, except for names.

And finally,

(6) In the *Tractatus* Wittgenstein *accepts* Russell's distinction between knowledge by description and knowledge by acquaintance, and Russell's theory of descriptions, and also Russell's radically simple correspondence theory of truth, but he *rejects* Russell's multiple-relation theory of judgment.

One of the initially most puzzling features of the *Tractatus* is its background metaphysics of *solipsism* and *pure realism* (*TLP* 5.64, 153). In my opinion, Wittgenstein's "solipsism" and "pure realism" are most correctly understood as reformulations and sublimations of Kant's Transcendental Idealism, in the specific sense of *Strong Transcendental Idealism* or STI, and also his "Empirical Realism," in the *Critique of Pure Reason*. This is clearly demonstrated, I think, by the second epigraph of this section—

> This is the way I have travelled: Idealism singles men out from the world as unique, solipsism singles me alone out, and at last I see that I too belong with the rest of the world, and so on the one side *nothing* is left over, and on the other side, as unique, *the world*. In this way idealism leads to realism if it is strictly thought out. (*NB*: 85ᵉ)

—and by these other texts from the *Notebooks 1914-1916* and the *Tractatus*:

> What has history to do with me? Mine is the first and only world.
>
> I want to report how *I* found the world.
>
> What others in the world have told me about the world is a very small and incidental part of my experience of the world.
>
> *I* have to judge the world, to measure things.
> The philosophical I is not the human being, not the human body or the human soul with the psychological properties, but the metaphysical subject, the boundary (not a part) of the world.
>
> The human body, however, my body in particular, is a part of the world among others, among beasts, plants, stones, etc., etc. (*NB*: 82e)
>
>   7.10.16
> And in this sense I can also speak of a will that is common to the whole world.
> But this will is in higher sense *my* will.
>
> As my representation is in the world, in the same way my will is the world-will. (*NB*: 85e)
>
> 5.62 This remark [= "What we cannot think, that we cannot think: we cannot therefore *say* what we cannot think."] provides a key to the question, to what extent solipsism is a truth.
>
> In fact what solipsism *means,* is quite correct, only it cannot be *said*, but it shows itself.
>
> That the world is *my* world shows itself in the fact that the limits of language (*the* language, which I understand) means the limits of *my* world.
>
> 5.621 The world and life are one.
>
> 5.63 I am my world. (The microcosm.)

5.631 The thinking, presenting subject: there is no such thing.

If I wrote a book, *The world as I found it*, I should also have to include a report on my body, and report which parts were subordinate to my will, and which were not, etc., this being a method of isolating the subject, or rather of showing in an important sense there is no subject; for it alone could *not* be mentioned in the book.--

5.632 The subject does not belong to the world but it is a limit of the world.

6.633 Where *in* the world is a metaphysical subject to be found?

You say that this case is altogether like that of the eye and the field of sight. But you do *not* really see the eye. And from nothing *in the field of sight* can it be concluded that it is seen from an eye.

5.634 This is connected with the fact that no part of our experience is also a priori.

Everything we see could also be otherwise.

Everything we can describe at all could also be otherwise.

There is no order of things a priori.

5.64 Here we see that solipsism strictly carried out coincides with pure realism. The I in solipsism shrinks to an extensionless point and there remains the reality co-ordinated with it.

There is therefore really a sense in which in philosophy we can talk of a non-psychological I.

The I occurs in philosophy through the fact that "the world is my world."

The philosophical I is not the man, not the human body or the human soul of which psychology treats, but the metaphysical subject, the limit--not a part of the world. (*TLP* 151-152)

As these texts show, solipsism as Wittgenstein understands it is essentially a creative reformulation and sublimation of Kant's STI, which as we have already seen, says:

(1) Things-in-themselves (a.k.a . "positive noumena," or Really Real things, i.e., things as they could exist in a "lonely" way, altogether independently of rational human minds or anything else, by virtue of their intrinsic non-relational properties) really exist and cause our perceptions, although rational human cognizers only ever perceive mere appearances or subjective phenomena.

(2) Rational human cognizers actually impose the non-empirical structures of their innate cognitive capacities onto the manifest natural world they cognize, i.e., necessarily, all the immanent forms or structures of the proper objects of human cognition are literally *type-identical to* the forms or structures that are non-empirically generated by our innately-specified spontaneous cognitive capacities.

and

(3) Necessarily, if all rational human cognizers went out of existence, then so would the manifest natural world they cognize.

What plays the specific ontological and epistemic roles of things-in-themselves or positive noumena in STI are Wittgenstein's *objects*:

2.014 Objects contain the possibility of all states of affairs.

2.0141 The possibility of its occurrence in atomic facts is the form of the object.

2.02 The object is simple.

2.0201 Every statement about complexes can be analyzed into a statement about their constituent parts, and into those propositions which completely describe the complexes.

2.021 Objects make up the substance of the world. Therefore they cannot be compound.

2.0211 If the world had no substance, then whether a proposition had a sense would depend on whether another proposition was true.

2.0212 It would then be impossible to form a picture of the world (true or false).

2.022 It is clear that however different from the real world an imagined one may be, it must have something—a form—in common with the real world.

2.023 This fixed form consists of the objects.

2.0231 The substance of the world *can* only determine a form and not any material properties. For these are first presented by the propositions—first formed by the configuration of the objects.

2.0232 Roughly speaking: objects are colorless. (*TLP* 35)

Correspondingly, what plays the specific ontological and epistemic roles of empirically real appearances, objective real phenomena, or objects of experience in STI are Wittgenstein's *atomic facts*:

1 The world is everything that is the case.

1.1 The world is the totality of facts, not of things.

1.11 The world is determined by the facts, and by these being *all* the facts.

1.12 For the totality of facts determines both what is the case, and also what is not the case.

1.13 The facts in logical space are the world.

1.2 The world divides into facts.

1.21 Any one can either be the case or not be the case, and everything else remain the same.

2 What is the case, the fact, is the existence of atomic facts.

2.01 An atomic fact is a combination of objects (entities, things).

2.011 It is essential to a thing that it can be a constituent part of an atomic fact. (*TLP* 31)

So in these ways, according to Wittgenstein, *I* am my world (*TLP* 5.63, 151) and the world is *my* world (*TLP* 5.641, 153), the subject does not belong to the world but is a limit of the world (*TLP* 5.631-5.632, 151), and the metaphysical subject is a non-psychological ego (*TLP* 5.633 and 5.641, 151 and 153). In other words, according to Wittgenstein, a specifically Schopenhauerian, creatively reformulated, and sublimated version of Kantian STI is true.

What basic reasons does Wittgenstein have for holding this specifically Schopenhauerian, creatively reformulated, and sublimated version of STI? The answer is that they follow directly from

(1) the Tractarian thesis that the world of facts is constructed *by the language-using subject,*

together with

(2) the Tractarian thesis that the positively noumenal objects, or Wittgensteinian things-in-themselves, are given as an independent constraint *on language and thought*,

together with

(3) the Tractarian thesis *that language is fundamentally a language of thought.*

In short, Wittgenstein's specifically Schopenhauerian, creatively reformulated, and sublimated version of Kantian Strong Transcendental Idealism, or STI, in the *Tractatus* is *Linguistic* Strong Transcendental Idealism, or LSTI for short:

5.6 *The limits of my language* mean the limits of my world. (*TLP* 149)

According to Wittgenstein, the totality of propositions is the language as a whole (*TLP* 4.01, 63). Among other things, this shows us that in the *Tractatus* he is treating language only insofar as it is an information-carrying medium, not insofar as it is a means

of communication, self-expression, or social action. The totality of true propositions is complete natural science (*TLP* 4.11, 75). This in turn shows us that Wittgenstein is treating positive atomic facts, or the truth-makers of atomic propositions, as ultimately reducible to the natural facts (ontological naturalism). Language, however, for Wittgenstein is not *merely* the set of public inscriptions or utterances or texts that pick out the natural facts: it also includes any proposition-constructing activity, internal or external. This Wittgenstein calls "thought" or "thinking" (*TLP* 3.1-3.11, 45, and 3.5-4.002, 61-62). So all thinking is also essentially linguistic and propositional in character. This has two important consequences.

First, all thinking, whether or not accompanied by utterance, occurs in a private language of thought. Thinking is inner propositional activity.

Second, natural language and cognition are both essentially propositional and thought-based in character, even though they may not *appear* to be such. The surface structure of inner or outer natural language (i.e., its surface grammar or psychological syntax) thoroughly disguises its real structure (i.e., its depth grammar or logical syntax) (*TLP* 4.002, 62). Only logical analysis can reveal this underlying structure. But, on the other hand, this logical analysis should not be regarded as a *reform* of language (prescriptive grammar, prescriptive syntax); on the contrary, everything in natural language is logically perfectly in order, just as it is (*TLP* 5.5563, 149). In this respect, early Wittgenstein's approach to logical analysis is sharply different from that of the early Logicists (e.g., Frege and Russell) and the Logical Empiricists (e.g., Carnap and the Vienna Circle), who were explicitly rational reformers of natural language. On the contrary, for Wittgenstein in the *Tractatus*, logical analysis is there merely to clarify what

we already *implicitly* fully understand. I will come back to these points when I discuss

transcendental philosophy's critique of self-alienated philosophy—*including* Frege's and

Russell's Logicism, Carnap's and the Vienna Circle's Logical Empiricism, and Quine's,

Reichenbach's, and Sellars's Scientific Naturalism—in section V.

Strikingly, Wittgenstein's solipsism, i.e., his Linguistic Strong Transcendental

Idealism, or LSTI, has two importantly distinct although fully complementary

dimensions:

(i) a solipsism/LSTI of the *representing* subject,

and

(ii) a solipsism/LSTI of the *willing* subject:

6.373 The world is independent of my will. (*TLP* 181)

6.43 If good or bad willing changes the world, it can only change the limits of the world, not the facts; not the things that can be expressed in language.

In brief, the world must thereby become quite another. It must wax or wane as a whole.

The world of the happy man is a different one from that of the unhappy man.

6.431 So too at death the world does not alter, but comes to an end.

6.4311 Death is not an event in life: we do not live to experience death.

If we take eternity to mean not infinite temporal duration but timelessness, then eternal life belongs to those who live in the present.

Our life has no limit in just the way in which our visual field has no limits. (*TLP* 185)

Wittgenstein's solipsism/LSTI of the *representing* subject says that all worldly

facts are metaphysically dependent on my mind in the double sense that linguistic form

(and its a priori essence, logical form) enters directly into the constitution of every fact,

and that language itself is constructed by the individual subject.

But Wittgenstein's solipsism/LSTI of the *willing* subject, by sharp contrast, says

that the specific internal nature of the objects is metaphysically dependent on my

attitudes, desires and volitions (i.e., my willing). The world of facts is independent of my will, but the *form* and *limits* of the world, i.e., the *global structure* of the world, which is partially constituted by the specific internal nature of the objects, is dependent on my will. Now the world and my life are the same thing. Thus the world can "wax or wane" as a whole, depending on my acts of willing, just as all the events of my life depend on my will. They do not however depend on my will in the sense that I can actually change any facts—I cannot—but in the sense that I can control the personal meaning or *value* of those facts, which is bound up essentially with the world's global stucture. So my will determines how I value the world and my life, which in turn partially determines the "substance" of the world by partially determining the nature of the objects, and thereby partially determining the global structure of the world. In this way, the world of the happy person, e.g., is metaphysically distinct from the world of the unhappy person. Here we can see that although the constitution of the facts is dual (with language on the one side, and the objects on the other) the metaphysical subject ultimately grounds both of the dual inputs by acting both as the language-user and also as the partial determiner of the specific character of the objects and of the world's global structure, i.e., of its *transcendental* structure. In that sense, whether I live in "the world of the happy," or not, *is solely up to me, and something for which I am alone fully responsible,* no matter what the natural facts may be. I will come back to the will's volitional independence from the facts when I discuss Kant's moral theology and early Wittgenstein's ethical mysticism in section III.

Now back to the dependence of the world on the individual representing subject. Wittgenstein wants to argue that his solipsism/LSTI, when properly understood, is in fact

a "pure realism." In order to make sense of this, we must remember is that the classical philosophical thesis of Realism comes in two very different versions:

> (i) **Noumenal Realism**: Things in the world have an essentially *mind-independent* existence and nature—i.e., they are things-in-themselves—hence they are *directly unknowable* by means of human cognition, and are at best *indirectly knowable* by means of sensory ideas, concepts, inferences, and theories,

> (ii) **Empirical Realism**: Things in the world have an essentially *mind-dependent* existence and nature, and are *directly knowable* by means of human cognition, and also *indirectly knowable* by means of concepts, inferences, and theories .

Classical Rationalists and classical Empiricists hold (i)—with sharply different degrees of epistemic confidence about the indirect knowability of things-in-themselves, to be sure— and reject (ii). By sharp contrast, Kant and Wittgenstein hold (ii) and firmly reject any version of (i), whether classical Rationalist or classical Empiricist, especially including Berkeley's subjective idealism, as a radical version of classical Empiricism according to which either direct or indirect knowledge of external or material things-in-themselves is impossible, although direct and indirect knowledge of minds or spirits is possible, indeed actual, insofar as we know our own minds directly and God's mind indirectly. Most importantly, in any case, Kant and Wittgenstein hold that *in order to be an Empirical Realist/pure realist, one must also be a Transcendental Idealist/solipsist, but not a subjective idealist.*

More precisely, Wittgenstein's pure realism is that *nothing mediates between our correct use of language and the facts we thereby know:* we cognize facts directly through the correct use of complete propositional symbols, and we cognize objects directly through the correct use of names. Then, provided that our judgments are true, we know the facts directly.

This does not, however, in and of itself tell us how solipsism (i.e., LSTI) *leads to* pure realism (i.e., Empirical Realism). Here Wittgenstein wants to say that his solipsism/LSTI is not a solipsism/LSTI of the *psychologically* individual subject, who is individuated by her body and her own personal history, but rather a solipsism/LSTI of the individual subject *considered as an anonymous or generic represener and language-user*. This anonymous or generic subject is an "extensionless point" precisely because s/he functions *only* as the means of representing the world. Here Wittgenstein uses the striking analogy of the visual field and the eye: the seeing eye is the necessary vehicle or means of vision, but it is not itself part of the visual field or its contents; rather the seeing eye is *presupposed by* the visual field and its contents. Similarly, the world contains all the facts, including the facts about my psychologically individual subject; but when *all* of these facts have been recorded, there is still something left over, namely, the anonymous or generic representing language-using subject *as such*, which is contentless, yet presupposed by all the facts. Then when we consider the world of facts from the standpoint of that contentless anonymous or generic representing language-using subject as such, we recognize that this entire world (my world, my life, the totality of facts) is directly presented to me and also fully knowable by me just insofar as I linguistically represent it. In section *V*, I will argue that although later Wittgenstein in the *Investigations* specifically rejects the *solipsism*, or individualism, of LSTI, as mistakenly Cartesian, he never rejects and in fact permanently continues to hold the *Transcendental Idealism* of LSTI, only now in a *communitarian* or social-practical version.[10] Otherwise put, although the later Wittgenstein purges himself of his earlier *Cartesianism*, he never rejects and in fact permanently continues to adhere to his earlier *Kantianism*.

III. If God's Existence is Unprovable, Then is Everything Permitted? Denying *Wissen* and Making Room for *Glauben*

> The famous ontological (Cartesian) proof of the existence of the highest being from concepts is only so much trouble and labor lost, and a human being can no more become richer in insight from mere ideas than a merchant could in resources if he wanted to improve his financial state by adding a few zeroes to his cash balance. (*CPR* A602/B630)

> A *postulate* of pure practical reason … [is] a *theoretical* proposition, though not one provable as such, insofar as it is attached inseparably to an a priori unconditionally valid *practical* law. (*CPrR* 5: 122)

> There are three existence-spheres: the esthetic, the ethical, the religious…. The ethical sphere is only a transition-sphere, and therefore its highest expression is repentance as a negative action. The esthetic sphere is the sphere of immediacy, the ethical the sphere of requirement (and this requirement is so infinite that the individual always goes bankrupt), the religious the sphere of fulfillment, but, please note, not a fulfillment such as when one fills an alms box or a sack with gold, for repentance has specifically created a boundless space, and as a consequence the religious contradiction: simultaneously to be out on 70, 000 fathoms of water and yet be joyful.

> --S. Kierkegaard[11]

> "Take that money away with you, sir," Smerdyakov said with a sigh.
> "Of course, I'll take it! But why are you giving it to me if you committed a murder to get it?" Ivan asked, looking at him with intense surprise.
> "I don't want it at all," Smerdyakov said in a shaking voice, with a wave of the hand. "I did have an idea of starting a new life in Moscow, but that was just a dream, sir, and mostly because 'everything is permitted'. This you did teach me, sir, for you talked to me a lot about such things: for if there's no everlasting God, there's no such thing as virtue, and there's no need of it at all. Yes, sir, you were right about that. That's the way I reasoned."

> --F. Dostoevsky[12]

> To believe in God means to understand the problem about the meaning of life. To believe in God means to see that the facts of the world are not the end of the matter. To believe in God means to see that life has a meaning. (*NB*: 74^e)

> Life is like a path along a mountain ridge; to right and left are slippery slopes down which you slide without being able to stop yourself, in one direction or the other. I keep seeing people slip like this and I say "How could a man help himself in such a situation!". And *that* is what "denying free will" comes to. That is the attitude expressed in this 'belief' (*'Glauben'*). But is it not a *scientific* (*wissenschaftlichen*) belief and has nothing to do with scientific convictions. (*CV*: 63^e)

III.1 Transcendental Existentialism

In my opinion, Kant's moral theology and early Wittgenstein's ethical mysticism

in the *Tractatus* are both investigations in a special kind of transcendental *Existential*

philosophy, which I will call *Transcendental Existentialism* for short. The word

'existential' in this descriptive label is intentionally and systematically ambiguous. This is because I want its meaning to comprehend both

> (i) the notion of "existential" in the specifically *logico-semantic* and *formal-ontological* sense of existential predication, particular quantification, or direct reference,

and also

> (ii) the notion of "existential" in the specifically *ethical, metaphysical, and religious* sense of Existentialism, which is concerned with our anxious search for a coherent, meaningful, and morally good life in an otherwise absurd, meaningless, and amoral world without a God, either because God has apparently withdrawn from Her Creation (theistic Existentialism—e.g., Augustine, Pascal, Spinoza, or Kierkegaard) or because God apparently does not exist (atheistic or humanistic Existentialism[13]—e.g., Nietzsche, Camus, or Sartre).

More precisely, in this section I want to show how seven deeply important Kantian and early Wittgensteinian ideas, which collectively constitute Transcendental Existentialism, can significantly illuminate some central issues in philosophical logic, semantics, formal ontology, ethics, the metaphysics of free will, action theory, morality, and philosophical theology. These deeply important Kantian and early Wittgensteinian ideas are:

> (1) Kant's argument for the impossibility of The Ontological Argument on logico-semantic grounds alone, which entails both the logical unprovability and scientific unknowability of God's existence and God's non-existence alike. This is Kant's *radical agnosticism*.

> (2) Kant's first "postulate of pure practical reason," *immortality*, which basically says that even though we cannot either logically prove or scientifically know either that our souls are immortal or that they are not immortal, we must nevertheless have "faith," *Glaube*, or moral certainty in the rational Idea that after our deaths we will have an endless human personal existence in a world that is wholly known and governed by God, and in which eventually all the morally virtuous people are made happy, and all the wicked people are punished *not for retributive reasons* but instead precisely so that they can finally face up to their sins, take full responsibility for them, and then change their lives for the better. Notice especially that there is no such thing as *capital punishment* in the God-governed world of immortal rational human animals: on the contrary, in the Kantian after-life, everyone gets *an infinite number of "second chances" to redeem themselves*.

> (3) Kant's third postulate of pure practical reason, *the existence of God*, which basically says that even though we cannot logically prove or scientifically know either that God exists or that God does not exist, we must nevertheless have "faith," *Glaube*, or moral certainty in the rational Idea that that God exists in order to unify happiness and virtue in a desperately nonideal world filled to the brim with "the crooked timber of humanity"

and "radical evil," and in which, it seems, nothing will ever be made straight, and "no good deed goes unpunished."

(4) Early Wittgenstein's thesis that to believe in God means to see "the facts of the world are not the end of the matter" and "that life has a meaning."

(5) Kant's second postulate of pure practical reason, *freedom*, which basically says that *because* we cannot logically prove or scientifically know either that God exists or that God does not exist, we must therefore have "faith," *Glaube*, or moral certainty in the rational Idea that we are both transcendentally free and also practically free in order to rule out

> (i) the self-stultifying threat to our transcendentally free intentional agency of a world in which, it seems, on the *one* hand, that the past is completely filled with deterministic and impersonal causes (the rock) and, on the *other* hand, that the future is randomly indeterministic and heart-breakingly completely filled with possibilities for bad luck (the hard place),

and also

> (ii) the self-stultifying threat to our practically free moral agency of a world in which, it seems, because God's existence is unprovable, then moral chaos reigns and "everything is permitted."

(6) Wittgenstein's essentially parallel and similar thesis that the *denial* of free will is also a special kind of "faith" or *Glaube*, and not a scientific belief.

and finally

> (7) Our fundamental value-commitments to (i) the non-denumerably absolutely intrinsically valuable dignity of rational animals or finite persons, which inheres in our *capacity* to be autonomous, and also to (ii) its perfection, "the highest good" or "the supreme good," which is a *good will* (*GMM* 4:393) (*CPrR* 5: 110), and also to (iii) its full realization, "the sole and complete good" (*GMM* 4:396) or "the whole, complete good" (*CPrR* 5: 111), which is *the coherent fusion of a good will and morally virtuous happiness*, would be impossible without our wholehearted affirmation of the three postulates of pure practical reason. But, in turn, this wholehearted affirmation is impossible without radical agosticism. In short, our rational human lives can be authentic and meaningful *only if* we accept radical agnosticism.

III.2 Kant's Moral Theology and The Incoherence Problem

Kant's moral theology is notoriously difficult to understand. This is principally

due to an apparent inconsistency between the four proper parts of his theory.

Part 1. First, Kant works out a devastating logical, semantic, and epistemological

critique of any possible proof for God's existence, including The Ontological Argument,

The Cosmological Argument, and The Design Argument—a.k.a. "the physico-theological argument" or The Telelogical Argument—which has the immediate further implication that any possible proof for God's *non*-existence is *also* impossible, including The Argument from Evil, in either its classical "metaphysical" version or its more modern "evidential" version. More precisely, Kant argues that God's existence or non-existence is not only *unknowable* but also *uncognizable*, although at the same time God's existence remains *thinkable*.

Now for Kant, "scientific knowing" or *Wissen* is the same as a true belief that P which is sufficiently justified by reasons in both a subjective sense (in which case it is "conviction" or *Überzeugung*) and also in an objective or universally intersubjective sense (in which case it is "certainty" or *Gewissheit*) (*CPR* A822/B850). Apart from sufficient justification by reasons, scientific knowing also has two further substantive necessary conditions, namely

> (i) truth or "objective reality," which is the formal correspondence of a cognition with an actual or real-world object,

and

> (ii) empirical meaningfulness or "objective validity," which is the necessary relatedness of any cognition to direct, non-conceptual sensory acquaintances or encounters with real individual worldly objects, i.e., "empirical intuitions" (*empirischen Anschauungen*).

By sharp contrast to scientific knowing, however, "cognition" or *Erkenntnis* is either

> (1) according to the very broad construal in the 1781 or A edition of the *Critique of Pure Reason*, any object-directed consciousness whatsoever (*CPR* A320/B376),

or else

> (2) according to the quite narrow construal in the 1787 or B edition of the *Critique of Pure Reason*, an empirically meaningful (objectively valid) judgment that P, which is the same as a "judgment of experience" or *Erfahrungsurteil* (*CPR* Bxxvi, B142, B147).

On either the (1) broad or the (2) narrow construal of *Erkenntnis*, however, it is possible for a cognition to be either not objectively valid (i.e., not empirically meaningful) or not objectively real (i.e. false). Hence the notion of cognition is not equivalent with the notion of knowledge, which on the contrary entails both objective validity (i.e., empirical meaningfulness) and objective reality (i.e., truth), in addition to *sufficient justification by reasons*. In any case, objective validity is a necessary and sufficient condition of the truth-valuedness of any belief, judgment, or statement. More specifically, the failure of objective validity for any putative belief, judgment, or statement entails that it is nothing but a mere *thought* which lacks a truth-value altogether—"thoughts without content are empty (*leer*)" (*CPR* A51/B75)—and thereby is a "truth-value gap."[14]

Correspondingly, "thinking" or *Denken* in the 1787 or B edition of the first *Critique* is minimal consistent conceivability, which entails the bare logical possibility of the object which is thereby thinkable, but does *not* guarantee the real or metaphysical possibility of that object, much less its actuality or reality (*CPR* Bxxvi), much less the truth-valuedness of that thought.

Part 2. Second, Kant works out a devastating logical, semantic, and epistemological critique of any possible proof for the existence or non-existence of an immaterial, substantial soul, which has the direct implication that any possible proof for the immortality or non-immortality of the soul is also impossible. More precisely, Kant argues that the immortality or non-immortality of the soul is not only scientifically knowable but also uncognizable, although at the same time the immortality of the soul remains thinkable.

Part 3. Third, Kant argues that the rational or reasons-responsive content of "believing" or "faith" (*Glauben*) in the existence of God and the immortality of the soul, and more generally the rational or reasons-responsive content of theology and religion, is strictly *moral or practical* in character, and not *scientific* or *theoretical* in character:

> I cannot even *assume God, freedom, or immortality* for the sake of the necssary practical use of my reason unless I simultaneously *deprive* speculative reason of its pretensions to extravagant insights; because in order to attain to to such insights, speculative reason would have to help itself to principles that in fact reach only to objects of possible experience, and which, if they were to be applied to what cannot be an object of possible experience, then they would always transform it into an appearance and thus declare all *practical extension* of pure reason to be impossible. Thus I had to deny *scientific knowing* (*Wissen*) in order to make room for *faith* (*Glauben*). (*CPR* Bxxix-xxx).

Part 4. Fourth—and most puzzlingly of all in view of the other parts of his theory—Kant also argues that believing in God's existence and the immortality of the soul are *necessary presuppositions* of morality, in the strong sense that without these rational commitments, not only would morality itself would be empty and pointless, but also my personal commitment to morality would be self-alienating and self-stultifying:

> I will inexorably believe in the existence of God and a future life, and I am sure that nothing can make these beliefs unstable, since my moral principles themselves, which I cannot renounce without becoming contemptible in my own eyes, would thereby be subverted. (*CPR* A828/B856)

Parts 1, 2, and 3 are clearly consistent. The equally clear problem is that Part 4 apparently contradicts Parts 1, 2, and 3: How can believing in God's existence and the immortality of the soul be constitutive presuppositions of morality, on the assumption that morality exhausts the rational content of theology and religion, if all proofs for God's existence and the immortality of the soul are impossible? For convenience, let us call this *The Incoherence Problem*. In order to make any headway at all towards solving The Incoherence Problem, I think that we must especially emphasize and understand three special features of Kant's theory.

The first special feature is the fact that Kant's critique of arguments for God's existence and for the immortality of the soul yield the conclusions that we *scientifically know* that neither God's existence nor God's non-existence can be proved, and also that we *scientifically know* that neither the immortality of the soul nor the non-immortality of the soul can be proved, although at the same time they remain thinkable. In other words, for Kant we *scientifically know* that God's existence or non-existence is unknowable and uncognizable, and we also *scientifically know* that the immortality or non-immortality of the soul is unknowable and uncognizable, although they remain thinkable. Let us call this feature *radical agnosticism*, since it is not ordinary agnosticism or *epistemic neutrality* as between opposing beliefs. On the contrary, it is a special form of *epistemic certainty* with respect to the inherent scientific unknowability and uncognizability alike of *both* members of certain contradictory or contrary belief-pairs, while at the same time accepting the thinkability of both propositions. Radical agnosticism is nothing more and nothing less than the permanent rational suspension of belief in a thinkable proposition (or doctrine) and its negation alike. Or otherwise put, *radical agnosticism is having objective epistemic certainty about that which is objectively epistemically uncertain.*

The second special feature is that for Kant the rational attitude of *believing-in* is *not* the same as the rational attitude of *believing-that.* For example, I can believe *in* global justice, and thereby be volitionally committed to global justice, and indeed be prepared to die for the sake of global justice, even if I also strongly believe *that* contemporary post-industrial capitalist world politics are inherently corrupt and evil, and *that* there is no true justice anywhere on the face of the earth. Conversely, I can believe with a priori rational intuitive certainty that 7+5=12 even if, as Kant very aptly remarked, I would not be

prepared to die for this belief.[15] Let us call this special feature *believing-in-as-volitional-commitment*.

And the third special feature of Kant's account is that for him there is a crucial distinction between

(I) *believing that P* when you have no sufficient epistemic justification for believing that P,

and

(II) *choosing or acting as if, counterfactually, you believe that P*, even though in fact you have no sufficient epistemic justification for believing that P.

Propositional attitude (I) cannot be rational in any sense. It cannot be epistemically rational to believe that P without sufficient epistemic justification for believing that P, nor can it be practically rational to believe that P without sufficient epistemic justification for believing that P. In other words, you cannot have a good practical reason to have an epistemic belief in a proposition you know you have no good epistemic reason to believe.

But by sharp contrast, propositional attitude (II) *can* indeed be fully *practically* rational:

Only in a practical relation…can taking something that is theoretically insufficient to be true be called believing (*Glauben*). (*CPR* A823/B852)

More precisely, what propositional attitude (II) says is that you choose or act in such a way that you *would* act, *were* you to believe it, even though either (i) you *do* not epistemically believe it or else (ii) you *cannot* epistemically believe it. Hence a case of propositional attitude (II) can be fully practically rational if you have a sufficient practical reason for comporting yourself in the same way as you *would* comport yourself, *were* you to epistemically believe a certain proposition that P, even though you scientifically know that this proposition is unknowable and uncognizable, and indeed

even though you scientifically know that the denial of this proposition that P is also

unknowable and uncognizable, although at the same time both the proposition and its

denial remain thinkable. In other words, propositional attitude (II) can be both fully

practically rational and also fully consistent with *radical agnosticism*. You can have a

sufficient practical reason to comport yourself *as if, counterfactually,* you epistemically

believe a proposition that P, even though you *scientifically know* you have no good

epistemic reason to believe that P or disbelieve that P. In this way, propositional attitude

(II) is not a *doxic* propositional attitude (i..e, an epistemic belief), but instead a

commissive propositional attitude (i.e., a practical belief). For example, I can have a

sufficient practical reason for comporting myself *as if, counterfactually,* I epistemically

believe that nearly all people are generous and good-hearted, since that way of

comporting myself keeps me committed to working towards global justice in the face of a

large body of otherwise very disheartening evidence which shows that the purely

decision-theoretic interests of multinational corporations will always trample on and

trump the basic human interests of ordinary people, and even though I have no good

epistemic reason whatsoever for believing or disbelieving that nearly all ordinary people

are generous and good-hearted. When a case of propositional attitude (II) has a

categorically sufficient, or *moral*, practical reason supporting it, then Kant calls *moral*

belief or *moral certainty*:

> [In moral belief] it is absolutely necessary that something must happen, namely, that I fulfill the
> moral law in all points. The end here is inescapably fixed, and according to all my insight there is
> possible only a single condition under which this end is consistent with all ends together and
> thereby has practical validity, namely, that there be a God and a future world; I also know with
> complete certainty that no one else knows of any other conditions that lead to this same unity of
> ends under the moral law…. The conviction is not *logical* but *moral* certainty, and, since it
> depends on subjective grounds (of moral disposition) I must not even say "*It is* morally certain
> that there is a God," etc., but rather "*I am* morally certain" etc. That is, the belief in a God and
> another world is so interwoven with my moral disposition that I am in as little danger of ever

surrendering the former as I am worried that the latter can ever be torn away from me. (*CPR* A828-829/B856-857).

Kant's notion of "moral certainty" plays a very interesting variation on Descartes's notion of "moral certainty" in his *Principles of Philosophy*, about which Descartes says:

> Moral certainty is certainty which is sufficient to regulate our behavior, or which measures up to the certainty we have on matters relating to the conduct of life which we never normally doubt, though we know it is possible, absolutely speaking, that they may be false.[16]

In the *Discourse on Method*, Descartes also explicitly contrasts moral certainty with "metaphysical certainty,"[17] i.e., with what Kant calls *logical* certainty. It is also importantly ironic that in that particular text in the *Discourse*, Descartes is explicitly contrasting the *metaphysical certainty* of his proof for the existence of God and the soul with the merely *moral certainty* of

> everything else of which [people] may think themselves more sure—such as their having a body, there being stars and an earth, and the like.

For Kant, by sharp contrast, there can be *no such thing* as metaphysical or logical certainty about the existence of God and the immortality of the soul; there really can be and is *scientific knowledge* about the existence of one's own body, the earth, the stars, "and the like"; and there really can be, and only *ought to be*, moral certainty about the existence of God and the immortality of the soul. In this sense, Kant can consistently hold that

> (i) It is cognitively impossible either to *believe-that* God exists or to *believe-that* God does not exist.

and also that

> (ii) It is morally obligatory to *believe-in* the rational Ideas of the existence of God and the immortality of the soul.

In this sense, Kant's most philosophically insightful and knowledgeable biographer,

Manfred Kuehn, is strictly speaking mistaken when he says that

> It was clear to anyone who knew Kant personally that he had no faith in a personal God. Having postulated God and immortality, he himself did not believe in either.[18]

Strictly speaking, what Kuehn *should* have written is that

> It was clear to anyone who knew Kant personally that he had no belief-that a personal God either exists or does not exist. Having postulated God and immortality, he himself did not believe-that either, although at the same time he also had overriding moral reasons for believing-in both.

I will come back again later to these fundamental points in sub-sections III.4 and III.5.

III.3 The Unprovability of God's Existence or Non-Existence

Kant's famous or notorious critique of "transcendental theology" (*CPR*

A631/B659) occurs in chapter three of the Transcendental Dialectic, and is called "The

Ideal of Pure Reason" (*CPR* A567-642/B595-670). There he argues for the logical

unprovability of God's existence in four steps by arguing that

(i) There cannot be an ontological proof.

(ii) There cannot be a cosmological proof.

(iii) There cannot be a physico-theological proof (i.e., a sound argument from design, or a sound teleological argument).

and that

(iv) There are only three possible proofs for God's existence.

In fact, Kant's critique of the ontological proof, *on its own*, suffices to show that

God's existence is logically unprovable and scientifically unknowable, since only the

ontological argument even purports to be a logical—or analytic a priori—argument for

God's actual existence. The cosmological proof, if sound, would yield God's actual

existence as a synthetic a priori truth; and the physico-theological proof or

design/teleological argument, if sound, would yield God's actual existence as a synthetic

a posteriori truth. But the negative criterion of the syntheticity of any proposition, whether synthetic a priori or synthetic a posteriori, is that its negation is logically consistent (*CPR* A150-158/B189-197). Therefore, even if the cosmological proof or the physico-theological proof were sound, this would not entail that God exists *in every logically possible world, including the actual world*. In other words, even if these proofs were sound, then logically and analytically speaking, God still might not have existed. But that leaves open an epistemological and ontological gap into which an atheistic skeptic can always introduce a significant doubt. So showing that the ontological proof is impossible suffices to show that God's existence is logically unprovable in the sense required for *epistemic necessity*, which according to Kant is a belief which involves not merely "conviction" (*Überlegung*), thereby having a subjectively sufficient justification, but also involves "certainty" (*Gewissheit*), thereby having an objectively sufficient justification (*CPR* A820-822/B848-850). In other words, showing that the ontological proof of God's existence is impossible also shows that authentic scientific *knowledge* of God's existence is impossible.

The chapter on the Ideal of Pure Reason follows the Paralogisms and the Antinomies, and completes Kant's transcendental logic of illusion, or the Dialectic of Pure Reason. The Dialectic is triadically organized according to three basic types of Idea of Pure Reason:

> (1) the Idea of an absolute subject of cognition, or the Cartesian immaterial soul (the Paralogisms),

> (2) the Idea of an absolute object of cognition, or nature as a cosmological totality (the Antinomies),

and

(3) the Idea of an absolute ground of both the subject and the object of cognition, or God (the Ideal).

The dialectical error in the Paralogisms is the invalid inference from the fact of transcendental apperception or the "I think," to the existence of a simple substantial immortal Cartesian soul; and in the Antinomies the dialectical error is failing to draw the fundamental ontological distinction between appearances or phenomenal entities and things-in-themselves or noumenal entities. In the Ideal, the error is the invalid inference from the fact that every part of the actual or real world is completely determined, to the existence of a single absolutely real being (God) which is the ground of (i.e., is necessary and sufficient for) the complete determination of the actual or real world.

Ideals, according to Kant, are the Ideas of Pure Reason *incarnate* or *reified*: they are individual beings which contain in themselves the completed totality of conditions that is represented by the content of every Idea insofar as it is a third-order "absolutizing" concept or "notion" that applies to the logically fundamental second-order concepts, or pure concepts of the understanding. The concept of God, in turn, depends on the very concept of a "concept." Logico-semantically speaking, a concept is a unified self-consistent inherently general semantic content that functions as a predicate of judgments. For every such concept (e.g., the concept of a cat, or the concept of the cat's being on the mat), given the unity and self-consistency of its semantic content, there is a corresponding logically possible object or logically possible state-of-affairs (e.g.,. a cat, or a cat's being on the mat). For every such concept, there is also a corresponding contradictory concept (e.g., the concept of a non-cat, or the concept of its not being the case that the cat is on the mat). Now consider the total set of all such concepts together with their contradictories: this constitutes our total human conceptual repertoire, or what

Kant calls "the sum total of all possibility" (*CPR* A573/601). From this repertoire, a *logically possible world* can be cognitively constructed as a total set of mutually consistent concepts such that the addition of one more concept to the set would lead to a contradiction. In the jargon of contemporary logic, this is called "maximality." So a logically possible world for Kant is nothing but a maximal consistent set of concepts.

Now consider the set containing every maximal consistent set of concepts. This is the set of all logically possible worlds. A "determination" for Kant is an empirical concept insofar as it is actually applied or at least applicable to an empirical object: in contemporary terms, a determination is a *property* of an object. Now according to Kant, everything that is actual or real must be completely determined. This means that for every actual or real thing, and for every concept of things, either the concept or its contradictory applies to the thing, but not both. Obviously this ontological principle corresponds directly to the classical logical Principle of Non-Contradiction (PNC) for first-order monadic predicate logic:

> For every property and for every thing, it is not the case that any thing both intantiates and also does not instantiate the same property, i.e.,

> ($\Box$P) ($\Box$x) ~ (Px & ~Px).

But the ontological significance of complete determination is that the reality or actuality of a thing expresses a logically complete systematic selection of properties from the totality of possible properties. Otherwise put, every actual or real thing is identical with the total set of mutually consistent concepts that apply to it. This corresponds to Leibniz's idea that every monad or metaphysically real individual has a *complete individual concept* that completely determines its essence. And this in turn corresponds to Leibniz's Laws: *The Identity of Indiscernibles*, which says that necessarily, any two things sharing

all properties in common are identical, and *The Indiscernibility of Identicals*, which says that necessarily, identical things share all their properties in common. Now according to Kant, the concept of God is the concept of a single (positively) noumenal being that is the ground of (i.e., is necessary and sufficient for) the complete determination of the actual or real world. Again, the concept of God is the concept of a single thing-in-itself that contains within its essence all of actuality or reality: hence Kant calls the concept of God the concept of the *ens realissimum* (*CPR* A577/B605).

Given this framework, the fallacy of the Ideal can be construed in two different ways: first, to infer invalidly from the objectively valid thesis of the complete determination of every actual or real thing, to the noumenal concept or Idea of a single "really real" being that completely determines all of actuality or reality (= false reification), or second, to infer invalidly from the concept of the *ens realissimum,* or the concept of the ground of the sum total of all possibilities, to the actual existence of what is described by that concept (= an unsound existence proof, or unsound ontological argument).

As I have mentioned already, The Ontological Argument, or The OA, is the analytic a priori argument from the concept of God to God's actual existence. The original version of The OA is to be found in Anselm's *Proslogion*. But probably the best known modern version of it is to be found in Descartes's fifth *Meditation*. Here are quick glosses of those two arguments:

Anselm's OA

(1) The concept of God is the concept of that-than-which-nothing-more-real-can-be-thought.

(2) That-than-which-nothing-more-real-can-be-thought could not exist merely *inside* the mind (as a concept or idea), for then it would be possible to think of something more real than it: i.e., its existing *outside* the mind.

(3) Therefore that-than-which-nothing-more-real-can-be-thought must not exist merely inside the mind (as a concept or idea). That is, it must also exist outside the mind.

(4) Therefore it is necessarily (i.e., logically, analytically a priori) true that God exists.

Descartes's OA

(1) The concept of God is the concept of a perfect being.

(2) The concept of a perfect being is the concept of a being whose essence contains all perfections.

(3) Existence is a perfection.

(4) Therefore the concept of God is the concept of a being whose essence entails its existence.

(5) Therefore it is necessarily (i.e., logically, analytically a priori) true that God exists.

And here is the substance of Kant's reply to The OA in any of its classical versions:

I answer: You have already committed a contradiction when you have brought the concept of its existence, under whatever disguised name, into the concept of a thing which you think merely in terms of its possibility. If one allows you to do that, then you have won the illusion of a victory, but in fact you have said nothing; for you have committed a mere tautology. I ask you: is the proposition This or that thing (which I have conceded to you as possible, whatever it may be) exists—is this proposition, I say, an analytic or synthetic proposition? If it is the former then with existence you add nothing to your thought of the thing; but then either the thought that is in you must be the thing itself, or else you haver inferred that existence on this pretext from its inner possibility, which is nothing but a miserable tautology. The word 'reality', which sounds different from 'existence' in the concept of the predicate, does not settle it. For if you call all positing (leaving indeterminate what you posit) 'reality', then you have already posited the thing with all its predicates in the concept of the subject and assumed it to be actual, and you only repeat that in the predicate. If you concede, on the contrary, as in all fairness you must, that every existential proposition is synthetic, then how would you assert that the predicate of existence may not be cancelled without contradiction?—since this privilege pertains only in the analytic propositions, as resting on its very character. I would have hoped to annihilate this over-subtle argumentation without any digressions through a precise determination of the concept of existence, if I had not found that the illusion consisting in the conusion of a logical predicate with a real one (i.e., the determination of a thing) nearly precludes all instruction. Anything one likes can serve as a logical predicate, even the subject can be predicated of itself; for logic abstracts from every content. But the determination is a predicate, which goes beyond the concept of a subject and enlarges it. Thus it must not be included in it already. Being is obviously not a real predicate, i.e., a concept of something that could add to the concept of a thing. It is merely the positing of a thing or of certain determinations in themselves. In the logical use it is merely the copula of a judgment…. [T]he little word 'is' is not a predicate, but only that which posits the predicate in relation to the subject. Now if I take the subject (God) together with all its predicates … and say God is, or there is a God, then I add no new predicate to the concept of God, but only posit the subject in itself with all its predicates, and indeed posit the object in relation to my concept. Both must contain exactly the

same, and hence when I think this object as given absolutely (though the expression, 'it is') nothing is thereby added to the concept, which expresses merely its possibility. Thus the actual contains nothing more than the merely possible. A hundred actual dollars do not contain the least bit more than a hundred possible ones. For since the latter signifies the concept and the former its object and its positing in itself, then, in case the former contained more than the latter, my concept would not express the entire object and thus would not be the suitable concept of it. But in my financial condition there is more with a hundred actual dollars than with the mere concept of them (i.e., their possibility). For with actuality the object is not merely included in my concept analytically, but adds synthetically to my concept (which is a determination of my state); yet the hundred dollars themselves that I am thinking of are not in the least increased through this being outside my concept. Thus when I think a thing, through whichever and however many predicates I like (even in its thoroughgoing determination), not the least bit gets added to the thing when I posit in addition that this thing is. For otherwise what would exist would not be the same as what I had thought in my concept, but more than that, and I could not say that the very object of my concept exists…. Now if I think of a being as the highest reality (without defect), the question still remains whether it exists or not…. Thus whatever and however much our concept of an object may contain, we must go out beyond it to provide it with existence. With objects of sense this happens through connection with some perception of mine in accordance with empirical laws; but for objects of of pure thinking thete is no means whatever for cognizing their existence, because it would have to be cognized entirely *a priori*, but our consciousness of all existence (whether immediately through perception, or through inference connecting something with perception belongs entirely and wihout exception to the unity of experience, and though an existence outside the field cannot be declared absolutely impossible, it is a presupposition that we cannot justify through anything. (*CPR* A597-601/B625-629)

Reduced to its essentials, Kant's critique of The OA consists of three distinct

parts.

(Part i) "Exists" is a logical and not a "real" or "determining" predicate: more precisely, "exists" is a second-order concept C_2 which says of some first-order concept C_1 that C_1 has at least one instance.

(Part ii) The category of existence, when schematized, yields the schematized category of reality or actuality (*Realität, Wirklichkeit*).

(Part iii) Objectively valid and true existence-judgments (e.g., "Socrates exists") are synthetic (hence their meaning and truth is based on intuition), not analytic (hence their meaning and truth is not based solely on concepts).

Each of these theses needs to be unpacked more. I will do that for them severally and

then re-combine them into a single thesis about The OA.

Re (Part i). According to Kant, *logical predicates* or *logical concepts* are those

concepts whose application to another concept does not change or augment the semantic

content of the second concept, although it may nevertheless change or augment the

second concept's psychological or logical form. For example, applying the logical

operation of analytical decomposition to the concept BACHELOR yields the several

ordered constituents of its conceptual microstructure, i.e.,

<UNMARRIED + ADULT + MALE>

but does not in any way change or augment the semantic content of that concept.

Nevertheless the decomposition operation itself *does* generate new semantic information,

i.e., direct insight into the microstructure of that concept. (This, by the way, would be the

key to a Kantian solution of "the paradox of analysis."[19]) Again, applying the logical

operation of negation to the concept CAT yields NON-CAT but does not in any way

change or augment CAT's semantic content. CAT's semantic content is its *intension*,

and this intension uniquely determines CAT's cross-possible-worlds *extension* or

semantic value, i.e., the set of all actual and possible cats. Nevertheless the negation

operation as applied to CAT itself *does* generate a new semantic value, namely the set of

all non-cats.

By contrast, *real predicates, determining predicates*, or *determining concepts* are

precisely those concepts whose application to another concept does indeed change and

augment the semantic content of the second concept. For example, RED is a real or

determining concept whose application to the concept ROSE modifies the latter's content

or "intension" (*Inhalt*) by further specifying it and also, correspondingly, narrowing its

non-empty cross-possible-experienceable-worlds extension or "comprehension"

(*Umfang*). More generally, however, real predicates, determining predicates, or

determining concepts are all and only first-order, objectively valid concepts with an

analyzable or decomposable intension and a non-empty comprehension.

Now EXISTS is merely a logical concept in that applying it to the concept of, say,

ONE HUNDRED DOLLARS, does not in any way change or augment the latter's

semantic content. Notice that Kant does *not* say that applying EXISTS to another concept is either meaningless or vacuous. Having an existent one hundred dollars in my pocket is quite different from a merely possible one hundred dollars. Similarly, Kant does *not* say that EXISTS is *not* an authentic predicate: on the contrary EXISTS *is* an authentic predicate. It is just that it is a logical predicate and not a real or determining predicate.

Q: What more precisely does the logical concept EXISTS mean when it is applied to another concept?

A: The concept EXISTS is a second-order concept which says that the concept to which it is applied has instances in some non-empty domain of objects. So EXISTS is a second-order predicate that functions in essentially the same way as the existential quantifier of first-order predicate logic.

Re (Part ii). For Kant, the concept EXISTS is empirically meaningful or objectively valid only when it is schematized by the representations of time and space, and then, as schematized, EXISTS says that the concept to which it is applied has empirically intuitable or sense-perceivable instances at some time or another in the empirical world. Otherwise put, the schematized concept EXISTS means the same as the concepts REAL and ACTUAL. The Anticipations of Perception tell us that for something to be real is for it to be an empirically intuitable object of sense-perception having some positive degree of intensive magnitude (force). And the Postulates of Empirical Thought tell us that for something to be actual is for it to be given in empirical intuition at some time or another.

Re (Part iii). If EXISTS is a logical predicate but not a real or determining predicate, and if the concept EXISTS is a second-order concept which means that the

concept to which it is applied has instances in some non-empty domain of objects, and if the schematized concept EXISTS means the same as REAL and ACTUAL, then to apply EXISTS to another concept in an objectively valid judgment (e.g., "Socrates exists") is to say of the second concept that it has empirically intuitable real or actual instances in the empirical natural world. Hence "X exists" is true if and only if something falling under the concept X has empirically intuitable real or actual instances in the empirical natural world. Any judgment whose whose meaning and truth depend on empirical intuition is synthetic. Hence every objectively valid and true existential judgment is synthetic.

Q: How does this all apply to The OA?

A: In two ways, as follows—

First, The OA fallaciously and fatally errs by treating the concept EXISTS as if it were a real or determining predicate. But EXISTS is neither a real or determining predicate nor is EXISTS ever contained analytically in any other real or determining predicate. Therefore all arguments purporting to show that the concept EXISTS is analytically contained in the concept GOD are fallacious and unsound.

The very same point also goes for NECESSARILY EXISTS. This is simply because NECESSARILY EXISTS, just like EXISTS, is a logical predicate but not a real or determining predicate. Hence the objection which is sometimes made against Kant's critique of The OA and in favor of some or another post-classical version of The OA,[20] to the effect that even if EXISTS is not a real predicate, nevertheless NECESSARILY EXISTS can be a real predicate of God, is doubly mistaken. <u>First</u>, it is mistaken because it wrongly assumes that Kant's technical term "real predicate" means the same as "authentic predicate," whereas as we have seen that in fact it means the same as *real or*

determining predicate, which is a completely different notion. And <u>second</u>, it is mistaken because if EXISTS is a second-order predicate which does not operate like a first-order real or determining predicate, then obviously NECESSARILY EXISTS is also *another* second-order predicate, which *also* does not operate like a first-order real or determining predicate.

Second, consider the judgment or proposition "God exists." It is true just in case (i) GOD is objectively valid, and (ii) GOD has empirically intuitable instances. But GOD is not objectively valid, hence "God exists" is a truth-value gap. Moreover even if, *per impossibile*, "God exists" *were* true, that judgment could only ever be synthetic, not analytic.

There are also three extremely important logical, semantic, and epistemological consequences of The OA.

First, as I noted above, the impossibility of The OA generalizes to the impossibility of any strict scientific proof or strict scientific knowledge of God's existence.

Second, Kant's critique of The OA also entails a general solution to a longstanding problem in philosophical logic: the problem of the correct analysis of negative existential propositions, a problem which goes back at least as far as Plato's *Sophist*, but which also seriously worried Frege, Russell, and many other major philosophical logicians after them. The problem is this: If a word has to have a reference in order for it to be meaningful, then how can existence ever be truly denied of anything? In other words, it seems paradoxical to assert "X does not exist" wherever what replaces 'X' is a meaningful word: e.g., "Superman does not exist." Kant's critique of The OA

shows us that wherever true or false existential predications are made, the subject-term of

the proposition *stands for an objectively valid concept, not an object.* And some

objectively valid concepts have a null real-world or actual-world extension, e.g., the

empirical concept SUPERMAN. So it is not generally true that a word has to have a

reference in order for it to be meaningful: words can stand for concepts, and concepts

need not be instantiated in the real or actual world. Then when a word—e.g.,

'Superman'—stands for a concept that has no real or actual instances, then it can be truly

and non-paradoxically said that X does not exist. Thus an existential proposition is true

just in case

> (i) the subject concept of the proposition is objectively valid and has some real or actual
> instances,

and a negative existential proposition is true just in case

> (ii) the subject concept of the proposition is objectively valid and has no real or actual
> instances.

Third, the concept SUPERMAN and the concept GOD are radically different

concepts. The concept SUPERMAN is an objectively valid empirical concept with (as it

so happens) a null real or actual world extension. But the concept GOD is *not* an

objectively valid concept, and therefore *not* an empirical concept. On the contrary, GOD

is a (positive) noumenal concept, i.,e., a "notion," or an Idea of Pure Reason. Hence GOD

is not cognizable, but instead only thinkable. This means that *neither* the proposition

"God exists" *nor* the proposition "God does not exist" has a classical truth-value: indeed,

both "God exists" and also "God does not exist" are *truth-value gaps.* This in turn means

that *atheism* is every bit as closed to logical proof or strict scientific knowledge as *theism*

or *deism* are. For example, and perhaps most importantly, both the metaphysical and

evidential arguments for atheism from the existence of evil are impossible, just as The

Ontological Argument is impossible. *Philosophical theology contains unprovable propositions.* As we will see in the next sub-section, this is a logico-semantically profound result which is comparable in its *moral and practical* significance to the *cognitive and theoretical* significance of Kurt Gödel's logico-semantic demonstration in the 1930s that the system of elementary or Peano arithmetic (i.e., elementary logic plus the five Peano axioms) contains unprovable (and undecidable) sentences.[21]

It is relevantly interesting and philosophically ironic in this connection that Gödel *also* developed a version of The Ontological Argument.[22] It seems clear that Gödel intended his version of The OA to be strictly a pump for *rational intuition*, which for the later Gödel can exceed logical provability in the narrower senses of either decidability or formal provability in elementary logic or elementary arithmetic[23]—hence Gödel did not hold that the existence of God is logically provable in those narrower senses. Nevertheless, this indirectly shows that Kant's strictures on *analytic* logical provability significantly anticipate and mirror Gödel's strictures on logical provability in the *narrower* senses.

III.4 The Immortality Postulate, The God Postulate, and What They Really Mean

We know from the Paralogisms and the Ideal of Pure Reason that both the idea of the human soul and the idea of God are *scientifically unknowable Ideas of pure reason.* Correspondingly, both the immortality of the soul and the existence of God are *logically and scientifically unprovable propositions.* Neither their truth nor their falsity can be demonstrated. Hence the correct philosophical attitude to take towards them *is radical agnosticism.* But the Ideas of immortality, of freedom, and of God's existence still can

have regulative, practical significance as *postulates of pure practical reason*. Here is

what Kant writes:

> The production of the highest good in the world is the necessary object of a will determinable by the moral law. But in such a will the *complete conformity* of dispositions with the moral law is the supreme condition of the highest good. This conformity must be just as possible as its object is, since it is contained in the sane command to promote the object. Complete conformity of the will with the moral law is, however, *holiness*, a perfection of which no rational being of the sensible world is capable at any moment of his existence. Since it is nevertheless required as practically necessary, it can only be found in an *endless progress* toward the complete conformity, and in accrdance with principles of pure practical reson it is necessary to assume such a practical progress as the real object of our will. This endless progress is, however, possible only on the presupposition of the *existence* and personality of the same rational being continuing *endlessly* (which is called the immortality of the soul). Hence the highest good is practically possible only on the presupposition of the immortality of the soul, so that this, as inseparably connected with the moral law, is a *postulate* of pure practical reason…. For a rational but finite being only endless progress from lower to higher stages of moral perfection is possible. *The eternal being*, to whom the temporal condition is nothing, sees in what is to us an endless series, the whole of conformity with the moral law, and the holiness that his command inflexibly requires in order to be commensurable with his justice in the share he determines for each in the highest good is to be found whole in a single intellectual intuition of the existence of rational beings. All that a creature can have with respect to hope for this share is consciousness of his tried disposition, so that, from the progress he has already made from the worse to the morally better and from the immutable resolution he has thereby come to know, he may hope for a further uninterrupted continuance of this progress, however long his existence may last, even beyond this life, and thus he cannot hope, either here or in any foreseeable future moment of his existence, to be fully adequate to God's will (without indulgence or dispensation, which do not harmonize with justice); he can only hope to be so only in the endlessness of is duration (which God alone can survey). (*CPrR* 5: 122-124)

> *Happiness* is the state of a rational being in the world in the whole of whose existence *everything goes according to his wish and will*, and rests, therefore, on the harmony of nature with his whole end as well as with the essential determining ground of his will. Now, the moral law as a law of freedom commands through determining grounds that are to be quite independent of nature and of its harmony with our faculty of desire (as incentives); the acting rational being in the world is, however, not also the cause of the world and of nature itself. Consequently, there is not the least ground in the moral law for a necessary connection between the morality and the proportionate happiness of a being belonging to the world as part of it and hence dependent upon it, who for that reason cannot by his will be a cause of this nature and, as far as his happiness is concerned, cannot by his own powers make it harmonize thoroughly with his practical principles. Nevertheless, in the practical task of pure reason, that is, in the necessary pursuit of the highest good, such a connection is postulated as necessary: we *ought* to strive to promote the highest good (which must therefore be possible). Accordingly the existence of a cause of all nature, distinct from nature, which contains the ground of this connection, namely of the exact correspondence of happiness with morality, is also *postulated*. However, this supreme cause is to contain the ground of the correspondence of nature not merely with a law of the will of rational beings but with the representation of this *law*, so far as they make it the *supreme determining ground of the will*, and consequently not merely with morals in their form but also with their morality as their determining ground, that is, with their moral disposition. Therefore the highest good in the world is possible only insofar as a supreme cause of nature having a causality in keeping with the moral disposition is assumed. Now a being capable of actions in accordance with the representation of laws is *an intelligence* (a rational being), and the causality of such a being in accordance with his representation of laws is his *will*. Therefore the supreme cause of nature, insofar as it must be presupposed for the highest good, is a being that is the cause of nature by *understanding* and *will* (hence its author), that is, *God*. Consequently, the postulate of the possibility of the *highest*

derived good (the best world) is likewise the postulate of the reality of a *highest original good*, namely of the existence of God. (*CPrR* 5: 124-125)

[The postulates of pure practical reason proceed from the principle of morality, which is not a postulate but a law by which reason determines the will immediately; and this will, just because it is so determined as a pure will, requires these necessary conditions for the observance of its precept. These postulates are not theroretical dogmas but *presuppositions* having a necessarily practical reference and thus, although they do not indeed extend speculative cognition, they give objective treality to the ideas of speculative reason in *general* (by means of their reference to what is practical) and justify its holding concepts even the possibility of which it could not otherwise presume to affirm. These postulates are those of *immortality*, of *freedom* considered positively (as the causality of a being insofar as it belongs to the intelligible world), and of the *existence of God*. The *first*, flows from the practically necessary condition of a duration befitting the complete fulfillment of the moral law; the *second* from the necessary presupposition of independence from the sensible world and of the capacity to determine one's will by the law of an intelligible world, that is, the law of freedom; the *third* from the necessity of the condition for such an intelligible world to be the highest good, through the presupposition of the highest independent good, that is, of the existence of God. (*CPrR* 5: 132)

As I mentioned above, the first postulate of pure practical reason, *immortality*, basically says that even though we cannot either logically prove or scientifically know either that our souls are immortal or that they are not immortal, we must nevertheless morally believe in the rational Idea that we will have an endless human personal existence in a world that is wholly known and governed by God, and in which eventually all the morally virtuous people are made happy and all the wicked people are punished *not for retributive reasons* but instead precisely so that they can finally face up to their sins, take full responsibility for them, and then change their lives for the better. Notice especially, again, that there is no such thing as *capital punishment* in the God-governed world of immortal rational human animals: on the contrary, in the Kantian after-life, everyone gets *an infinite number of "second chances" to redeem themselves*.

And the third postulate of pure practical reason, *the existence of God*, basically says that even though we cannot logically prove or scientifically know either that God exists or that God does not exist, we must nevertheless morally believe in the Idea that God exists in order to unify happiness and virtue in a desperately nonideal world filled to

the brim with "the crooked timber of humanity" and "radical evil," and in which, it seems, nothing will ever be made straight, and "no good deed ever goes unpunished."

I will come back explicitly to the second postulate, *freedom*, in sub-section III.5. So bracketing The Freedom Postulate for the moment, what does Kant mean by all this? He certainly does not hold that we have *logical or scientific justification* for believing either that personal immortality is really possible or that God exists. Moreover, neither personal immortality nor God's existence can be "proved through experience" in an essentially non-conceptual, directly volitional way, as practical freedom can (*CPR*: A802-803/B831). Hence neither personal immortality nor God's existence has *practical reality* in the sense that freedom has practical reality—i.e., there is no "Fact of Reason" for either personal immortality or God's existence, as there is for freedom:

> The consciousness of this fundamental law [of pure practical reason, which says: so act that the maxim of your will could always hold at the same time as a principle of universal law giving] may be called a fact of reason, since one cannot ferret it out from antecedent data of reason, *such as the consciousness of freedom* (for this is not antecedently given), and since it forces itself upon us as a synthetic proposition a priori based on no pure or empirical intuition… In order to regard this law without any misinterpretation as given, one must note that it is not an empirical fact, but the sole fact of pure reason, which by it proclaims itself as originating law. (*CPrR* 5: 31, underlining added—see also *CPrR* 5: 42, 47, and 55-56)

So here is what I think The Immortality Postulate and The God Postulate really mean. I think that Kant's moral theology is radically different from any form of transcendental or noumenal theology. More precisely, I think that in order to solve The Incoherence Problem, then Kant's moral theology should be understood as *Transcendental Existentialism*. I also believe that Transcendental Existentialism is both

(i) later developed explicitly by early Wittgenstein in the *Notebooks 1914-1916* and the *Tractatus*,

and also

(ii) independently philosophically defensible, quite apart from Kant's and Wittgenstein's texts.

In any case, here is a preliminary sketch of Transcendental Existentialism in eight steps.

First, Transcendental Existentialism contains Parts 1-4 of Kant's philosophical theology as I described them in section II.

Second, Transcendental Existentialism contains the three special features I also described in section II: (1) radical agnosticism, (2) believing-in-as-volitional-commitment, "faith," or *Glauben*, and (3) moral certainty.

Third, Kantian radical agnosticism means my taking the scientific knowledge that God's *non*-existence is scientifically unknowable and uncognizable every bit as seriously as I take the scientific knowledge that God's *existence* is scientifically unknowable and uncognizable, although still thinkable.

Here is where classical arguments for atheism from the existence of natural evil and moral evil become directly relevant to Transcendental Existentialism.[24] The classical *Metaphysical Argument for Atheism from the Existence of Evil* runs as follows:

> 1. Assume that God exists and is omnipotent, omniscient, and omnibenevolent. Or in other and fewer words, assume that *a 3-O God* exists. (Premise.)
>
> 2. Assume that evil exists in the world—both natural evil (e.g., disasters and disease) & also moral evil (wicked choices and acts, or just bad things that happen to people). (Premise.)
>
> 3. Then EITHER a 3-O God is responsible for the existence of evil, in which case a 3-O God is Her/Himself evil and not all-good, which is a contradiction with God's assumed 3-O-ness. (From 1 and 2.)
>
> 4. OR a 3-O God is not responsible for the existence of evil and yet knew that it was going to happen and could not prevent it—so a 3-O God is not all-powerful, which is also a contradiction with assumed God's 3-O-ness. (From 1 and 2.)
>
> 5. OR a 3-O God would have prevented evil but did not know it was going to happen, and is not all-knowing, which is another contradiction with God's assumed 3-O-ness. (From 1 and 2.)
>
> 6. Therefore, given the existence of evil, necessarily a 3-O God does not exist. (From 1, 2, 3, 4, and 5.)

If The Metaphysical Argument from Evil were sound, then it would show that it is analytically necessary that a 3-O God does not exist.

In the classical theistic critical response to The Metaphysical Argument from Evil, it is claimed that it is at least logically possible that God has a sufficient reason for permitting evil that we are either capable of knowing or else simply incapable of knowing, given our finite epistemic powers. Perhaps this sufficient reason is The Best of all Possible Worlds doctrine; perhaps it is free will; perhaps it is moral progress; perhaps it is all three of these taken together; or perhaps it is something else completely unfathomable by us. Let us call this classical response *Theodicy*. In response to Theodicy, the neo-classical *Evidential Argument for Atheism from the Existence of Evil* says that even if it is logically possible that God has a sufficient reason for permitting evil, nevertheless it is significantly more rationally justified to believe that God does not exist, than to believe that God exists.

But as we have seen in sub-section III.3, for strictly logico-semantic reasons, neither God's existence *nor* God's non-existence is scientifically provable. Hence not only The Metaphysical Argument from Evil, but also Theodicy, as well as The Evidential Argument from Evil, are *equally* rationally ungrounded. This radically agnostic fact, in turn, puts The Problem of Evil in a completely new light. If natural evil and moral evil both exist, and there is lots of it all over the place, but God's existence and non-existence are both scientifically unprovable, then natural evil and moral evil *are entirely up to us to deal with*. We and we alone must deal with natural evil and moral evil, as best we can, by cleaning up or fixing up the natural world when it breaks down and Stuff Happens, and by trying our damnedest to be morally good in a desperately nonideal world. Either God

does not exist, and evil is simply a massive natural and moral challenge for us; or else God does exist, it is all part of God's plan, and we must do God's work. But since both options are *equally* scientifically unprovable, and yet at the same time *equally* intelligible and relevant to us, then we must comport ourselves *accordingly*. I will come back briefly to this fundamental point again at the end of sub-section III.5.

Fourth, it is *not* Kant's view that we have a sufficient practical reason to believe what we do not have sufficient epistemic to reason to believe, namely that God exists and that there is immortality of the soul. Instead it is Kant's view that we have sufficient practical reason for *choosing and acting as if, counterfactually, we believe* that God exists and that there is immortality of the soul, even while also *scientifically knowing* that these propositions are scientifically unknowable and uncognizable, while still remaining thinkable. This is the same as *believing-in* the Idea of God's existence, i.e., the same as having moral *certainty*, *faith*, or *Glaube* in God's existence.

Fifth, according to Kant, given radical agnosticism, the notion of believing-in-as-volitional-commitment, and the notion of moral certainty, then for me to believe-in God's existence and for me to believe-in the immortality of the soul, is *non-cognitively equivalent* to my believing that life itself has absolute moral meaning and also to my believing that my own life has an absolute moral meaning, which in turn are *non-cognitively equivalent* to my being morally certain that life itself has absolute moral meaning and also to my being morally certain that my own life has an absolute moral meaning. This extremely deep idea was either earlier anticipated or later expressed by the developers of theistic Existentialism, particularly by Augustine, Pascal, Spinoza, and

Kierkegaard, and also—as we have already anticipated in the sixth epigraph of this section—by the early Wittgenstein:

> To believe in God means to understand the problem about the meaning of life. To believe in God means to see that the facts of the world are not the end of the matter. To believe in God means to see that life has a meaning. (*CV* 74ᵉ)

Sixth, therefore according to Kant, my *soundly but non-scientifically proving* that God exists and that there is immortality of the soul is non-cognitively equivalent to my *soundly but non-scientifically proving* that life itself has absolute moral meaning and that my own life has an absolute moral meaning, which in turn are non-cognitively equivalent to my *becoming morally certain* that life itself has absolute moral meaning and that my own life has an absolute moral meaning.

Seventh, therefore according to Kant, *the only acceptable way* of soundly but non-scientifically proving that life itself has absolute moral meaning and that my own life has an absolute moral meaning, which in turn are non-cognitively equivalent to my becoming morally certain of, or having faith, or *Glaube*, in life itself's having an absolute moral meaning and also in my own life's having an absolute moral meaning, is *by actually going forth and having a morally meaningful life* by means of my autonomous power of choice, and the pursuit of good willing and the worthiness to be happy, under the constitutive presuppositions that I choose and act *as if, counterfactually, I believe* that God exists and that there is immortality of the soul

Eighth, therefore according to Kant, the only acceptable way of soundly non-scientifically proving that God exists and that there is immortality of the soul is *by soundly non-scientifically and non-cognitively proving myself as an autonomous moral agent*—i.e., by actually going forth and having a morally meaningful life, finally facing up to and taking full responsibility for all my sins, and changing my life for the better—

and thereby *by actually becoming the kind of person I would be* if God *were* to exist and

if there *were* immortality of the soul, although I *scientifically know* that these

propositions are both scientifically unknowable and uncognizable, yet still thinkable. This

emotional and practical project begins as a fundamental "loss of faith" or anxiety (*Angst*),

particularly in the face of The Problem of Evil, but it ends as a fundamental moral "leap

of faith" or groundless affirmation, as in Kierkegaard's sublime version of the story of

Abraham and Isaac,[25] when this is interpreted in terms of radical agnosticism, the notion

of believing-in-as-volitional-commitment, and the notion of moral certainty, faith, or

Glaube.

This Kantian-Kierkegaardian "moral leap of faith" is not a scientific proof of

God's existence and the immortality of the soul in the *classical logical* or analytic sense

of a valid or sound argument in first-order bivalent polyadic quantified logic with

identity, but it is not *irrational* either. For logic is *not* exhausted by classical logic: there

is still *non*-classical logic. Thus the Kantian-Kierkegaardian "moral leap of faith" is a

non-scientific and non-cognitive but still fully practically rational and non-classically

logically acceptable proof in the special sense of *Intuitionistic constructivist* mathematics

or logic,[26] which provides for an inherently ruled-governed step-by-step generation of an

actual token of the ideal type whose existence you are demonstrating. And just as in

Intuitionistic logic, so too in Transcendental Existentialism the classical logical Principle

of Excluded Middle does *not* apply to the proposition that God exists, for as we have

seen, it is strictly logically unprovable and scientifically unknowable either that God

exists or that God does not exist. Because it is strictly logically unprovable and

scientifically unknowable whether God exists or not, then precisely the right emotional

and practical attitude to take towards the question of God's existence or non-existence is radical agnosticism. But for someone to believe-in God's existence or believe-in the immortality of the soul, and then to non-scientifically and non-cognitively prove it, is for her to prove non-scientifically and non-cognitively, in an Intuitionistic constructivist way, that her life has a meaning by virtue of its categorically normative moral content, via her steady step-by-step pursuit of a life of wholehearted commitment to her own projects, along with other rational human agents, or real human persons, as fully embedded in the larger natural world, under absolute moral principles, thereby producing *an actual token in her own life* of the ideal types whose actual existence she is non-scientifically and non-cognitively demonstrating.

If these eight steps are cogent, then Kant's philosophical theology is not a classical transcendental or noumenal theology in any sense, but sharply on the contrary, it is an *Existentialism* of an altogether unique and transcendental kind, specifically because of its background metaphysics of Transcendental Idealism and also because of its important formal analogies with non-classical, Intuitionistic constructivist mathematics and logic. For not only does modern Existentialism clearly flow *historically* from Kant's Transcendental Idealism and his philosophical theology, but much more importantly it also seems to me that contemporary moral theorists and philosophical theologians could significantly learn from Transcendental Existentialism, in view of its solid foundations in independently defensible Kantian philosophical logic, cognitive semantics, epistemology, Transcendental Idealist metaphysics, and Intuitionistic mathematics and logic, and also in view of the very real importance of Kantian ethics in contemporary moral theory.

In any case, what ultimately matters according to Transcendental Existentialism is my actively believing-in and having moral certainty, faith, or *Glaube* about the real possibility that my life has a meaning and categorically normative moral content, via radical agnosticism about about God's existence and the immortality of the soul alike. If a person actively believes-in the real possibility that her life has a meaning and categorically normative moral content, then just by virtue of that moral commitment itself, and her Intuitionistic constructive non-scientific and non-cognitive proof of this, then her life necessarily *does* have a meaning and categorically normative moral content. This is a truly remarkable *Existential bootstrapping* feature of the moral metaphysics of rational human agency. Unlike moral virtue, which, as everyone knows, and as Kantians always emphasize, can often be extremely lonely, self-repressing and therefore depressing, unpleasant, and very unrewarding in an everyday sense, despite its non-denumerable absolute intrinsic value, rational human agency genuinely *can be* and *is* its own reward. So *ought* implies *can*, and with active believing-in, moral certainty, faith, or *Glaube*, then *can* also implies *is*.

III.5 The Freedom Postulate, Wittgenstein's Mystical Compatibilism, the Two Fallacies of Freedom-Inauthenticity, and Our Faith in Free Will

I turn now to the second postulate of pure practical reason, *freedom*, which basically says that *because* we cannot logically prove or scientifically know either that God exists or that God does not exist, we must *therefore* morally believe in the rational Idea of our own transcendental freedom and practical freedom in order to rule out:

> (i) the self-stultifying threat to our transcendentally free intentional agency of a world in which, it seems, on the *one* hand, that the past is completely filled with deterministic and impersonal causes (the rock) and, on the *other* hand, that the future is randomly indeterministic and heart-breakingly completely filled with possibilities for bad luck (the hard place), and

(iii) the self-stultifying threat to our practically free moral agency of a world in which, it seems, because God's existence is unprovable, than moral chaos reigns and "everything is permitted."

In order to understand this two-part doctrine properly, we I will briefly unpack some preliminary points in Kant's metaphysics of free will, and then in the metaphysics of free will more generally.

Kant's theory of transcendental freedom is his metaphysics of free will. Transcendental freedom is how a person can, "from itself" (*von selbst*) (*CPR* A533/B561), be the spontaneous mental cause of certain natural events or processes. If I am that person, then insofar as I am transcendentally free, it follows that certain events or processes in physical nature are *up to me*—or to use Kant's own phrase, *in meiner Gewalt* (literally: "in my control" or "in my power"; *CPrR* 5: 94-95). So otherwise put, transcendental freedom is deep freedom of the will, or up-to-me-ness (as it were, *In-Meiner-Gewalt-Sein*).

Transcendental freedom is the same as absolutely spontaneous mental causation:

> By freedom in the cosmological sense … I <u>understand the faculty of beginning a state *from itself* (*von selbst*), the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature</u>. Freedom in this signification is a pure transcendental idea, which, first, contains nothing borrowed from experience, and second, the obhect of which cannot be given determinately in any experience…. But since in such a way no absolute totality of [natural] conditions in causal relations is forthcoming, <u>reason creates the idea of a spontaneity, which could start to act from itself, without needing to be preceded by any other cause that in turn determines it to action according to the law of causal connection</u>. (*CPR* A533/B561, underlining added)

Although transcendental freedom is a particularly robust kind of *mental* causation, in the second *Critique* Kant sharply distinguishes distinguishes transcendental freedom from mere *psychological* freedom:

> These determining representations [i.e., instincts or motives] themselves have the ground of their existence in time and indeed in the *antecedent state*, and in a preceding state, and so forth, these determinations may be internal and they may have psychological instead of mechanical causality, this is, <u>produce actions by means of representations and not by bodily movements; they are always *determining grounds* of the causality of a being insofar as its existence is determinable in time and</u>

therefore under conditions of past time, which are thus, when the subject is to act, *no longer within his control* and which may therefore bring with them psychological freedom (if one wants to use this term for a merely internal chain of representations in the soul) but nevertheless natural necessity, leaving no room for *transcendental freedom* which must be thought of as independence from everything empirical and so from nature generally, whether regarded as an object of inner sense in time only or also as an object of outer sense in both space and time; without this freedom (in the latter and proper sense), which alone is practical a priori, no moral law is possible and no imputation in accordance with it. (*CPrR* 5: 96-97, underlining added)

Otherwise put, psychological freedom is the subject's subjective experience or consciousness of choosing or acting without being prevented, and without inner or outer compulsion. As Kant explicitly points out, and as Hume and Leibniz also noted in anticipation of contemporary "Soft Determinist" versions of Compatibilism, it is both logically and metaphysically possible to be psychologically free without being transcendentally free. This is what Kant very aptly and famously calls "the freedom of a turnspit" (*CPrR* 5: 97). So psychological freedom is not a sufficient condition of transcendental freedom. Nevertheless, according to Kant psychological freedom remains a *necessary* condition of transcendental freedom. And this seems independently highly plausible. No one could be transcendentally free and also at the same time undergo the subjective experience or consciousness of being prevented from choosing or acting, or of being inwardly or outwardly compelled to choose or act. Indeed, as the second Analogy of Experience explicitly shows, psychological freedom is necessarily built into the mental representation of *any* objective causal sequence, via what Kant calls the "the subjective sequence of apprehension," whose ordering is always subjectively experienced as "entirely arbitrary" (*ganz beliebig*) and not necessitated (*CPR* A193/B238).

When we ascribe transcendental freedom specifically to the will of a real human person, then in addition to the positive factor of absolute spontaneity, which confers deep freedom or up-to-me-ness on the real human person's choices and acts, and psychological freedom, which guarantees the subjective experience or consciousness of being

unprevented and uncompelled in one's choices and acts, there is also a negative

dimension of freedom which guarantees the person's choices and acts occur

independently of all "alien causes," that is, independently of all pathological inner and

unowned outer sources of nomologically sufficient compulsion:

> The will is a kind of causality that living beings have so far as they are rational. *Freedom* would then be that property whereby this causality can be active, indepedently of alien causes *determining* it; just as *natural necessity* is a property characterizing the causality of all non-rational beings—the property of being determined to activity by the influence of alien causes. The above definition of freedom is *negative*. (*GMM* 4: 446, underlining added)

This is where practical freedom comes on the scene. Practical freedom presupposes but

also exceeds transcendental freedom, in that practical freedom is the absolute spontaneity

of the will independently of all alien causes and also independently of *all sensible*

impulses (empirical desires):

> *Freedom in the practical sense* is the independence of the power of choice (*Willkür*) from necessitation by impulses of sensibility. For a power of choice is sensible insofar as it is pathologically affected (through moving-causes of sensibility); it is called an animal power of choice (*arbitrium brutum*) if it can be pathologically necessitated. The human power of choice is indeed an *arbitrium sensitivum*, yet not *brutum*, but *liberum*, because sensibility does not render its action necessary, but in the human being there is a faculty of determining oneself from oneself, independently of necessitation by sensible impulses. (*CPR* A534/B562, underlining added)

But this is merely a negative characterization of practical freedom. As positively

characterized, practical freedom also involves the capacity for *self-legislation* in

conformity with the Categorical Imperative or moral law. Or in other words, practical

freedom is necessarily equivalent with autonomy (*GMM* 4: 440-441, 446-463).

It may seem, on the face of it, that there should be no direct connection

whatsoever between the person's absolutely spontaneous, psychologically free,

autonomous will and her existence in physical nature. But in fact Kant himself explicitly

asserts otherwise:

> Practical freedom can be proved through experience. For it is not merely that which stimulates the senses, i.e., immediate affects them, that determines human choice, but we always have a capacity to overcome impressions on our sensory faculty of desire by representations of that which is useful

or injurious even in a more remote way; but these considerations about that which in regard to our whole condition is desirable, i.e., good and useful, depend on reason. Hence this also yields laws that are imperatives, i.e., objective *laws of freedom*, and that say *what ought to happen*, even though it never does happen…. We thus cognize practical freedom through experience, as one of the natural causes, namely a causality of reason in the determination of the will. (*CPR* A802-803/B830-831, underlining added)

Now although there is an incalculable gulf fixed between the domain of the concept of nature, as the sensible, and the domain of the concept of freedom, as the supersensible …: yet the latter *should* have an influence on the former, namely the concept of freedom should make the end that is imposed by its laws real in the sensible world; and nature must consequently also be able to be conceived in such a way that the lawfulness of its form is at least in agreement with the possibility of the ends that are to be realized in it in accordance with the laws of freedom. (*CPJ* 5: 176, underlining added)

In other words, Kant is explicitly saying that transcendental freedom is both really (i.e., synthetic a priori) possible and real.

Now *Universal Natural Determinism* is the doctrine that the complete series of settled past events, together with the general causal laws of nature, causally or nomologically necessitate the existence and specific character of all future events, including all the choices and acts of persons. This can be formulated even more carefully. Let us adopting the following symbolic conventions, where 'p' stands for an arbitrarily chosen proposition about the natural world:

C/N-NEC: It is causally or nomologically necessary that
Pa: All settled past events are taken together as a complete series
Ln: All the general causal laws of nature are conjoined
FEp: Every fact that p about every future event is fixed

Then Universal Natural Determinism can be explicitly stated as:

(C/N-NEC) [(Pa & Ln) → FEp]

If Universal Natural Determinism is true, then it specifically follows that whatever I am choosing or doing now is causally or nomologically necessitated by the Big Bang, or by whatever it was that actually constituted and determined the causal and nomological origins of the physical world. Furthermore, Universal Natural Determinism entails that

> Causally or nomologically necessarily, if any two events E_1 and E_2 have *exactly the same past*, then E_1 and E_2 will also have *exactly the same future*.

Let us call this *The Closed Future Rule*. The basic idea of The Closed Future Rule is that the future of the larger natural world and all the persons in it is antecedently fixed with causal necessity, and that natural history and the lives of persons do not contain any inherently random factors. Everything is part of the Big Parade, and Stuff never *just* happens. It also follows directly from Universal Natural Determinism that if someone were able to know all the relevant natural facts about the past and also the general causal laws of nature, then she would be able to predict all future events a priori with scientific certainty.

For clarity's sake, it is crucial to distinguish Universal Natural Determinism from a much stronger doctrine which says that the complete series of settled past events, together with the general causal laws of nature, *logically necessitate* the existence and specific character of all future events, including all the choices and acts of persons. This is *Fatalism*. Let us also adopt this convention:

L-NEC: It is logically necessary that

Then Fatalism can be explicitly stated as:

(L-NEC) [(Pa & Ln) → FEp]

In other words, according to Fatalism there *is no logical contingency whatsoever in the causal processes of natural history or the lives of persons.* Otherwise put, according to Fatalism all the *causal links* in nature or in us are also *logically necessary links*. It follows directly from Fatalism that if someone were able to know all the relevant natural facts about the past and also the general causal laws of nature, then she would be able to predict all future events a priori with *logical* certainty.

While Fatalism is consistent with Universal Natural Determinism, nevertheless Universal Natural Determinism does not entail Fatalism. You can consistently affirm Universal Natural Determinism and deny Fatalism. As later Wittgenstein very correctly and crisply puts it:

> Fate is the antithesis of natural law. A natural law is something you try to fathom and make use of, but not fate. (*CV*: 61ᵉ)

Even if every moment's existence and specific character is in itself *logically contingent*, in the sense that it logically could have been otherwise, given all the actual facts about the past and the laws of nature, Universal Natural Determinism can still be true. Universal Natural Determinism says only that any later event in time is *causally or nomologically* necessitated to exist and have a certain specific character, *given* that the past exists in the specific way that it does exist, and *given* the specific character of the general causal laws of nature. But the past did not *logically have to be* just that way, nor did the general causal laws of nature *logically have to be* just that way. To be sure, the logical necessity of the past and the logical necessity of the general causal laws of nature are not automatically entailed by Fatalism. Yet they are still *consistent with* Fatalism.

Moreover Fatalism does not entail Universal Natural Determinism, on at least one interpretation of Fatalism. If it turned out that both the past and the general laws of nature were *logically* necessary—if, in effect, the essence of the physical world directly mirrored a system of classical logic, as e.g., in the *Tractatus*—then this ultra-Fatalism could hold true even if Universal Natural Determinism were false. Indeed, in the *Tractatus* early Wittgenstein claims that all necessity is logical necessity and that causal or nomological necessity is not only impossible but even unintelligible:

> 5.133 All inference takes place a priori.

5.134 From an elementary proposition no other can be inferred.

5.135 In no way can an inference be made from the existence of one state of affairs to the existence of another entirely different from it.

5.136 There is no causal nexus which justifies such an inference.

5.1361 The events of the future *cannot* be inferred from those of the present. Superstition is the belief in the causal nexus. (*TLP* 109)

6.37 A necessity for one thing to happen because another has happened does not exist. There is only *logical* necessity. (*TLP* 181)

Early Wittgenstein's extremely interesting philosophical response to his own ultra-Fatalism is what I will call *Mystical Compatibilism*:

6.421 It is clear that ethics cannot be expressed.

Ethics is transcendental.

(Ethics and aesthetics are one.) (*TLP* 183)

6.422 Of the will as the subject of ethics we cannot speak.

And the will as a phenomenon is only of interest to psychology.

6.43 If good or bad willing changes the world, it can only change the limits of the world, not the facts; not the things that can be expressed in language.

In brief, the world must thereby become quite another. It must so to speak wax or wane as a whole.

The world of the happy is quite another than the world of the unhappy. (*TLP* 185)

6.44 The intuition (*Anschauung*) of the world sub specie aeterni is its intuition as a limited whole.

The feeling of the world as a limited whole is the mystical feeling. (*TLP* 187)

Wittgenstein's ultra-Fatalism clearly brings out the crucial point that Universal Natural Determinism is about the *causal or nomological* necessity of the future, not about the *logical* necessity of the future. Similarly, Universal Natural Determinism cannot logically guarantee that any particular moment of time will actually exist. For all that Universal Natural Determinism says, it is logically possible that the world *might never have existed*. Of course, the world does actually exist now. So either the world always existed, or perhaps the world started to exist and then continued to exist until now, or else the world

pops in and out of existence discontinuously. But in any case, it is always logically

possible that it might also *fail* to exist at any later time.

In 1919, in a letter to Ludwig von Ficker, editor of the journal *Der Brenner*,

Wittgenstein glossed the *Tractatus* as follows:

> The book's point is an ethical one. I once meant to include in the preface a sentence which is not in fact there but which I will write out for you here, because it will perhaps be a key to my work for you. What I meant to write then, was this: My work consists of two parts; the one presented here plus all that I have *not* written. And it is precisely this second part which is the important one. My book draws limits to the sphere of the ethical from the inside as it were, and I am convinced that this is the *ONLY rigorous* way of drawing these limits. In short, I believe that where *many* others today are just *gassing*, I have managed in my book to put everything firmly in place by being silent about it. [27]

This letter has often been dismissed by commentators as an intentionally misleading

attempt by Wittgenstein to interest a non-philosopher in publishing the *Tractatus*. And it

is true that at the time he was having difficulties getting the *Tractatus* published. Even so,

I think that it would be a big mistake not to take these remarks seriously, as a self-

commentary on the following propositions about aesthetics, ethics, God, and the meaning

of life in the *Notebooks* and the *Tractatus* (some of which I have already quoted just

above):

<div align="right">21.7.16</div>

> What really is the situation of the human will? I will call 'will' first and foremost the bearer of good and evil. (*NB*: 76^e)

> Let us imagine a man who could use none of his limbs and hence could, in the ordinary sense, not exercise his *will*. He could, however, think and *want* and communicate his thoughts to someone else. He could therefore do good or evil through the other man. Then it is clear that ethics would have validity for him, too, and that he in the *ethical sense* is the bearer of a *will*. (*NB*: 76^e-77^e)

> The World and Life are one.
> Physiological life is not of course "Life." And neither is psychological life. Life is the world.

> Ethics does not treat of the world. Ethics must be a condition of the world, like logic.
> Ethics and aesthetics are one. (*NB*: 77^e)

> It seems one can't say anything more than: Live happily!
> The world of the happy is a different world from that of the unhappy.
> The world of the happy is *a happy world*.

I keep on coming back to this! simply the happy life is good, the unhappy bad. If I now ask myself: but why should I live *happily*, then this of itself seems to me to be a tautological question; the happy life seems to be justified, of itself, it seems that it *is* the only right life.

But this is really in some sense deeply mysterious! It is clear that ethics cannot be expressed!

What is the objective mark of the happy, harmonious life? Here it is again clear that there cannot be any such mark, that can be described.

This mark cannot be a physical one but only a metaphysical one, a transcendental one. (*NB*: 78ᵉ)

Ethics is transcendental.
How things stand, is God.
God, is how things stand.
Only from the consciousness of the uniqueness of my life arises religion ... and art.

2.8.16

And this consciousness is life itself.
Can there be any ethics if there is no living being but myself?
If ethics is supposed to be something fundamental, there can.
If I am right, then it is not sufficient for the ethical judgment that a world is given.
Then the world in itself is neither good nor evil.

Good and evil enter only through the subject. And the subject is not part of the world, but a boundary of the world.

As the subject is not a part of the world but a presupposition of its existence, so good and evil which are predicates of the subject, are not properties in the world. (*NB*: 79ᵉ)

5.621 The world and life are one.

5.63 I am my world. (*TLP* 151)

6.37 A necessity for one thing to happen because another has happened does not exist. There is only *logical* necessity.

6.373 The world is independent of my will. (*TLP* 181)

6.374 Even if everything we wished were to happen, this would only be, so to speak, a favour of fate, for there is no *logical* connection between will and world, which would guarantee this, and the assumed physical connection itself we could not again will. (*TLP* 181)

6.4 All propositions are of equal value.

6.41 The meaning of the world (*Sinn der Welt*) must lie outside the world. In the world everything is as it is and happens as it does happen. *In* it there is no value--and if there were, it would be of no value.

If there is a value which is of value, it must lie outside all happening and being-so. For all happening and being-so is accidental.

What makes it non-accidental cannot lie *in* the world, for otherwise this would again be accidental.

6.42 It must lie outside the world.

Hence also there are no ethical propositions.

Propositions cannot express anything higher.

6.421 It is clear that ethics cannot be expressed.

Ethics is transcendental.

(Ethics and aesthetics are one.)

6.422 The first thing in setting up an ethical law of the form "thou shalt…" is : And what if I do not do it? But it is clear that ethics has nothing to do with reward and punishment in the ordinary sense. The question as to the *consequences* of an action must therefore be irrelevant. At least these consequences will not be events. For there must be something right in that formulation of the question. There must be some sort of ethical reward and ethical punishment, but this must lie in the action itself.

(And this is clear also that the reward must be something acceptable, and the punishment something unacceptable.)

6.423 Of the will as the subject of the ethical we cannot speak.

And the will as a phenomenon is only of interest to psychology.

6.43 If good or bad willing changes the world, it can only change the limits of the world, not the facts; not the things that can be expressed in language.

In brief, the world must thereby become quite another. It must wax or wane as a whole.

The world of the happy man is a different one from that of the unhappy man.

6.431 So too at death the world does not alter, but comes to an end.

Death is not an event in life: we do not live to experience death.

If we take eternity to mean not infinite temporal duration but timelessness, then eternal life belongs to those who live in the present.

Our life has no limit in just the way in which our visual field has no limits.

6.4312 The temporal immortality of the human soul, that is to say, its eternal survival after death, is not only in no way guaranteed, but this assumption in the first place will not do for us what we always tried to make it do. Is a riddle solved by the fact that I survive forever? Is this eternal life not as enigmatic as our present one? The solution to the riddle of life in space and time lies *outside* space and time.

(It is not problems of natural science which have to be solved.)

6.432 *How* the world is, is completely indifferent for what is higher. God does not reveal himself in what is higher.

6.4321 The facts all belong only to the problem and not to the solution.

6.44 Not *how* the world is, is the mystical, but *that* it is.

6.45 The intuition (*Anschauung*) of the world sub specie aeterni is its intuition as a limited whole.

The feeling of the world as a limited whole is the mystical feeling.

6.5 For an answer which cannot be expressed, the question too cannot be expressed.

The *riddle* does not exist.

If a question can be put at all, then it *can* also be answered.

6.51 Skepticism is *not* irrefutable, but palpably senseless, if it would doubt where a question cannot be asked.

For doubt can exist only where there is a question; a question only where there is an answer, and this only where something *can* be *said*.

6.52 We feel that even when all *possible* scientific questions have been answered, the problems of life remain completely untouched. Of course there are then no questions left, and this itself is the answer.

6.521 The solution to the problem of life is seen in the vanishing of this problem.

(Is this not the reason why those who have found after a long period of doubt that the meaning of life (*Sinn des Lebens*) became clear to them have been unable to say what constituted that sense?)

6.522 There is indeed the inexpressible. This *shows* itself; it is the mystical. (*TLP* 185-187)

For early Wittgenstein, willing (wanting, preferring, choosing, etc.) and feeling (including emotional attitudes) are one and the same. That is one reason why ethics and aesthetics are one. Another reason is that for early Wittgenstein neither aesthetics nor ethics has a propositional or logical component, i.e., neither of them has anything to do with scientific knowing or *Wissen*. Correspondingly, another way of putting this is that the metaphysical subject has two essentially different mental capacities:

(1) a *theoretical* or fact-representing-based, thinking, and logical capacity,

and

(2) a *practical* or feeling-based, willing, and ethical capacity.

But although these two capacities are exercised with respect to the same set of objects (the world of facts, or life), their contents are wholly divergent.

Moreover, the world of facts is modally independent of feeling and willing, and cannot be changed by the will: "mental causation"—whereby a mental event is a

sufficient cause of some physical event—is impossible, because all connections between facts in space and time are either logically contingent or logically necessary. But all value, all good and evil, inheres in the will of the metaphysical subject. This means that early Wittgenstein is positing a radically sharp *theoretical world* vs. *practical world* dichotomy: the world as represented through propositions and natural science is wholly factual and logically-governed, but without any value, whereas the will has fundamental value, yet the value properties of the will are not properties that can be represented propositionally, because although my will is always directed towards my own life, which (given solipsism/LSTI) is the same as my world, those value properties attach only to the metaphysical subject, which is not a part of the world but rather is a necessary presupposition of the world's existence and specific character, and belongs only to *the logical and ethical transcendental structure* of the world.

This radical theoretical world vs. practical world dichotomy has two crucial consequences.

The *first* is that natural science (the totality of contingent truths about the world of facts) and logic are absolutely value-neutral. So even if the world were to be completely described and all of its logical truths made manifest, the problem of the *value* of life (and in particular the value of *my* life, hence "the meaning of life") would not have been touched. This problem of *the value and meaning of my life*, which is the basic problem of aesthetics and ethics, consists precisely in how the subject is to be good or happy, and science has nothing to do with it, at least not directly.

Second, the ethical problem, or the problem of the value and meaning of my life, i.e., how to be good or happy, is radically unlike any scientific problem that can be

propositionally formulated and then (at least in principle) solved. Indeed, early Wittgenstein is saying precisely that the fundamental barrier to solving the problem of the value and meaning of my life is to treat the issue of my goodness or my happiness *as if it were sort of scientific problem to which factual answers could be given*. Only when I am able to realize that the problem of life is *not* a problem in the factual or scientific-logical sense, and that there simply is no such problem of life in this sense, can my will be radically converted into a bearer of goodness and happiness.

According to Wittgenstein, how does this conversion happen? There are two parts to this.

First, we must realize that the theoretical world of scientific facts, and its a priori essence, logic, are in themselves valueless. But second, because we cannot change or in any way affect the facts *in* the world, we must instead change our volitional stance *towards* the world as a whole. This in turn can determine a radically different world in a transcendental structural sense.

On the metaphysical side, early Wittgenstein is saying here that the willing subject can jointly re-constitute the objects and its own language, and thus bring about the existence of a distinct world of facts, which again cannot themselves be changed or affected by our will. This is my will conceived as the "world-will," or what I called Wittgenstein's "solipsism/LSTI of the willing subject."

But on the first-personal side, early Wittgenstein is also saying that to change the world and my own life is not to change any facts whatsoever but rather fundamentally to change the internal configuration of my will so that it becomes internally coherent or harmonious (goodness, happiness) rather than internally incoherent or discordant

(badness, unhappiness). Or in other words, to change the world and my own life is not to change any facts in the world whatsoever, but rather to carry out a radical personal conversion to some essentially new set of ethical values or commitments, which in turn changes the transcendental structure of the world, comparable to that described by Augustine in the *Confessions*, by Pascal in the *Pensées*, by Spinoza in the *Ethics*, or by Kierkegaard in "Purity of Heart is to Will One Thing."[28]

Moreover, when Wittgenstein says that "ethics cannot be expressed," "ethics is transcendental," "ethics and aesthetics are one," and that "of the will as the subject of ethics we cannot speak," I think he is asserting essentially the same thing that Kant is asserting at the end of the *Critique of Practical Reason*:

> Two things fill the mind with ever new and increasing admiration and reverence, the more often and more steadily one reflects on them: *the starry heavens above me* [i.e., nature] *and the moral law within me* [i.e., freedom]. I do not need to search for them and merely conjecture them as though they were veiled in obscurity or in the transcendent region beyond my horizon; I see them before me and connect them immediately with the consciousness of my existence. (*CPrR* 5: 161-162, underlining added)

In other words, what Kant and early Wittgenstein are both asserting, in a shared Transcendental Existentialist spirit, is that is my sense of myself as an intentional and moral agent is *an indispensably necessary and affectively salient phenomenal character* of "the consciousness of my own existence." What early Wittgenstein calls "the world of the happy," and so very sharply contrasts with "the world of the unhappy," is then, for all intents and purposes, a volitionally-oriented and freedom-oriented version of the moral-psychological phenomenon that the Existentialists call *authenticity*. Correspondingly but negatively, what early Wittgenstein calls "the world of the unhappy," and so very sharply contrasts with "the world of the happy," is then, for all intents and purposes, a volitionally-oriented and freedom-oriented version of the moral-psychological phenomenon that the Existentialists call *inauthenticity*.

Strikingly, the moral-psychological phenomenon of inauthenticity also appears in Kant's writings, in at least three slightly different guises.

The first is the almost shockingly stark picture of the person who dogmatically and slavishly accepts the precepts of some existing philosophical system such as the Wolffian philosophy:

> He has formed himself according to an alien reason, but the faculty of imitation is not that of generation, i.e., the cognition did not arise from reason in him, and although objectively it was certainly a rational cognition, subjectively it is still merely historical. He has grasped and preserved well, i.e., he has learned, and is a plaster cast of a living human being. Rational cognitions that are objectively so (i.e., could have arisen originally only out of the reason of human beings themselves) may also bear this name subjectively only if they have been drawn out of the universal sources of reason, from which critique, indeed even the rejection of what has been learned, can also arise, i.e., from principles. (*CPR* A836-837/B864-865, underlining added)

The second is the equally stark picture of the essentially immature and cowardly person who refuses to acknowledge the fundamental moral idea behind "enlightenment" or *Aufklärung*, which is to think for yourself with resolution and courage:

> Enlightenment is the human being's emergence from his self-inflicted immaturity. Immaturity is the inability to use one's own understanding without the guidance of another. This immaturity is self-inflicted if its cause is not lack of understanding, but lack of resolution and courage to use it without the guidance of another. The motto of enlightenment is therefore: *Sapere aude!* Have the courage to use your own understanding! …. [O]nce the germ on which nature has lavished most care—the human being's inclination and vocation to think freely—has developed within its hard shell, it gradually reacts upon the mentality of the people, who thus gradually become increasingly able to *act freely*. Eventually, it even influences the principles of governments, which find that they can themselves profit by treating the human being, who is *more than a machine*, in a manner appropriate to his dignity. (*WiE* 8: 35 and 41-42, underlining added)

And the third is the perhaps even more stark picture of the person who hides from himself the self-defining fact of his own "radical evil" (*radicale Böse*) by pretending that evil is nothing but *bad historical consequences of human activity*, and not the direct result of our deep or transcendental freedom of the will, or in effect, the direct result of what Augustine so aptly called *the perversity of the will*, our transcendentally free ability to do the morally wrong thing *just because* it is the morally wrong thing:

> This dishonesty (*Unredlichkeit*), by which we throw dust in our own eyes and which hinders the establishment in us of an authentic moral disposition (*ächter moralischer Gesinnung*), then

extends itself also externally, to falsity or deception of others. And if this dishonesty is not to be called malice, it nonetheless deserves at least the name of unworthiness. It rest on the radical evil of human nature which (inasmuch as it puts out of tune the moral ability to judge what to think of a human being, and renders any imputability uncertain, whether internal or external) constitutes the foul stain of our species—and so long as we do not remove it, hinders the germ of good from developing as it otherwise would. A member of the English Parliament exclaimed in the heat of debate: "Every man has his price, for which he sells himself." If this is true (and everyone can decide for himself), if nowhere is a virtue which no level of temptation can overthrow, if whether the good or evil spirit wins us over only depends on which bids the most and affords the promptest pay-off, then, what the Apostle says might indeed hold true of human beings universally, "There is no distinction here, they are all under sin—there is none righteous (in the spirit of the law), no, not one." (*Rel* 6:38-39, underlining added)

Human practical reason is our living, spontaneous capacity to exercise the power of choice for the sake of instrumental or non-instrumental principles. In these ways, inauthenticity in the Kantian sense is just to comport yourself *as if you were nothing but a machine*, wholly determined by natural causal laws, and neither alive nor practically free. Or in other words, inauthenticity in the Kantian sense is the *self-automating* denial of your own capacity for practical freedom:

> [I]f the freedom of our will were nothing else than [an *automaton spirituale* when it is impelled by representations], i.e., psychological and comparative and not at the same time transcendental or absolute, it would in essence be no better than the freedom of a turnspit, which when once wound up also carries its motions from itself. (*CPrR* 5: 97, underlining added)

In relation to our capacities for transcendental and practical freedom, there are two different and yet also intimately related ways in which someone can fall into self-automating inauthenticity.

The first way is what I call *The Fallacy of The Rock and The Hard Place*. This is the fallacy of philosophically looking backwards towards the past and also forwards towards the future in ways that self-stultifyingly deny the actual existence of one's own inherent teleology as a rational human intentional agent whose innermost life is aimed at

> (i) the non-denumerable absolutely intrinsic value of rational animals or finite persons, which inheres in the capacity for autonomy,

> (ii) the perfection of this fundamental value, the highest or supreme good, the good will, i.e., choosing and doing the morally right thing for the morally right reasons (the Categorical Imperative),

and finally also at

> (iii) the full realization of the good will, which is the sole, whole, and complete good, namely the hylomorphic fusion of the good will and happiness.[29]

Leaving out the inherently teleological character of one's own rational human innermost life, it can seem on the *one* hand, that the past is completely filled with deterministic and impersonal causes which simply flow through one, shut down all genuine possibilities for choice, and thereby make authentic intentional agency at any present moment impossible (The Rock, a.k.a. "Source Incompatibilism with respect to Determinism") and, on the *other* hand, that the future is randomly indeterministic and heart-breakingly completely filled with possibilities for bad luck and disaster, which thereby make any sort of authentic agential causal efficacy impossible (The Hard Place, a.k.a. "Source Incompatibilism with respect to Indeterminism"). In so doing, one simply *loses heart*, and then in effect tragically *dies* as an authentic intentional and moral agent, even if neurobiological and psychological life continues on. In this sense, the ground of all bad faith and inauthenticity in the Transcendental Existentialist sense is just *the seemingly "scientific" and therefore seemingly "Enlightened" belief that Natural Mechanism, which is the disjunctive combination of Universal Natural Determinism and Universal Natural Indeterminism, i.e., the thesis that every natural event is either determined or indeterministic, is really true*.

The right and authentic Transcendental Existentialist response to The Fallacy of The Rock and The Hard Place is what Kant calls "believing-in," "moral certainty," "faith," or *Glaube*, and what early Wittgenstein also calls "the mystical," as specifically applied to the chain of past events and future events. It is, more specifically, a *teleological believing-in, moral certainty, faith, Glaube, or mystical intuition*

*(Anschauung) in the physico-theological sense.*You morally must resolutely choose and act as if, counterfactually, you believe that the world is designed for us by an all-knowing, all-powerful, and all-good God, *and also that all* your choices and acts in that world are really and truly *up-to-you* and therefore really and truly *free choices for which you are ultimately responsible*, even though you scientifically know, via radical agnosticism, that the existence or non-existence of such a God is both scientifically unknowable and uncognizable. In short, you morally must resolutely choose and act in such a way that you thereby convert the world in which you choose and act into *the world of the happy.*

In this way, believing-in, moral certainty, faith, *Glaube* in Kant's sense, or mystical intuition or *Anschauung* in early Wittgenstein's sense, is when we finally stop theoretically generating all those scientific words and thoughts, achieve some degree of purity of heart, freely will the right thing, and *thereby* enter the world of the good and the happy. Or in other words, it is when we undergo a radical personal conversion, and fundamentally *change our lives for the better*. This is also what Kant calls a *good will*:

> There is no possibility of thinking of anything at all in the world, or even out of it, which can be regarded as good without qualification, except a *good will*….A good will is good not because of what it effects or accomplishes, not because of its fitness to attain some proposed end; it is good only through its willing, i.e., it is good in itself. (*GMM* 4: 393-394)

The second way to fall into freedom-inauthenticity is what I call *Smerdyakov's Fallacy*. It is a standard strategy for critics of Universal Natural Determinism, whether intentionally or not, to confuse Universal Natural Determinism with Fatalism, whether "ordinary" Fatalism or ultra-Fatalism. For example, if someone sincerely says

> "If everything is naturally determined, then whatever *has* happened, was *strictly fated* to happen, and whatever *will* happen, *strictly must* happen, no matter what I choose or do,"

then he is confusing Universal Natural Determinism with Fatalism.

It is equally crucial to distinguish Universal Natural Determinism from another stronger doctrine which says that nature is initially created and also sustained at every later moment by the irresistible causal powers of an all-knowing and all-good deity. This stronger doctrine is *Universal Divine Determinism,* a.k.a. "Theological Determinism". While Universal Divine Determinism is both consistent with Universal Natural Determinism and indeed *entails* Universal Natural Determinism as a trivial consequence, nevertheless Universal Natural Determinism does *not* entail Universal Divine Determinism. Even if an all-powerful, all-knowing, all-good, world-creating, and world-sustaining deity does *not* exist, Universal Natural Determinism can still be true.

In this connection, and corresponding to the fallacy of confusing Universal Natural Determinism with Fatalism, there is an important two-part fallacy that consists in confusing Universal Natural Determinism with Theological Determinism, and then unsoundly inferring universal anarchy from the denial of Theological Determinism, i.e., Smerdyakov's Fallacy:

"If God is dead, then everything is permitted."

Smerdyakov's Fallacy is of course so-dubbed because of the famous passage in Fyodor Dostoevsky's *The Brothers Karamozov* that I included as the fourth epigraph of this section:

> "Take that money away with you, sir," Smerdyakov said with a sigh.
> "Of course, I'll take it! But why are you giving it to me if you committed a murder to get it?" Ivan asked, looking at him with intense surprise.
> "I don't want it at all," Smerdyakov said in a shaking voice, with a wave of the hand. "I did have an idea of starting a new life in Moscow, but that was just a dream, sir, and mostly because 'everything is permitted'. This you did teach me, sir, for you talked to me a lot about such things: for if there's no everlasting God, there's no such thing as virtue, and there's no need of it at all. Yes, sir, you were right about that. That's the way I reasoned."

We also know, as Wittgenstein's biographer Ray Monk tells us, that

> [In 1916] Wittgenstein read [Dostoevsky's *The Brothers Karamazov*] so often he knew whole passages of it by heart, particularly the speeches of the elder Zossima, who represented for him a powerful Christian ideal, a holy man who could 'see directly into the souls of other people'. [30]

So here is the crucial point. From the standpoint of Kant's and early Wittgenstein's

Transcendental Existentalism, the *moral significance* of someone's sincerely asserting

> "If everything is naturally determined, then whatever *has* happened, was *strictly fated* to happen, and whatever *will* happen, *strictly must* happen, no matter I choose or do,"

and Smerdyakov's Fallacy are exactly the same. He has thereby given himself a license

to choose and do whatever he feels like choosing and doing, without any regard for non-

self-interested, non-selfish, non-hedonic, and non-consequentialist moral principles, and

constrained only by natural causal laws. He thereby comports himself as if he were

nothing but a fleshy deterministic or indeterministic Turing-machine, running a decision-

theoretic program for satisfying self-interested, selfish, hedonic, or consequentialist

desires, and therefore *not morally responsible*. This sort of highly self-deceived and

highly self-serving reasoning—ironically and tragically enough, only a really and truly

free agent could ever engage in this sort of duplicitous reasoning—is the quintessence of

freedom-inauthenticity and *bad-faith-about-freedom* in the Kantian sense. And as

Wittgenstein so insightfully points out, this freedom-inauthenticity and bad-faith-about-

freedom is indeed *itself* a kind of "faith," believing-in, or *Glaube*, and not truly a

scientific belief of any kind:

> Life is like a path along a mountain ridge; to right and left are slippery slopes down which you slide without being able to stop yourself, in one direction or the other. I keep seeing people slip like this and I say "How could a man help himself in such a situation!". And *that* is what "denying free will" comes to. That is the attitude expressed in this 'belief' (*'Glauben'*). But is it not a *scientific* belief and has nothing to do with scientific convictions. (*CV*: 63e)

By your invoking the seemingly "scientific" and therefore seemingly "Enlightened"

belief in either Universal Naturalism Determinism, Natural Mechanism, or Fatalism, then

like poor benighted Smerdyakov, as philosophically programmed by Ivan Karamazov,

thereby "denying free will," you not only refuse *to hold others responsible*, but also much more fundamentally you thereby deny your own capacity *to take responsibility and to change your life*. This is *not* Enlightenment in the Kantian and Wittgensteinian sense—indeed, it is precisely *the opposite* of Enlightenment in that sense. It is, in fact, nothing but the quintessence of what Kierkegaard so aptly calls *double-mindedness*.[31]

III.6 Why *Not Everything* is Permitted

I am now at last in a position to re-raise the quasi-Dostoevskian question posed in the title of this section: If God's existence is unprovable, then is everything permitted, i.e., is human morality really impossible? The Kantian and Wittgensteinian Transcendental Existentialist answer I am offering is: *No*, and in fact the truth of the matter is precisely the other way around. *Only if* God's existence or non-existence is logically unprovable and scientifically unknowable, is rational human morality really possible; *only if* God's existence or non-existence is logically unprovable and scientifically unknowable, will we be able to face up to The Problem of Evil adequately; and *only if* God's existence or non-existence is logically unprovable and scientifically unknowable, can life have a moral meaning either generally or first-personally. Transcendental Existentialism and its radical agnosticism tell us that a morally meaningful human life *begins* at the inherent limits of classical logic and natural science, and that it also *presupposes* those limits. You cannot rationally be *either* a theist/deist *or* an atheist. The logic of moral life is *deeply* non-classical, Intuitionistic, and constructivist. The classical logical facts and the natural facts, real as they are, are not *all* the facts about rational human life. Beyond those facts are the *further* facts of the Ethical, the Mystical, and the world of the happy, which *can* be constructed, but *only* by you and

by you *alone*. In that sense, all these further non-classical-logical and non-natural facts about rational human life are equally *up-to-you* and also *down-to-you*.

This Transcendental Existentialist doctrine is therefore *neither* theism/deism *nor* atheism—on the contrary, it is the doctrine that there ought to be and therefore morally must be, for each and every one of us who is capable of seriously considering these matters, a rational, freely chosen, and entirely wholehearted step-by-step non-classical-logical, Intuitionist, and constructivist *personal transformation* from the logical unprovability and scientific unknowability of God's existence or non-existence, to moral authenticity:

> One of the things Christianity says, I believe (*glaube ich*), is that sound doctrines are all useless. One must change one's *life*. (Or the *direction* of one's life.) (*CV*: 53ᵉ)[32]

IV. How and Why Logic is Transcendental

> [The logic of the general use of the understanding] contains the absolutely necessary rules of thinking, without which no use of the understanding takes place, and it therefore concerns these rules without regard to the difference of the objects to which it may be directed…. Now general logic is either pure or applied logic. In the former we abstract from all empirical conditions under which our understanding is exercised…. A general but pure logic therefore has to do with strictly *a priori* principles, and is a *canon of the understanding* and reason, but only in regard to what is formal in their use, be the content what it may…. A general logic, however, is called applied if it is directed to the rules of the use of the understanding under the subjective empirical conditions that psychology teaches us…. In general logic the part that is to constitute the pure doctrine of reason must therefore be entirely separated from that which constitutes applied (though still general) logic. The former alone is properly science…. In this therefore logicians must always have two rules in view. 1) As general logic it abstracts from all contents of the cognition of the understanding and of the difference of its objects, and has to do with nothing but the mere form of thinking. 2) As pure logic it has no empirical principles, and thus draws nothing from psychology …. It is a proven doctrine, and everything in it must be completely *a priori*. (*CPR* A52-54/B76-78)

> 6.13 Logic is not a theory but a reflexion of the world. Logic is transcendental. (*TLP* 169)

IV.1 The L-is-T Thesis

Both Kant and early Wittgenstein held the perhaps surprising thesis *that logic is transcendental*. I will call this *The L-is-T Thesis*. The L-is-T Thesis says that

> logic is objectively necessarily true, a priori, and also *transcendentally explains* (a.k.a. "is the condition of the possibility of") all rational human cognition and thought.

Here, in turn, is the relevant notion of a *Transcendental Explanation*, via the preliminary notion of a *Transcendental Argument*.

An *argument* is a set of sentences or statements Γ (and possibly Γ = the null set of sentences or statements), i.e., the premises, such that a sentence or statement *S* (which may or may not be a member of Γ), i.e., the conclusion, is held to follow validly or soundly from Γ. Then an argument is a Transcendental Argument if and only if

> (i) Some version of Transcendental Idealism, whether Strong Transcendental Idealism (STI) or Weak or Counterfactual Transcendental Idealism (WCTI), is assumed to be true.

and

> (ii) That argument advances from a sentence or statement *S*, taken as a single premise, to an a priori necessary presupposition *APNP* of *S*—i.e., "*a* condition of the possibility" of *S*—taken as a single conclusion, as follows:

(1) S
(2) S presupposes $APNP$.
(3) Therefore, $APNP$.

For example, let S = "There are 7 martinis sitting on the kitchen table" and let

$APNP$ = "3+4=7."

(1) There are 7 martinis sitting on the kitchen table.

(2) The sentence or statement that there are 7 martinis sitting on the kitchen table presupposes the a priori necessary truth that 3+4=7. For if it were *not* the case that 3+4=7, that is, if it were *not* the case that the primitive recursive functions over the natural numbers, like addition, hold, then it would be *neither* true that there are 7 martinis sitting on the kitchen table *nor* false that there are 7 martinis sitting on the kitchen table.

(3) Therefore, 3+4=7. (From (1) and (2).)

An $APNP$ can be either analytic a priori (indeed, trivially, every analytic truth is

presupposed by every sentence or statement whatsoever) or synthetic a priori,

but in either case it is known by basic authoritative philosophical rational intuition.[33]

In turn, an *explanation* is a set of sentences or statements Γ (and Γ cannot be the

null set of statements) and another sentence or statement S (which cannot be a member of

Γ, on pain of circularity), such that some sort of necessitation relation is held to obtain

between Γ and S, i.e.,

$\sim (\Gamma \rightarrow S]$

Then an explanation is a Transcendental Explanation if and only if there is an a priori

necessary presupposition $APNP$ of a sentence or statement S such that $APNP$, when taken

together with some or another set of true general and specific claims (C1, C2, C3 ….Cn)

derived from natural science and/or Wide Reflective Equilibrium in the Rawlsian sense,

is also related to S in the following way:

Syn Ap $\sim [\{APNP \& (C1, C2, C3 …Cn)\} \sim \rightarrow S]$

or in other words,

Synthetically a priori necessarily, if *APNP* and also some or another set of general and specific claims (C1, C2, C3 ….Cn) derived from natural science and/or Wide Reflective Equilibrium in the Rawlsian sense all *were to be* true, then *S would be* true.

Thus a sound Transcendental Explanation demonstrates an synthetic a priori subjunctive conditional relation between an *APNP*, which is known by basic authoritative philosophical rational intuition, and an *S*, which is known by any other reliable method of knowledge, via some body of fundamental knowledge claims provided by natural science and/or Wide Reflective Equilibrium in the Rawlsian sense. Otherwise put, a sound Transcendental Explanation demonstrates that *APNP* is "*the* condition of the possibility" of *S*.

For example, let *S* = "There are 7 martinis sitting on the kitchen table," let *ANPP* = "3+4=7," and let "(C1, C2, C3 … Cn)" be a set of relevant general and specific claims taken from natural science and Wide Reflective Equilibrium in the Rawlsian sense, about martinis, tables, their causal-dynamic relations, and the nature of the sitting-on relation. Then the following is a sound Transcendental Explanation:

(1) There are 7 martinis sitting on the kitchen table.

(2) Synthetically a priori necessarily, if "3+4=7" and also a set of relevant general and specific claims about martinis, tables, their causal-dynamic relations, and the nature of the sitting-on relation, all *were to be true*, then it *would* be true that there are 7 martinis sitting on the kitchen table.

Now Kant held The L-is-T Thesis because he held that pure general logic is the strictly universal and a priori science of the laws of thought. Early Wittgenstein, by a significant contrast, held The L-is-T Thesis because he held that the classical second-order logic of Frege's *Begriffsschrift*, and Russell and Whitehead's *Principia Mathematica*, is built into the very nature of my language and also into the very nature of the world my language represents.

I fully agree with Kant and early Wittgenstein that The L-is-T Thesis is true. But two things about the The L-is-T Thesis are quite obscure in Kant's and early Wittgenstein's writings in philosophical logic:

(1) Precisely which argument, or arguments, can adequately justify The L-is-T Thesis?

(2) Precisely what are the basic implications of The L-is-T Thesis?

In this section, first, I will present five Kantian arguments for The L-is-T Thesis, and also spell out their basic implications. Then, second, I will unpack both early and later Wittgenstein's views on the nature of logic.

IV. 2 Kantian Argument 1: First-Order Monadic Logic and Pure General Logic are both Transcendental

The first argument is intended to show that both first-order monadic logic and pure general logic are, in addition to being objectively necessarily true, also a priori necessary presuppositions (*APNP*s) of all rational human cognition and thought, hence "transcendental" in the sense I specified at the beginning of the section.

It is both relevant and important to note that as early as C.I. Lewis's seminal 1918 book *Survey of Symbolic Logic,* there was a fundamental distinction in the 20th century logical tradition between

(i) *formal* or *symbolic* logic, which is essentially a rigorous development of Kant's notion of pure general logic,

and

(ii) what Russell aptly called *mathematical* logic, which is *second-order* because it includes whatever logical or semantic machinery is needed to quantify over and talk about functions, predicates, and relations, and also other characteristically mathematical furniture like sets, numbers, and spaces.[34]

The reason that this distinction is philosophically important is that for Kant, it is also possible to have a pure or completely a priori logic that is topic *specific,* or systematically

sensitive to special ontological commitments, which is what he calls *transcendental* logic (*CPR* A62/B87). Strikingly, early Wittgenstein seems to have had, in effect, the very same idea about transcendental logic in the *Tractatus*, as we saw in section II and again in this section's second epigraph:

> Logic is not a theory but a reflexion of the world. Logic is transcendental.

In this way, *mathematical logic* in Russell's sense would count as a transcendental logic for both Kant and the Tractarian Wittgenstein.

Transcendental logic in Kant's sense, however, also inherently contains *necessarily true synthetic a priori statements*, which would not have been allowed by Wittgenstein in his Tractarian period. Nevertheless, from a Kantian standpoint, it seems that if early Wittgenstein had admitted necessarily true synthetic a priori statements into *his* transcendental logic, then this would have made it possible for him to provide a coherent account for the logico-semantic status of the infamous Two Colours Proposition, or The TCP. Here is what early Wittgenstein says explicitly about The TCP in the *Tractatus*:

> For two colours … to be at one place in the visual field, is impossible, logically impossible, for it is excluded by the logical structure of colour.[35]

In this way, early Wittgenstein regards The TCP—i.e., "For two colours … to be at one place in the visual field, is impossible"—as a logical truth of elementary logic. But this forces him into the dilemma of *either* giving up the logical independence of atomic propositions—e.g., the logical independence of the atomic propositions

> (R) Point *P* in visual space is red all over.

and

> (G) Point *P* in visual space is green all over.

—*or else* devising some analysis of propositions like (R) and (G) which smoothly

converts them and all their analogues into complex or molecular propositions, in order to

be able to assert that that the obvious mutual exclusion relation between (R) and (G) is a

purely logical relation. But for early Wittgenstein, facing up to this dilemma also means

giving up certain basic commitments of his account of the nature of logic and logical

analysis in the *Tractatus*—in particular, giving up the mutual logical independence of

atomic propositions and correspondingly, giving up the mutual logical independence of

atomic facts—which is precisely what post-Tractarian Wittgenstein more or less

explicitly does in 1929 in "Some Remarks on Logical Form," by claiming that atomic

propositions *can* be mutually logically contradictory,[36] and then by later observing to

Waismann that this 1929 move, in turn, leads to absurdity:

> Now suppose the statement "An object cannot be both red and green" were a synthetic judgment
> and the words 'can not' meant logical impossibility. Since a proposition is the negation of its
> negation, there must also exist the proposition, "An object can be red and green." This proposition
> must also be synthetic. As a synthetic proposition it has sense, and this means that the state of
> things represented by it *can obtain*. If 'can not' means *logical* impossibility, we therefore reach the
> consequence that the impossible *is* possible.[37]

From a Kantian standpoint, however, it seems to me obvious that the correct way

out of this dilemma is to allow for two *essentially different* kinds of necessity, namely,

> (1) analytic or logical a priori necessity, i.e., the necessity which flows from the nature of
> concepts (a.k.a. "conceptual necessity," or "weak metaphysical a priori necessity"),

and

> (2) synthetic or non-logical a priori necessity, i.e., the necessity which flows from the
> nature of things in the world (a.k.a. "essentially non-conceptual necessity," or "strong
> metaphysical a priori necessity"),

which is the same as to hold the thesis of *Modal Dualism*. Given Modal Dualism, and

given the fact that impossibility is definable in terms of necessity and negation, one can

coherently hold that (R) and (G) are logically independent propositions and yet also non-

logically mutually exclusive propositions, by holding that the mutual exclusion relation between them is one of *synthetic or non-logical a priori impossibility*, not analytic or logical a priori impossibility.

In any case, as I have mentioned already, Kant holds that the truths of arithmetic and geometry are synthetic a priori, not analytic. One reason he does so is because he thinks that the representational content of mathematics rests on logic *plus* our a priori representations of the formal structures of asymmetrically-directional time (for the purposes of representing Primitive Recursive Arithmetic and its conservative extensions, including Peano Arithmetic[38]) or orientable 3-D Euclidean space (for the purposes of representing Euclidean geometry and its conservative extensions, including classical Non-Euclidean geometry[39]). But another, and ultimately equivalent, way of expressing the synthetic apriority of arithmetic and geometry is to point out that the logic which represents them must contain irreducibly *relational* predicates whose satisfaction conditions require the existence of at least one object in the actual world (e.g., in the case of identity) or otherwise the existence of at least two objects in the actual world, and in some cases (e.g., the case of the relational predicates needed to represent the standard Peano axioms for arithmetic) the existence in the actual world of at least a denumerably infinite number of objects. Thus all the logical truths of the first-order, inherently polyadic, and multiply-quantified part of Frege's logic—i.e., classical first-order predicate logic with identity—in Kantian (or at least, contemporary neo-Kantian) terms, are *synthetic a priori*, not analytic.

Frege's logic includes set theory, as well as an axiom that allows for the unrestricted formation of sets, and of course it leads directly to Russell's Paradox about

the logically explosive (a.k.a. "impredicative") status of the set *K* of all sets that are not

members of themselves, whose existence yields the unhappy paradoxical result that *K* is a

member of itself if and only if it is not a member of itself. Russell's mathematical logic

includes a principle—*The Vicious Circle Principle*—which stipulatively rules out the

impredicativity that afflicts unconstrained iterative set theory.[40] But Russell's

mathematical logic also includes something called *The Axiom of Infinity*, which posits the

existence of at least a denumerably infinite number of objects in the domain of discourse,

and which is arguably not a purely logical principle. Moreover, and in any case, Russell's

mathematical logic still threatens to allow for paradoxical impredicativity with respect to

functions, predicates, and relations, even if it stipulatively rules out impredicative sets,

unless one makes a further empirical and clearly non-logical assumption Russell calls *The*

Axiom of Reducibility.[41]

In other words, the crucial issue here is whether the rational core of classical logic

should be taken to be second-order logic in either the Fregean or Russellian sense, or

instead is *elementary logic*: i.e., bivalent first-order polyadic predicate calculus with

identity.[42]

Tarski, e.g., both emphatically and explicitly supported the thesis that elementary

logic, not second order logic, is the core classical logic:

> The terms 'logic' and 'logical' are used [by most contemporary logicians] in a broad sense, which
> has become almost traditional in the last decades; logic is here assumed to comprehend the whole
> theory of classes and relations (i.e., the mathematical theory of sets). For many different reasons I
> am personally inclined to use the term 'logic' in a much narrower sense, so as to apply it only to
> what is sometimes called "elementary logic," i.e., to the sentential calculus and the (restricted)
> predicate calculus.[43]

But even *elementary logic* contains some arguably non-logical factors. For example,

since

(1) a=a

is an instance of the law of identity and can be introduced into any line of a proof as a theorem of logic, and thus as depending on the empty set of premises, it follows immediately that

(1) $(\Box x)\, x = x$

which says *that something exists*, is *also* a theorem of logic, which seems highly implausible. Why couldn't there be logically possible worlds that with no individual objects in them (i.e., the empty domain of discourse); and furthermore, why couldn't there be logically possible worlds in which *nothing whatsoever* exists?[44]

Quine, significantly, holds that identity is indeed *part* of the rational core of classical logic, yet also *excludes* set theory from this core:

> The upshot is, I feel, that identity theory has stronger affinities with its neighbors in logic than with its neighbors in mathematics. It belongs in logic.

> We turn now from identity to set theory. Does it belong in logic? I shall conclude not.[45]

By sharp contrast, for Kantians, both Frege's logic and also Russell's mathematical logic, and indeed *any* logic that is an inherently relational or polyadic logic and also includes identity, hence elementary logic, and also any logic that includes set theory, and any logic that is a second-order logic more generally, will all count as synthetic a priori *transcendental* logics, not pure general logics, precisely because they all include special ontological commitments that take them significantly beyond the scope of pure general logic. To the same effect, in the specific case of set theory, Quine accurately and aptly points up the significant philosophical advantages of Kant's pure general logic over Frege's logic:

> Altogether, the contrasts between elementary logic and set theory are so fundamental that one might well limit the word 'logic' to the former… and speak of set theory as mathematics in a sense exclusive of logic. To adopt this course is merely to deprive 'ε' of the status of a logical word. Frege's derivation of arithmetic would then cease to count as a derivation from logic; for he used set theory. At any rate we should be prepared to find that [Carnap's] linguistic doctrine of

logical truths holds for elementary logic and fails for set theory, or vice versa. Kant's readiness to see logic as analytic and arithmetic as synthetic, in particular, is not superseded by Frege's work (as Frege supposed), if "logic" be taken as elementary logic. And for Kant logic certainly did not include set theory. [46]

And basically the very same points could be made for the comparison and contrast between Kant's logic and Russell's mathematical logic, just by uniformly substituting 'Russell' for 'Frege' and 'second-order logic' for 'set theory' in that quotation from Quine.

This brings me to the heart of the matter. Kant thinks of *pure general logic* as the core classical logic because it is analytic, a priori, and strictly universal, but also more fundamentally because it is *topic-comprehensive* or *topic-synoptic*, and holds equally for empty domains of discourse, and for worlds with nothing whatsoever in them, as well as for occupied domains, and worlds containing sets, functions, or relations. Now Kant's pure general logic, as it happens, is a *second-order intensional monadic logic*. It is second-order and intensional because it both includes and quantifies over finegrained, decomposable concepts, as well as possible-worlds extensions.[47] By another sharp contrast, Quine's and Tarski's elementary logic is an extensional logic, and not an intensional logic; moreover, elementary logic is also inherently polyadic or relational, and it includes identity. Nevertheless, where Kant's pure general logic and elementary logic *fully overlap* is precisely in *first-order monadic logic*, which is bivalent truth-functional logic together with a restricted predicate logic employing quantification over individuals and into one-place predicates only.[48] In empty domains, or in completely empty possible worlds, *first-order monadic logic collapses to truth-functional logic.*

Therefore, if we zero in on first-order monadic logic and explicitly take into account how it collapses into truth-functional logic in empty domains and empty worlds,

it follows that in first-order monadic logic we have before us an ultra-pasteurized version

of Kant's pure general logic that is also the perfect candidate for being "sheer logic" in

Quine's sense:

> If sheer logic is not conclusive, what is? What higher tribunal could abrogate the logic of truth
> functions or of quantification?[49]

In part, this is because of the following highly significant historical intersection of

doctrines in the philosophy of logic:

> (1) Kant implicitly accepts first-order monadic logic as belonging to the rational core of
> classical logic,
>
> (2) Frege implicitly accepts first-order monadic logic as belonging to the rational core of
> classical logic,
>
> (3) Russell implicitly accepts first-order monadic logic as belonging to the rational core
> of classical logic,
>
> (4) The Tractarian Wittgenstein implicitly accepts first-order monadic logic as belonging
> to the rational core of classical logic,
>
> (5) Tarski implicitly accepts first-order monadic logic as belonging to the rational core of
> classical logic,

and

> (6) Quine implicitly accepts first-order monadic logic as belonging to the rational core of
> classical logic.

Furthermore, as Quine implicitly showed us, first-order monadic logic is also *the*

paradigm of logical analyticity. Therefore first-order monadic logic, as being logic in a

way about which Kant, Frege, Russell, early Wittgenstein, Tarski and Quine could all

fully agree, is *pure general, paradigmatically analytic, core classical, "sheer" logic*.

Indeed, when we realize that it was *precisely* the pure generality, paradigmatic

analyticity, core classicality, and sheerness of first-order monadic logic that Kant

implicitly had in mind when he wrote

> That from the earliest times logic has traveled this secure course [of a science] can be seen from
> the fact that since the time of Aristotle it has not had to go a single step backwards…. What is

further remarkable about logic is that until now it has also been unable to take a single step forward, and therefore seems to all appearances to be finished and complete. (*CPR* Bxviii-xix),

then we can clearly see that Kant's notorious remark was entirely apt, arguably self-evidently true, and precisely the reverse of outrageous.

Following out Kant's deep thoughts about the nature of pure general logic and (implicitly) first-order monadic logic, then, let us call the pure logical properties of truthful consistency, soundness, completeness, decidability, and logical truth or analyticity *The Logical Perfections.* As in standard treatments of contemporary logic, consistency is the property of the formal non-contradictoriness of statements, or alternatively the property of there being at least one interpretation in which all members of a given set of statements are true (a.k.a. the set of statements "has a model"). Soundness is the property such that all provable sentences or theorems in a logical system are logically true or tautologous. Completeness is the property such that all tautologies are theorems, or provable sentences. And decidability is the property such that there is a finite recursive procedure for determining tautologousness. By the perhaps slightly unfamiliar notion of the *truthful consistency* of given logical system Σ, moreover, I specifically mean that:

> (i) Σ never includes arguments that lead from true premises to false conclusions (= truth-preservation),

and

> (ii) Σ never includes contradictions as theorems of logic (= non-dialetheism—i.e., no "truth-value gluts" or "true contradictions" allowed).

We can think of truthful consistency as the *Highest or Supreme Good* of logic, and we can also think of this systemic feature together with all the other Logical Perfections as proper parts of the *Complete Good* of logic.

The Logical Perfections collectively specify the highest logical standards of rational normativity, and in this sense, logic, like Kant's "metaphysics of morals," is *a categorically normative moral science*. But it is also true that each of The Logical Perfections is not independently essential to logic. *Dialetheic paraconsistent* logical systems are possible,[50] in which contradictions can occur as true sentences or statements or even as theorems of logic (= dialetheism), and such systems are thereby not truthfully consistent, provided that the system also contains an axiom that prevents every sentence or statement whatsoever from being entailed by any given contradiction (= paraconsistency), a logical phenomenon that is called "Explosion." For example, arguably both The Liar Sentence (which asserts its own falsity)[51] and The Gödel Sentence (which provably asserts its own unprovability)[52] are true contradictions, and these true contradictions can arguably be allowed into logical systems as true sentences or even theorems, provided that Explosion is ruled out.

Correspondingly, some logical systems are not sound, e.g., dialetheic paraconsistent systems. Some logical systems are sound but not complete, e.g., elementary logic plus the standard Peano axioms for arithmetic. And some logical systems are undecidable, e.g., elementary logic. As Gödel showed, undecidability and indeed also logical unprovability both apply to some individual true statements in any formal system rich enough to contain elementary logic plus (enough of) the standard Peano axioms for arithmetic, and such systems are consistent if and only if they are incomplete and have their ground of truth outside the system. Decidability on its own, however, can also apply to a formal system consisting entirely of what Kant would have

regarded as irreducibly synthetic a priori truths, e.g., the truths of Primitive Recursive Arithmetic or PRA.[53]

More generally, it is only in the context of a logic of *analyticity* that decidability closes the tight systemic circle of all The Logical Perfections. Indeed, when we see that the tight circle of The Logical Perfections can actually be exemplified in at least two logics—i.e., either classical *truth-functional logic* or *first-order monadic logic*, both of which are truthfully consistent, sound, complete, decidable, and analytic—then we realize that each of these logics constitutes an *maximal, ideal, highest, or supreme rational normative standard* of systematicity. This maximal, ideal, highest, or supreme rational normative standard of systematicity, as Kant points out, necessarily guides all rational and scientific inquiry in a *regulative* way. But this ideal must *not* also be regarded as *constitutive* in Kant's sense. For the tragically mistaken thesis that the maximal, ideal, highest, or supreme rational normative standard of systematicity realized by classical truth-functional logic or first-order monadic logic applies to any *other* set of statements or body of knowledge will inevitably lead to fundamental metaphysical errors and insoluble logical paradoxes and puzzles, as the Transcendental Dialectic clearly shows in great detail (*CPR* A293-A704/B349-732). I will come back to this crucial point about logic in particular and philosophy more generally in section V.

In the Introduction to the *Jäsche Logic*, Kant himself uses the term "logical perfections" (*logische Vollkommenheiten*) in essentially the same way I have just used it (*JL* 9: 33-81). But Kant of course did not know about meta-logic. Now since Kant did not know about meta-logic, he also did not know that the first-order monadic logic that is embedded in his pure general logic is truthfully consistent, sound, complete, and

decidable, although he did of course (at least implicitly) know that first-order monadic

logic is analytic a priori, since (again, at least implicitly) he knew that second-order

intensional monadic logic is analytic a priori. Strikingly, and by contrast, classical first-

order predicate logic with polyadic predicates and multiple quantification is truthfully

consistent, sound, and complete, but *not* decidable, and (as we have seen) *not* analytic.

What are we to make of the fact that first-order monadic logic—or logic in a

sense that Kant, Frege, Russell, Tarski, and Quine all implicitly but fully affirm as

belonging to the rational core of classical logic—is provably truthfully consistent, sound,

complete, decidable, *and also* analytic a priori? One possibility is that first-order monadic

logic is the logic *which best captures our most unshakeable and thus authoritative*

"obviousness" intuitions[54] *about logical analyticity in natural language*. Indeed, even

Quine himself implicitly admits this, which can be easily enough seen by recalling his

initial definition of analyticity, adding one minor qualifier to his famous remark about

"sheer logic," and then juxtaposing these two seminal Quinean texts:

> [Analytic statements] fall into two classes. Those of the first class, which may be called *logically true*, are typified by:
>
> (1)   No unmarried man is married.
>
> The relevant feature of this example is that it not merely is true as it stands, but remains true under any and all reinterpretations of 'man' and 'married'. If we suppose a prior inventory of *logical* particles, comprising 'no', 'un-', 'not', 'if', 'then', 'and', etc., then in general a logical truth is a statement which is true and remains true under all reinterpretations of its components other than the logical particles.[55]
>
> If sheer logic is not conclusive, what is? What higher tribunal could abrogate the logic of truth functions or of [monadic] quantification?[56]

Notice here that all analytic statements of the same form as "No unmarried man is

married" involve first-order *monadic* quantification only. And not only the logic of truth

functions but also the logic of first-order monadic quantification each counts as

conclusive, sheer logic. But first-order monadic logic is the logic of truth functions plus

the logic of first-order monadic quantification. So according to Quine, at least implicitly, first-order monadic logic must be the logic which best captures our most unshakeable and thus authoritative "obviousness" intuitions about logical analyticity in natural language.

Now if first-order monadic logic is the logic which best captures our most unshakeable and thus authoritative "obviousness" intuitions about logical analyticity in natural language, then it is arguable that pure general logic, insofar as it inherently contains first-order monadic logic, along with fine-grained, decomposable intensions and possible-worlds extensions, is the *Universal Natural Logic* of human natural languages insofar as it best captures our most unshakeable and thus authoritative "obviousness" intuitions about *all kinds of analyticity* in natural language, just as Chomsky's *Universal Grammar* best captures our most unshakeable and thus authoritative "obviousness" intuitions about *all kinds of grammaticality* in natural languages.[57]

Here we need also to consider a distinct although, ultimately, closely related point. One of the great advances of 20th century logic was the discovery and development of non-classical logics. Non-classical logics are of two distinct kinds:

(i) *extended logics,* which preserve all the tautologies, theorems, inference rules, syntactic rules, and semantic rules of classical logic, but add some new ones,

and

(ii) *deviant logics*, which reject some of the tautologies, theorems, inference rules, syntactic rules, or semantic rules of classical logic, and may also add some new ones.[58]

Extended non-classical logics are *conservative*, while deviant non-classical logics are *radical*. For example, second-order logic and classical modal logic are extended logics, whereas Intuitionist logic (which rejects the universal principle of excluded middle, or PEM) and dialetheic paraconsistent logic (which as I mentioned above, rejects the universal principle of non-contradiction, or PNC, and accepts the existence of "truth-

value gluts" or "true contradictions," provided that it also contains an axiom that it rules out the entailment of every statement whatsoever by any given contradiction, a.k.a. "Explosion") are deviant logics.

Given the distinction between extended and deviant non-classical logics, and assuming the plausibility of my earlier claim that pure general logic, insofar as it inherently contains first-order monadic logic together with finegrained, decomposable intensions and possible-worlds extensions, is the logic which best captures our most unshakeable and thus authoritative "obviousness" intuitions about all kinds of analyticity in natural language, and is arguably the Universal Natural Logic of all natural languages, then I think that we can now also see that pure general logic plausibly arguably captures *the a priori essence of logic*, in the threefold sense that

> (i) Synthetically a priori necessarily, if *anything* counts as a logic, then pure general logic, insofar as it inherently contains first-order monadic logic, will count as a logic.

> (ii) Synthetically a priori necessarily, if *anything* is either an extended or a deviant logic, then it is nothing but either a conservative extension or a deviant of pure general logic, insofar as it inherently contains first-order monadic logic.

and

> (iii) Plausibly arguably, the conservative extension of first-order monadic logic to pure general logic captures the a priori essence of *logical analyticity*, since pure general logic is just second-order intensional monadic logic and plausibly arguably best captures our most unshakeable and thus authoritative "obviousness" intuitions about all kinds of analyticity in natural language.

IV.3 Kantian Argument 2: The Absolute Unrevisability Argument

The second argument is intended to show that at least one logical principle which is fully presupposed by first-order monadic logic and pure general logic alike, is itself absolutely unrevisable, and therefore, in addition to being objectively necessarily true, is

also an a priori necessary presupposition (*APNP*) of all rational human cognition and thought, and thus is also "transcendental" in the sense specified at the beginning of this section. Here is that logical principle in two equivalent alethic versions:

~ (□S) (S & ~ S)

Not every sentence or statement in every language or logical system is both true and false (*The Minimal Logical Meta-Principle of Non-Contradiction*).

The Minimal Logical Meta-Principle of Non-Contradiction also has a deontic version formulated as a *logical categorical imperative*:

You categorically ought to accept as truths in any language or logical system only those sentences or statements which do not entail that it and all other sentences or statements in that language or logical system are both true and false.

This logical categorical imperative version of The Principle, in turn, guarantees what I will call *minimal truthful consistency*. Truthful consistency, as such, means that you must accept as truths in a language or logical system only those sentences or statements which do not entail that *any* argument in that language or system leads from true premises to false conclusions. By contrast, *minimal* truthful consistency means that you must accept as truths in any language or logical system only those sentences ir statements which do not entail that *every* argument in that language or system leads from true premises to false conclusions. This latter notion of course is consistent with holding that *some* arguments in that language or system lead from true premises to false conclusions, and indeed is also consistent with holding that some arguments in the language or system lead from the *null* set of premises to *necessarily false* conclusions. If so, then some sentences or statements in that language or system are both true and false, hence are truth-value gluts or "true contradictions." So minimal truthful consistency is consistent with dialetheic paraconsistency.[59] In other words, then, The Minimal Logical

Meta-Principle of Non-Contradiction essentially secures minimal truthful consistency, and rules out Explosion. It is not a strictly *truth*-preserving logical principle, and not even a strictly *consistency*-preserving logical principle, but it nevertheless strictly rules out *global inconsistency*, i.e., logical anarchy or chaos, which is the ultimate result of Explosion: If every sentene or statement whatsoever follows from a contradiction, then the negation of every sentence or statement whatsoever also follows from a contradiction, and therefore every sentence or statement whatsoever is a truth-value glut or true contradiction.[60]

In the 1980s, Hilary Putnam very plausibly argued that the negative version of this minimal logical meta-principle is the one absolutely indisputable a priori truth:

> I shall consider the weakest possible version of the principle of [non-] contradiction, which I shall call the minimal principle of [non-] contradiction. This is simply the principle that *not every statement is both true and false*... [I]f, indeed, there are no circumstances in which it would be rational to give up our belief that *not every statement is both true and false*, then there is at least one *a priori truth*.[61]

Although the 1980s Putnam apparently held a sharply different theory of apriority from Kant's,[62] nevertheless Kant's and Putnam's accounts do converge perfectly on the *transcendental* logico-semantic status of the statement that $\sim (\Box S) (S \mathrel{\&} \sim S)$, i.e., on the *transcendental* logico-semantic status of The Minimal Logical Meta-Principle of Non-Contradiction.

IV.4 Kantian Argument 3: The Logocentric Predicament Argument

The third argument is intended to show that, if the two first two arguments are sound and if I am correct that first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all "transcendental" in the sense specified at the beginning of this section, then this fact can be used to provide an adquate solution to the very hard philosophical problem of *The Logocentric Predicament*.

So we can then conclude that logic is transcendental *by an inference to the best philosophical explanation.*

The Logocentric Predicament is this: How can logic ever be justified or explained, if logic must be presupposed and used in order to justify or explain logic? This problem is essentially the same as the one that the Harvard logician Harry Sheffer—known best for his discovery of the Sheffer stroke function—called "the logocentric predicament" in a 1926 review of the second edition of *Principia Mathematica*:

> The attempt to formulate the foundations of logic is rendered arduous by a … "logocentric" predicament. In order to give an account of logic, we must presuppose and employ logic.[63]

In 1895 Lewis Carroll had pointed up a closely related worry in "What the Tortoise Said to Achilles," by arguing that the attempt to generate the total list of premises required to validly deduce the conclusion of an argument leads to a vicious regress.[64] Carroll's argument was resuscitated in 1936 by Quine in "Truth by Convention," where he pointed out that the attempt to define logical (or analytic) truth on the basis of syntactic meta-logical conventions alone is viciously circular in a Tortoise-like fashion, because pre-conventional logic is already required to generate the truths from the conventions.[65] And in 1976 Susan Haack raised what is in effect the same worry, but this time in the form of a worry about the very idea of a justification of logical deduction, by arguing as follows:

> (1) All justification is either non-deductive (e.g., inductive) or deductive.

> (2) On the one hand a non-deductive justification of deduction is too weak and on the other hand a deductive justification of deduction is circular.

> (3) Therefore, deduction cannot be justified.[66]

Philosophers of logic have attempted various solutions to The Logocentric Predicament, the Tortoise regress problem, and the problem of justifying deduction. I will not canvass these attempts here, although I do cover them and critically analyze them in

detail in another place.[67] My intention here is just to suggest how we could use the notions of first-order monadic logic and pure general logic to solve The Logocentric Predicament and its associated problems. Suppose that pure general logic really does capture the a priori essence of logic *just because*, insofar as it contains first-order monadic logic, and also falls under The Minimal Logical Meta-Principle of Non-Contradiction, it thereby adequately captures all The Logical Perfections—truthful consistency, soundness, completeness, decidability, and above all, analyticity—and it is also The Universal Natural Logic. Then since all rational theorizing, explanation, and justification whatsoever presuppose logic, it follows that pure general logic must also be the a priori essence of all rational theorizing, explanation, and justification whatsoever.

More explicitly, it solves The Logocentric Predicament by showing us that pure general logic is the explanatory and justificatory *unique categorically normative theoretical primitive*. Pure general logic is *the one and only science necessarily by virtue of which and in terms of which* every judgment, belief, claim, inference, science, or more generally any theoretical activity or product that is in any way justifiable or explicable by reasons, *categorically ought to be* to be explained or justified. Pure general logic is then both adequately explained and justified when we learn that every explanation and justification *whatsoever*, including the explanation and justification *of every other logic*, both *has to* presuppose and use pure general logic, and has to presuppose and use it *alone*, and also *rightly does so*. Pure general logic—The Universal Natural Logic, the paradigm of logical analyticity—is that logic which, uniquely, we alethically must and also categorically ought to presuppose and use in order to *construct* any other logic, in order to *construct* any rational explanation whatsoever, in order to *construct* any rational

justification whatsoever, and in order to *construct* any rational theory whatsoever. Hence the Kantian ethicist Onora O'Neill very aptly calls this line of argument "a constructivist vindication of formulas of logic."[68]

The philosophical thesis of *Constructivism*, whether inside or outside of ethics, holds that human agents or the human mind play an active, basic role in determining and generating the content of all beliefs, truths, knowledge (especially including the knowledge of language), desires, volitions, act-intentions, and logical or moral principles. In this way, The Logocentric Predicament, the Tortoise regress problem, and the problem of justifying deduction are just ways of *showing us* pure general logic's primitive and unique a priori status in any cognitive, scientific, or more generally theoretical constructive activity or product, and in particular its absolutely unique a priori categorically normative status in *all* constructive theoretical explanation and justification whatsoever, including any attempt to construct theoretically an explanation or justification of logic itself. Pure general logic is *the one and only categorically normative a priori condition of the possibility* of all constructive theoretical explanation and justification whatsoever. Otherwise put, pure general logic *must be presupposed and used* in every constructive theoretical explanation and justification whatsoever. And *that* is why logic must be presupposed and used in any attempt to justify or explain logic. It is partially constitutive of our rational human animality. In this sense, pure general logic is not only *transcendental*,[69] but also our rational human logical *duty*.

IV.5 Kantian Argument 4: The Non-Supervenience Argument

The fourth argument is intended to show, again, that first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all

a priori in the Kantian sense of that notion, but in a way that is interestingly distinct from that of the first argument.

But before I spell out The Non-Supervenience Argument, I need to say something about the very idea of *apriority*. In the first *Critique*, Kant says that

> Although all our cognition commences with experience, yet it does not on that account all arise from experience.… It is therefore a question requiring closer investigation , and one not to be dismissed at first glance, whether there is any such cognition independent of all experience and even of all impressions of the senses. One calls such cognitions *a priori*, and distinguishes them from empirical ones, which have their sources *a posteriori*, namely in experience. (*CPR* B1-2)

I think that these remarks express a deep insight. Kant's deep insight is that apriority is in fact the *strict underdetermination* of the semantic content, truth, and justifiability of a statement *S* by all actual or possible sensory experiences and contingent facts, or what is the same thing, *the failure of the strong supervenience of the content, truth, and justifiability of S on all sensory experiences and contingent facts*, and NEITHER

> (1) the supposed fact of the *strict exclusion* of sense experience by the content, truth, or justifiability of *S* = The Classical Rationalist Conception of the A Priori,

NOR

> (2) the supposed fact that some beliefs—e.g., a belief expressed by *S*—are *armchair beliefs* in that a believer or community of believers resolves to hold those beliefs in such a way as to make them immune from empirical disconfirmation = The Pragmatic or Quinean Conception of the A Priori.[70]

Correspondingly, to say that a statement *S* is a posteriori is to say that the semantic content, truth, or justifiability of *S* is strictly determined by or strongly supervenient on some actual or possible sense experiences or contingent facts, and NEITHER merely that *S*'s content must bear a relation to experience, NOR merely that the truth of *S* must be learned or confirmed by means of experience, NOR merely that *S*'s justification must be supported by experiential evidence and established by experimental methods. Otherwise put, Kant's deep insight is that there is no such thing as semantic content, truth, or

objective knowledge (i.e., fully justified true belief) that *altogether excludes* sensory experiences or contingent facts, but that it does *not* follow from this that any version of full-strength Empiricism is true—i.e., that semantic content, truth, and justifiability are either strictly determined by/strongly supervenient on or (even more radically) reducible to actual or possible sensory experiences and contingent facts. That is clearly and simply a *non sequitur.*

Just to be perfectly clear and explicit about a familiar idea, strong supervenience[71] is a strict determination-relation between sets of properties of different ontological "levels," a relation that is weaker than strict property-identity, and is usually taken to be asymmetric, although two-way or bilateral supervenience is also possible. But assuming for the purposes of simpler exposition that supervenience is asymmetric, then, more precisely, B-properties (= the higher level properties) strongly supervene on A-properties (= the lower-level properties) if and only if

> (i) for any property F among the A-properties had by something X, F necessitates X's also having property G among the B-properties (upwards necessitation),

and

> (ii) there cannot be a change in any of X's B-properties without a corresponding change in X's A-properties (necessary co-variation).

It follows from strong supervenience that any two things X and Y share all their A-properties in common only if they share all their B-properties in common (indiscriminability). Facts are just actual or possible instantiations of properties. Hence strong supervenience for properties entails strong supervenience for facts, and failures of strong supervenience for properties correspondingly entails failures of strong supervenience for facts.

Now even if the *existence* of necessary truths logically strongly supervened on everything, it would not follow that their *specific character* logically strongly supervenes too. For although all logically necessary truths in first-order monadic logic and pure general logic are necessarily equivalent, their *structural senses* are different in virtue of their inherently different logical logical forms. For example, "P→ P" does not have the same structural sense as "Pv ~P" because its logical form is inherently different. It is in virtue of *transformation* rules—e.g., De Morgan's Equivalences—that we are able to move logically at will, or logically spontaneously, from one logical truth having a certain structural sense, to another logical truth having a distinct although necessarily equivalent structural sense. So their structural senses can vary independently of their being logically necessarily true, and this intensional fact is made manifest by the application of transformation rules. In turn, therefore, their structural senses do not logically strongly supervene on whatever it is that their existence logically supervenes on, under the supposition that their existence logically strongly supervenes on everything. And that is true in every logically possible world: logically necessary truths in first-order monadic logic and pure general logic with inherently different logical forms are all intensionally non-equivalent. So their specific character does not *logically* strongly supervene on *anything*, except of course on first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction themselves. Nor does their specific character *merely strongly* supervene on anything, except of course on first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction themselves. If their specific character does not either logically or merely strongly supervene on anything but first-order monadic logic, pure general logic,

and The Minimal Logical Meta-Principle of Non-Contradiction themselves, then since

none of these is strongly supervenient on actual or possible sense experience and

contingent facts, *then they are all a priori*.

IV.6 Kantian Argument 5: The Weak Transcendental Ideality Argument

Suppose that I am correct that first-order monadic logic, pure general logic, and

The Minimal Logical Meta-Principle of Non-Contradiction are all "transcendental" in the

sense specified at the beginning of this section. The fifth argument is intended to *explain*

why this is so, by showing that first-order logic, pure general logic, and The Minimal

Logical Meta-Principle of Non-Contradiction are all *weakly or counterfactually*

transcendentally ideal, or WC-ly TI, for short

So now I am going to argue explicitly that first-order monadic logic, pure general

logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all WC-ly TI.

1. First-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are either (i) physical, (ii) platonic, (iii) sense-experiential, (iv) conventional or social, or (v) transcendentally ideal. (Premise, justified by either transcendental argument or transcendental explanation)

2. If either first-order monadic logic, pure general logic, or The Minimal Logical Meta-Principle of Non-Contradiction were physical, then they would be contingent. But first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all necessary. So first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are not physical. (Premise, justified by either transcendental argument or transcendental explanation)

3. If either first-order monadic logic, pure general logic, or The Minimal Logical Meta-Principle of Non-Contradiction were platonic, then they would be unknowable by Benacerraf's Dilemma considerations.[72] But first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all knowable a priori. So first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are not platonic. (Premise, justified by either transcendental argument or transcendental explanation)

4. If either first-order monadic logic, pure general logic, or The Minimal Logical Meta-Principle of Non-Contradiction were sense-experiential, then they would be a posteriori. But first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all a priori. So both first-order monadic logic and pure

general logic are not sense-experiential. (Premise, justified by either transcendental argument or transcendental explanation)

5. If either first-order monadic logic, pure general logic, or The Minimal Logical Meta-Principle of Non-Contradiction were conventional or social, then they would be either physical, sense-experiential, logically strongly supervenient on physical facts or sense-experiential facts, or merely strongly supervenient on physical facts or sense-experiential facts. But neither first-order monadic logic, nor pure general logic, nor The Minimal Logical Meta-Principle of Non-Contradiction is either physical, sense-experiential, logically strongly supervenient on physical facts or sense-experiential facts, or merely strongly supervenient on physical facts or sense-experiential facts. So neither first-order monadic logic, nor pure general logic, nor The Minimal Logical Meta-Principle of Non-Contradiction is conventional or social. (Premise, justified by either transcendental argument or transcendental explanation)

6. Therefore first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all transcendentally ideal. (From 1-5, and Disjunctive Syllogism)

7. If something is transcendentally ideal, then it is either strongly TI or else WC-ly TI. (Premise, justified by either transcendental argument or transcendental explanation)

8. Strong TI is false. (Premise, justified by either transcendental argument or transcendental explanation)

9. Therefore first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all WC-ly TI. (From 7, 8, and Disjunctive Syllogism)

The fifth and final argument I have just spelled out is clearly valid, since it is in the form of two simple disjunctive syllogisms. But at the same time, it is equally clear that its soundness rests on the seven premises, each of which is justified by *transcendental* argumentation or explanation, whose rational support is therefore only *fairly* reliable, and does not flow from the highest kind of evidence, i.e., the self-evidence of authoritative rational intuition.[73] Nevertheless, I do think it can still be truly said that this argument makes a *fairly plausible case* for the weak or counterfactual transcendental ideality of first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction.

If the five arguments I have just spelled out are all in fact sound, then The L-is-T Thesis is true for first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction. Now if first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all necessary, a priori, and do not logically supervene on anything but themselves, then none of them logically supervenes on anything physical, contingent, sense-experiential, or conventional or social. This in turn entails that not everything logically supervenes on the physical world, the contingent natural world, the sense-experiential natural world, or the social world. So Scientific Naturalism is false, Physicalism is false, and Empiricism is false, including Logical Empiricism. If first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all categorically normative for all rational human cognition and thought, then they are necessarily presupposed by, and also conditions of the possibility of, all rational human cognition and thought. Because first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all weakly or counterfactually transcendentally ideal, it also follows that Platonism about logic is false. And finally, because first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all transcendental in *all* senses of that notion as I specified it at the beginning of this section, it follows that human rationality, human cognition, human thought, first-order monadic logic, pure general logic, and The Minimal Logical Meta-Principle of Non-Contradiction are all essentially bound up with one another, and stand or fall together. As Kant and early Wittgenstein so brilliantly saw, philosophical logic bottoms out in serious Transcendental Idealist metaphysics.

IV.7 The Nature of Logic in the *Tractatus*

As I pointed out earlier in sub-section IV.4, one fundamental problem in the philosophy of logic is The Logocentric Predicament. And as I also pointed out in that discussion, one of the essential variants of The Logocentric Predicament is *the problem of justifying deduction*:

> (1) All justification is either non-deductive (e.g., inductive) or deductive.
>
> (2) On the one hand a non-deductive justification of deduction is too weak and on the other hand a deductive justification of deduction is circular.
>
> (3) Therefore, deduction cannot be justified

In his 1913 *Notes on Logic*, Wittgenstein saw this problem clearly:

> Deductions only proceed according to the laws of deduction but these laws cannot justify deduction. (*NB*: 93)

One way of interpreting early Wittgenstein's theory of the nature of logic, then, is that it is essentially an extended attempt to solve the problem of justifying deduction, and therefore also essentially an extended attempt to find an acceptable way out of The Logocentric Predicament. This is the overarching interpretive frame that I shall adopt in the rest of this section.

Early Wittgenstein's theory of the nature of logic, understood as an extended attempt to solve the problem of justifying deduction and thereby find an acceptable way out of The Logccentric Predicament, also flows from his answer to what is perhaps the basic question in philosophical logic: *what is a proposition*? Generally speaking, of course, a proposition is both

> (1) the "content" of a judgment, assertion, or belief, or "what it says,"

and also

> (2) the "truth-bearer," or what has a classical truth-value (T or F) essentially.

According to Kant, propositions are complex mental representations, systematically built up out of intuitions and concepts, inherently and non-empirically unified under the original synthetic unity of apperception (a.k.a. "transcendental apperception"), intrinsically governed by logical forms and laws, that in turn constitute mind-dependent empirical or non-empirical states of affairs insofar as they are well-formed, intersubjectively sharable, and based on the actual givenness of objects to human sensibility. According to Frege, by contrast, propositions are the strongly mind-independent, platonically abstract, complex senses of indicative sentences, and also the direct objects of assertions or judgments, systematically composed under maximally general and categorically normative logical laws that are inherently about the True and the False. And according to early Russell, by another contrast, propositions are complexes of objects of different types—e.g., simple individuals, 1-place first-order properties, first-order relations or n-place first properties, functions, higher-order properties and relations, and logical constants—to which judging subjects are multiply psychologically related in a certain order, under non-psychological, universal, necessary, and a priori laws of deductive consequence.

Early Wittgenstein's theory of propositions draws more or less eclectically from all of these sources, but depends crucially on his theory of atomic facts (i.e., actual "states of affairs" or *Sachverhalten*, built up out of Tractarian objects, logical forms of objects, classical logic, and the active contribution of the representing and language-using subject). In a nutshell, for early Wittgenstein, *a proposition is a non-atomic or molecular linguistic fact, presupposing logic, that is directly correlated by a thinking and language-using subject with another (usually non-linguistic) molecular or atomic fact in order to*

cognize that fact, such that the first (linguistic) fact "pictures" the other fact and thereby significantly represents that fact.

Since for early Wittgenstein, as we have seen, logic is transcendental and since language necessarily mirrors the world, then necessarily for every actual or possible atomic fact there is also a correlated "picturing" proposition, and for every atomic proposition there is also a correlated "pictured" atomic fact. Moreover, necessarily for every actual atomic fact (= what is positively the case) there is a true atomic proposition, and for every merely possible atomic fact (= what is negatively the case) there is a false atomic proposition. This also implies that, necessarily

(1) For every positive or negative atomic fact there is a correlated true or false atomic proposition, directly related by negation to its bipolar logical opposite.

and also necessarily

(2) For every non-atomic or molecular fact there is a correlated non-atomic or molecular proposition that is a truth-function of atomic propositions.

It should also be noted in this connection that for early Wittgenstein falsity is the logical complement of truth (i.e., it applies to everything in the world other that is other than what is actually the case), and negation is a logical operation that reverses the truth-value of any proposition to which it is applied. So if a proposition ~P is true, then it follows logically that P is false, and that P is true. But if P is true, it does not follow that we can determine a unique proposition P = Q that is false, for ~P merely tells us that the whole world is actually other than it would have been if (contrary to actual fact) P had been the case.

Since the world bottoms out in atomic facts—objects, like things-in-themselves according to early Wittgenstein's solipsism/LSTI in the *Tractatus*, are the *substance* of the world, not the *world* itself—and since language and logic both mirror the world, then

the set of all propositions bottoms out in atomic propositions. Atomic propositions are linguistic sequences of names of objects, occurring in a certain fixed order, as used by a talking and thinking subject. Names are the simple or undecomposable atoms of propositions, just as objects are the simple or undecomposable bits of reality. Considered apart from its use, the linguistic sequence of names of objects is a propositional "sign." The propositional sign, in turn, is itself a non-atomic or molecular fact in the world. Signs, whether names or propositional signs, become meaningful "symbols" only by being *applied* and *used* by rational human linguistic cognizers (*TLP* 3.262, 3.326, 3.328, 51, 57). Each name has its meaning by directly picking out an object, and the atomic proposition "pictures" by having each of its names correlated one-to-one with an object, and also by having an isomorphism relation—i.e. a two-way, or bijective, sameness-of-structure relation—between the configuration of names in the propositional sign, and a corresponding configuration of the objects picked out by the names. Picturing is thus a meaning-conferring isomorphism between a non-atomic or molecular linguistic fact (i.e., the atomic propositional sign) and an atomic fact in the world. And in this way, each atomic proposition is a itself *meaningful linguistic model or diagram* of a positive or negative atomic fact. For example, the true propositional symbol

Frege is taller than Wittgenstein

linguistically models the real-world relation of relative height between Frege and Wittgenstein. What the proposition and the fact share in common—the same structure they both have—is called the "the form of representation" or *Form der Abbildung* (*TLP* 2.151, 39, 43).

One crucial thing to note here is that the atomic propositional symbol

Frege is taller than Wittgenstein

which can be formally symbolized as

$T^2 fw$

and which pictures the positive atomic fact that Frege is taller than Wittgenstein, contains

not just *two* but *three* names, i.e., '$T^2 xy$', 'f', and 'w'. In other words, strange as it may

seem, '$T^2 xy$' *is a Tractarian name*. More precisely, '$T^2 xy$', or (in English) 'x is taller

than y', *names a 2-place relation*. More generally, the class of Tractarian objects includes

individuals, 1-place properties, and n-place relations, and every meaningful expression in

the atomic proposition, whether a proper name, a monadic predicate, or a polyadic

predicate, *is a Tractarian name*. So

Frege is a philosopher

which is formally symbolized as

$P^1 f$

contains *two* names, namely '$P^1 x$' or 'x is a philosopher', which is a name that refers to

the 1-place property of being a philosopher, and also 'f' or 'Frege', which is a name that

refers to Frege.

This in turn allows me to construct a highly simplified, or toy, model of

Wittgenstein's theory of picturing in atomic propositions, or what I will call "The

Spheres-and-Hooks-and-Hangers model," or The S-H-H model for short. Here is how

The S-H-H model goes. Think of individual Tractarian objects as small spheres

connected to hooks, rather like Christmas ornaments. For each such sphere there must be

one hook. Hooks, in turn, occur on hangers. Hangers are *also* Tractarian objects. But each

hanger can have either one hook or more than one hook. Think of a 1-hook hanger as a 1-

place property, and think of a many-hook hanger as a many-place property or relation.

Then to each of the *n* names in a proposition there are *n*-1 spheres, each of which (in my

Tractarian toy model) is an individual object named by that name; and for each 1-place predicate or relational predicate in one of these propositions, we have in (in my Tractarian toy model) a *n*-1 hooked hanger which is either a 1-place property or else a many-placed property or relation, named by that predicate.

Consider, e.g.,

```
|-----------------|------------------|
    ∈                ∉               ∠
  Frege gives the beer bottle to Russell.
```

This propositional symbol is formally symbolized as

G^3fbr

Here, the three individual names are 'Frege', 'the beer bottle', and 'Russell', and the fourth name of the 3-place relation is 'G^3xyz', or (in English) 'x gives y to z'. Here is another example:

```
    |------------------------|
    ∈                        ∉
  Frege is taller than Wittgenstein.
```

This propositional symbol is formally symbolized as:

T^2fw

Here the two individual names are 'Frege' and "Wittgenstein', and the third name of the 2-place relation is 'T^2xy', or 'x is taller than y'. Finally, for the lower bound case of facts, the fact constructed of an individual and a 1-place property, here is the last example, whose toy model includes a 1-place hanger and a single sphere:

```
    |---------------------
    ∈
  Frege is a philosopher.
```

This sentence is formally symbolized as:

P^1f.

Here the individual name is 'Frege', and the name of the 1-place property is 'P^{1}x', or 'x is a philosopher'.

One important logico-philosophical point conveyed by The S-H-H model is that *Tractarian objects are not homogeneous in character*. On the contrary some Tractarian objects are what Frege would have called "saturated" entities (complete individual objects or particulars), whereas other Tractarian objects are what Frege would have called "unsaturated" entities (incomplete non-individual objects or universals—in this case, Fregean concepts, or functions from objects to truth-values, i.e., 1-place properties). Frege postulated a radical ontological difference between individual objects and functions, and this got him into unresolvable difficulties—e.g., that the concept HORSE is not a concept, because 'the concept HORSE' is a meaningful name and must pick out a complete individual object, yet concepts are functions and so cannot be complete individual objects. Wittgenstein effectively avoids and thereby (dis)solves this problem by allowing in complete individual objects *and also* unsaturated objects as Tractarian objects.

The other important logico-philosophical point made by The S-H-H model is that it is now quite easy to see how atomic propositions picture facts, and do not merely sequentially name objects (i.e., make lists of objects). That is, atomic propositions represent not only objects, but also represent *specifically how objects of one type go together with other objects of another type in order to form structured atomic facts*. This requires, in effect, that the user of language must always *project* in a certain way *from* the several names making up the propositional sign *onto* the configuration of objects. Otherwise it would be impossible to explain why

Frege is taller than Wittgenstein

is a propositional sign, whereas

Frege Russell Wittgenstein

is not a propositional sign.

To grasp precisely how the propositional sign is to be projected onto the correlated objects by the user of the sentence, is to grasp the *sense* of that proposition. Signs for early Wittgenstein, as I noted above, are perceptible and real parts of language—i.e., words—and in particular are fundamentally *types* of words, and only derivatively *tokens* of words. (The type-token distinction for letters in the English alphabet can be easily conveyed by asking "how many letters are there in 'aardvark'?": if you say 8 then you are counting letters as tokens, but if you say 5 then you are counting them as types.) As I also noted above, for early Wittgenstein, symbols, by contrast to signs, are signs with a meaning (whether sense or Meaning, i.e., reference, *Sinn* or *Bedeutung*) via their application and use by rational human linguistic cognizers. In natural or ordinary language, it is possible to have two or more different signs with the same meaning (synonymy), and it is also it is possible that one sign has two or more different meanings (ambiguity). Also it is possible for signs to lack a meaning (empty names or predicates). Hence two or more different signs can express the same symbol, the same sign can express different symbols, and some signs do not express symbols. Ambiguity or semantic emptiness can lead to confusion, unsoundness in arguments, contradiction, or even paradox. In a logically *perspicuous* language, however (a.k.a. an "ideal language"), every sign has a meaning, and each sign has one and only one meaning.

Signs for early Wittgenstein , as I have emphasized, do not have a meaning on their own: *they have to be applied and used* by creatures like us Otherwise put, signs

without a use are semantically "dead." In order to be meaningful and get a semantic "life"
of their own they must be semantically "vivified" by a *correct* application and usage of
those signs by some or another rational human linguistic cognizer, and can be
semantically vivified *only insofar as* they are correctly used by some or another rational
human linguistic cognizer. In this sense, meaningful language is *strongly
transcendentally ideal.*

Not just *any old "mess of words"* (*Wörtergemisch*) (*TLP* 3.141, 45) will count as
a propositional sign, however: instead, only sequences that obey precise logico-
grammatical formation rules (a.k.a. "rules of logical syntax") will be allowed as
propositional signs. For example,

is and but dog

Frege

and

sweetly sweetly Frege whisky

are not propositional signs. In this way, propositional significance for early Wittgenstein
is strictly constrained by *transcendental logical syntax.*

In this connection, it should be noted that there are two different kinds of logico-
grammatical formation constraints, hence two distinct levels of the transcendental logical
syntax of language. The first level of transcendental logical syntax has merely to do with
the forming or ordering of names in atomic propositions, and also of other words in non-
atomic or molecular propositions. For example,

Frege is a philosopher

is a transcendentally well-formed or well-ordered propositional sign, whereas

sweetly sweetly Frege whisky

is not. Such violations are are nothing but a priori logico-syntax nonsense. The second level of transcendental logical syntax has to do with the *logical types* of names, and correspondingly with the *logical forms* of objects correlated with those names, e.g., *space*, *time*, and *color* or *coloredness* (*TLP* 2.0251, 37). Importantly, violations of these second-level rules of transcendental logical syntax can emerge even when the sequence of names obeys the first-level rules of well-formation or well-ordering, e.g.,

Frege is nothing but a colorless green idea who sleeps furiously

or

Frege and quadruplicity together drink procrastination.

Such second-level violations of transcendental syntax are nowadays called instances of *sortal incorrectness*, and constitute logico-syntactical nonsense of a higher order than first-level ill-formedness or ill-ordering. In a Tractarian framework, these violations constitute *transcendental sortal incorrectness*, and correspond directly directly to violations of what Kant calls *Pure Concepts of the Understanding*, or *Categories* (*CPR* A64-83/B89-116), which, according to Kant, collectively capture the unique and complete set of basic logical functions of the rational human faculty of understanding. Correspondingly, such violations are also of great philosophical import for early Wittgenstein, because according to him virtually all philosophical errors consist in transcendental sortal incorrectness.

One profound difficulty about transcendental sortal incorrectness is that it may superficially *seem* perfectly logically or philosophically acceptable. For example, according to early Wittgenstein, the well-formed sequences

Two is a number

Frege is identical to Frege

Tully is identical to Cicero

and

Frege judges (asserts, believes, thinks) *P*

are all logico-philosophically subtle cases of transcendental sortal incorrectness, even though they seemed perfectly acceptable to both Frege and Russell. Among other things, this means that while our rational intuitions about linguistic well-formedness appear to provide a reliable basis for theorizing about syntax (as Chomsky pointed out), by contrast our rational intuitions about sortal correctness apparently are *not* reliable.

In any case, the every formation of propositional signs with a sense falls necessarily under a set of a priori logico-grammatical rules of transcendental syntax. But the essence of sense is the picturing relation as it occurs in atomic propositions. Every complex sense of a complex or molecular proposition is systematically inherited from the senses of its constituent atomic propositions. But what is the "sense" of sense? That is, what is its larger meaning, or rational purpose? The larger meaning or rational purpose of sense is to convey true or false information about atomic or molecular facts. An atomic proposition is true if and only if the fact that it pictures is a positive or actual atomic fact. Otherwise it is false. And a molecular proposition is true if and only if the truth-function of its component atomic propositions assigns it the truth-value "True." Early Wittgenstein's theory of truth for atomic propositions can be regarded as the purest version of the correspondence theory of truth, since from the grasp of the structure of the atomic propositional sign it is possible literally to read off the structure of its correlated atomic fact. Furthermore, early Wittgenstein's theory avoids the main problem in the traditional theory of correspondence, i.e., that the cognizer is required to justify the similarity between sign and object, which requires a vicious regress of higher-order

correspondence relations and justifications. On early Wittgenstein's theory, the isomorphism between propositional symbol and fact is *built right into the sense of the proposition*: then the issue of truth or falsity is merely the issue of whether that fact actually exists or not, which is external to the representing subject.

This brings us to the way in which early Wittgenstein's theory of truth adequately expresses his "pure realism," or Empirical Realism in Kant's sense. More precisely, early Wittgenstein's theory of truth introduces *no representational intermediary whatsoever* between language and the truth-making facts, hence it is fully immediate or direct. The correct use of the sentence maps the rational human linguistic cognizer onto the atomic fact, and then that atomic fact either actually exists or not, *and that's it*. In section II, I noted two fundamentally different senses of realism: Noumenal Realism, and Empirical Realism. But this does not exhaust the basic possibilities for philosophy, and the very idea of "realism" in the philosophical sense has at least six different kinds that are not always carefully distinguished:

(1) absolute mind-independence of the facts (Noumenal Realism),

(2) moderate *mind-independence* of the facts, together with either strong or transcendentally idealistic *mind-dependence* of the facts (Empirical Realism),

(3) indirect or mediated cognition of the facts via intermediary mental representations, e.g., sense data or concepts (Epistemic Indirect Realism),

(4) direct or unmediated cognition of the facts via perception or judgment (Epistemic Direct or Naïve Realism),

(5) indirect or mediated linguistic representation of the facts via intermediary descriptions or theories (Semantic Indirect Realism),

and

(6) direct or unmediated linguistic representation of the facts via non-descriptive, directly referential terms or propositions (Semantic Direct Realism).

Early Wittgenstein's theory of truth is *not* realistic in senses (1), (3), or (5), but it *is* realistic in senses (2), (4), and (6). This threefold counjunction of Empirical Realism, Epistemic Direct or Naïve Realism, and Semantic Direct Realism comes out in two basic ways.

First, early Wittgenstein's picture theory of meaning entails a *truth-maker semantics.* That is, positive and negative facts are themselves the worldly *truth-makers* and *falsity-makers* of atomic propositions, and senses of atomic propositions are intrinsically bound up with the bipolarity of the proposition, i.e., its classical truth or classical falsity. Given a sense of an atomic proposition, it completely divides the world into the atomic fact it pictures, its truth-maker, and everything else other than that atomic fact, its falsity maker.

Second, early Wittgenstein's picture of meaning entails a *transparency theory of judgment.* That is, for early Wittgenstein, the judger becomes nothing but a transparent cognitive conduit to the facts. How, more precisely, does his transparency theory of judgment work?

What I will call "primitive sense" is the semantic information conveyed by the actual way in which (or: the actual rule according to which) an atomic propositional sign in ordinary language or thought is used by a talking and thinking subject as a symbol in order to picture an atomic fact with a view to truth. On early Wittgenstein's theory, then, there is no need to add a further special act of judgment or belief in order to account for picturing, since when a propositional sign is used in the correct way, it just *is* a judgment or assertion or belief. In this way, the correct use of a propositional sign on the one hand, with its primitive sense, and propositional activity in ordinary language (i.e., judgment,

assertion, statement-making, belief, or thinking) on the ther hand, *are one and the very same thing*. So, for early Wittgenstein, to say that I judge or assert or believe or think *P*, is only to say:

'P' says P

or:

'P' has a primitive sense.

No psychological verbs are required, and in fact the explicit addition of psychological verbs produces transcendental sortal incorrectness or syntactic nonsense, rather like those trite bumper-stickers that one used to see sometimes in the 1990s, which said

If you can read this, thank a teacher.

In order to avoid *trite transcendental nonsense* according to early Wittgenstein, what those trite bumper stickers should have said was

Thank a teacher

As we have just seen, propositions for early Wittgenstein are at once

(1) the "content" of judgments, assertions, statements, beliefs, and thoughts

(2) bipolar truth-bearers,

(3) logical pictures of atomic facts,

and

(4) the vehicles of primitive sense.

Nevertheless, Tractarian propositions have at least seven other necessary features as well.

First, *propositions are the proper subjects of logical analysis*, in that they are essentially decomposable into their simple symbols, or names, and the way in which those names are configured into a propositional structure. Each proposition has a unique

complete decomposition (*TLP* 3.25, 49). Logical "elucidation" is then the activity of decomposing a proposition into its simple constituent symbols (*TLP* 3.263, 51).

Second, in the reverse direction, *propositions are essentially compositional*, in that each propositional symbol is a function of its component simple symbols or expressions. This compositionality of the proposition entails what is sometimes called the "creative" or "productive" aspect of language, namely that an infinitely large number of new propositions ("new sense") can be constructed from a finite set of simple symbols ("old words), together with and according to transcendental syntactical rules for construction (*TLP* 4.026-4.03, 69). This constructive compositional capacity for language, in turn, is an innate rational human endowment (*TLP* 4.002, 61-62).

Third, *propositions are essentially generalizable*, in that each meaningful part of the proposition can be replaced by a variable while other parts are held constant, thus producing a distinctive class of propositions of that form. For example, the proposition

Frege is taller than Wittgenstein

can be generalized as

x is taller than Wittgenstein

and then there will be a class of propositions determined by substituting different individual constants for the variable 'x'. Alternatively, the original proposition can also be generalized as

Frege is taller than y

or as

Frege bears R to Wittgenstein

or as

x bears R to Wittgenstein

or as

x bears R to y

and so-on. The absolutely general form of a proposition is the propositional variable '*P*', which simply means: "Such and such is the case" (*TLP* 4.5, 103).

Fourth, *propositions are the primary or primitive units of meaning*, in that all other symbols, including names (*TLP* 3.3, 51), have meaning (i.e., either sense or *Sinn*, or else Meaning, reference, or *Bedeutung*) only in the context of propositions.

Fifth, *propositions are semantically self-intimating*, in that they convey sense and are the vehicles of sense, but propositional senses cannot described or named: they can *only* be shown (*TLP* 4.022, 67), and what can be *shown* cannot be said (*TLP* 4.1212, 67)

This raises the equally difficult and important question: for early Wittgenstein, what is showing (*zeigen*)? It seems to me that the best overall characterization of showing is that it covers all the basic types of linguistic meaning or discursive significance *other than* describing facts, which is early Wittgenstein specifically calls "saying" (*sagen*). In other words, the following eight types of linguistic meaning are each different basic types of *significant discursive non-saying*, i.e., different basic types of Tractarian *showing*—

(i) intensional discourse (i.e., discourse about meanings),

(ii) reflexive discourse (i.e., self-referring discourse),

(iii) non-literal discourse (e.g., discourse expressing jokes, metaphors, puns, or other wordplay),

(iv) speech-act-expressing discourse (i.e., discourse that conveys types of speech acts, e.g., imperatives or questions),

(v) aesthetic discourse (i.e., discourse expressing esteem of the beautiful or disgust at the ugly, etc.),

(vi) emotive discourse (e.g., discourse expressing approval or disapproval, commitment or non-commitment, positive or negative evaluation, etc.),

(vii) practical discourse (e.g., discourse expressing reasons for action, deliberation, or decisions),

and

(viii) moral discourse (e.g., discourse expressing moral obligation, permissibility, or impermissibility).

Notice that according to this conception of Tractarian showing as "significant discursive non-saying," even Tractarian

(ix) *naming* (e.g., discourse via proper names, demonstratives and other indexicals, 1-place predicates, and n-place predicates),

as a non-descriptive, non-fact-stating yet still importantly significant form of discourse, is *also* a kind of showing, and thereby constitutes a *ninth* basic type of showing.

Sixth, *propositions are semantically non-reflexive*, in that they cannot refer to or describe themselves (*TLP* 3.332, 57). Correspondingly, functions cannot contain themselves as arguments (*TLP* 3.333, 57). Together these two forms of non-reflexivity automatically rule out the possibility of The Liar paradox and other classical semantic paradoxes, and also the set theoretic paradoxes.

Seventh, *propositions are essentially first-order*, in that when we take the fifth and sixth features together, it follows that although complex or molecular propositions are possible, no higher-order propositions are possible: there are no senses of senses (hence no hierarchy of senses), and there are no propositions about propositions (hence no hierarchy of propositions). Early Wittgenstein's rationale here seems to be that such hierarchies are in themselves irrational or viciously regressive. On the other hand however, by banning senses of senses and propositions about propositions, he puts his own Tractarian discourse in jeopardy. How can he himself, as the author of the *Tractatus*,

meaningfully talk about propositions? I will have more to say about this famous Tractarian problem of philosophical reflexivity, and how Wittgenstein proposes to deal with it, in section V.

As we saw under the third point just above, early Wittgenstein's theory of the logical form of single propositions says that there are as many ways of generalizing propositions as there are ways of abstracting out names as variables and holding the other propositional elements fixed (*TLP* 3.31-3.317, 51-52). The general form of a proposition is the limit case of abstraction in which all names are replaced by variables and the proposition itself is considered as a single variable (*TLP* 4.5, 101).

A *deduction*, by contrast to a single proposition, is a sequence of propositions that are related by "laws of inference," such that the last proposition in the sequence (the conclusion) is a logical consequence of the other propositions in the sequence (the premises), according to those laws of inference. Early Wittgenstein's idea is that the conclusion of every such deduction is "internally related" to the complex proposition which is the true conjunction of all its premises and thus "contained" in that complex proposition (*TLP* 5.131, 107). Another way of putting this is to say that in deduction the conditions under which all the premises are true will suffice for the truth of the conclusion (*TLP* 5.11, 5.123-5.124, 105, 107). But the most perspicuous way of putting this is to say that the logical *structure* of the complex proposition that is the true conjunction of all the premises, guarantees the truth of the conclusion (*TLP* 5.13, 107).

What, more precisely, is going on here? This involves our getting a handle on early Wittgenstein's general theory of *the logical form of propositions*. This has four basic parts:

(1) All propositions can be reduced to logical operations on atomic or elementary propositions (*TLP* 5.21-5.3, 114-119).

(2) All logical operations on propositions and also all logical relations between propositions (represented by "logical constants") are truth-functional operations and truth-functional relations (*TLP* 4.3-4.45, 5, 91-93, 103).

(3) The truth definition of the universal quantifier is that it is an extended conjunction of all the atomic propositions generated by replacing the individual variables by individual constants (logical product), and the truth definition of the existential quantifier is that it is an extended disjunction of all the atomic propositions generated by replacing the individual variables by individual constants (logical sum) (*TLP* 5.521-5.524, 135-137).

(4) All truth-functional relations between propositions can be reduced to the single Sheffer stroke function (a.k.a. "continuous negation"):

$$P|Q \ / \ \sim P \ \& \sim Q \ \text{[or alternatively: } P|Q \ / \sim P \ v \sim Q].$$

Using the De Morgan equivalences relating negation, conjunction, and disjunction, i.e.,

$$(P \& Q) \ / \ \sim P \ v \sim Q$$

$$(P \ v \ Q) \ / \ \sim P \ \& \sim Q$$

and also the equivalence between the conditional, and negation, disjunction, and conjunction, i.e.,

$$P \rightarrow Q \ / \ \sim P \ v \ Q$$
$$P \rightarrow Q \ / \ \sim (P \ \& \sim Q)$$

it is easy enough to see informally how every truth-functional relation can be expressed as a function of the Sheffer stroke, e.g.,

$$P|(Q|Q) \ / \ P \rightarrow Q.$$

Now this four part theory of logical form entails that every valid deduction can be represented by a truth-table showing that for every assignment of truth-values to the atomic propositions of the premises, their true conjunction will suffice to guarantee the truth of the conclusion. So early Wittgenstein's general thesis is that since logical deductions are fully guaranteed by the internal truth-functional structure of complex propositions, then "laws of inference" are in fact unnecessary and deduction is thereby internally justified a priori (*TLP* 5.132-5.133, 109). This, in turn, solves the problem of justifying deduction: for early Wittgenstein, the justification of deduction is *self-intimating* or *self-manifesting*, i.e., justification is supervenient on the truth-functional connections underlying the deductive structure of the valid argument. The error in the original problem of justifying deduction was the implicit assumption that the justification had to be "said": on the contrary, it is simply *shown*.

This solution to the problem of justifying deduction is closely related to another striking Tractarian doctrine, namely that logic is explanatorily self-contained. And this, in turn, can be construed as the Tractarian solution to the general Logocentric Predicament. In a certain way it is impossible to make mistakes in logic, because if we have indeed already cognitively constructed a logic, then this logic is perfectly in order, just as it is. *"Self-evidence" is thus entirely internal to the process of cognitively constructing a logic, and requires no external justification*:

> 5.473 Logic must take care of itself.
>
> A *possible* sign must also be able to signify. Everything which is possible in logic is also permitted. ("Socrates is identical" means nothing because there is no property which is called "identical". The propositiom is senseless because we have not made some arbitrary determination, not because the symbol is in itself impermissible.)
> In a certain sense we cannot make mistakes in logic.
>
> 5.4731 Self-evidence, of which Russell has said so much, can only be discarded in logic by language itself preventing every logical mistake. That logic is a

priori consists in the fact that we *cannot* think illogically.

In other words, the Logocentric Predicament was premised on the false assumption that logical explanation or logical justification has to be *external* to logic itself.

Another basic feature of the Tractarian view of the nature of logic is early Wittgenstein's theory of tautologies and contradictions. All propositions are either atomic (elementary) or molecular (complex). But the total class of all atomic and molecular can be cross-classified into two disjoint classes: contingent (sometimes true and sometimes false) *vs.* necessary (always true or always false). Propositions which are always true are tautologies, and they can be *shown* to be tautologous by constructing their truth-tables, which come out T under every line of their main connective. Tautologies are also called "propositions of logic" because they are true by virtue of their logical form alone. But there are no "primitive propositions of logic": every tautology is equally primitive (*TLP* 6.1271, 167-169). Propositions which are always false are contradictions, and similarly they can be *shown* to be contradictory by constructing their truth-tables, which come out F under every line of their main connective.

There is a deep connection between valid deductions and tautologies: for every valid deduction there is a tautology in "modus ponens" form consisting of a conditional proposition which contains all of the premises as a single conjunction in its antecedent, and the conclusion in its consequent. What this means is that the tautologies logically encode and "give" all the valid deductions of logic, thereby yielding the entire theory of deduction via the theory of tautologies (*TLP* 6.124-6.127, 167). Proof is of only psychological relevance, to help us recognize tautologies, but not in any way necessary for logic itself (*TLP* 6.1262, 167).

Now early Wittgenstein is committed to the fundamental thesis that the logical

constants do not *represent* anything:

> 4.0312 The possibility of propositions is based upon the representation of objects by signs.
>
> My fundamental thought (*Grundgedanke*) is that the "logical constants" do not represent. That the logic of the facts cannot be represented. (*TLP* 69)
>
> 4.441 It is clear that to the complex of the signs "F" and "T" no object (or complex of objects) corresponds; any more than to horizontal and vertical lines or to brackets. There are no "logical objects." (*TLP* 95)
>
> 5.32 All truth-functions are results of the successive application of a finite number of truth-operations to elementary propositions.
>
> 5.4 Here it becomes clear that there are no such things as "logical objects" or "logical constants" (in the sense of Frege and Russell).
>
> 5.41 For all the results of truth-operations on truth-functions are identical, which are one and the same truth-function of elementary propositions. (*TLP* 119-121)

Hence the logical relations between propositions cannot be represented, and in particular

the logically necessary relations between the parts of tautologies or contradictions cannot

be represented. In other words, tautologies and contradictions cannot *picture* atomic facts,

and cannot *represent* complex facts. As a consequence, tautologies and contradictions are

"senseless" or *sinnlos* in that they "say" nothing and do not provide logical pictures of

reality (*TLP* 4.461-4.462, 97-99). But although they are strictly speaking "non-sense" in

that they are *other* than what itself conveys a sense, at the same time they are also not

logically unacceptably "nonsensical" or *unsinnig*:

> 4.4611 Tautology and contradiction are, however, not nonsensical; they are part of the symbolism, in the way that "0" is part of the symbolism of arithmetic.

In this way, early Wittgenstein is implicitly using a distinction between

> (1) logically unacceptable nonsense or *Unsinnigkeit* (e.g., violations of transcendental syntax)

and

> (2) logically acceptable nonsense or *Sinnlosigkeit* (e.g., tautologies and contradictions).

This distinction is crucial for understanding the *Tractatus*, since it turns out that what can "only be shown and not said" is equivalent with *logically acceptable nonsense*.

So according to early Wittgenstein, what is logic, really? As we have seen one part of early Wittgenstein's answer to this question is: logic is how we make the various propositional forms and deduction cognitively manifest to ourselves, and in this sense logic "must take care of itself" in that it explains and justifies deduction in particular and logic itself more generally *internally* via the manifestation of tautologies. But early Wittgenstein, as we have also seen, also wants to connect the nature of logic directly with the nature of language, the nature of thought, and the nature of the world:

> 3.221 Objects I can only *name*. Signs represent them. I can only speak *of* them. I cannot *assert them*. A proposition can only say *how* a thing is, not *what* it is. (*TLP* 49)

> 5.471 The general form of the proposition is the essence of the proposition.

> 5.4711 To give the essence of the proposition means to give the essence of all description, therefore the essence of the world. (*TLP* 127)

> 5.552 The "experience" which we need to understand logic is not that such and such is the case, but that something *is*; but that is *no* experience.

> Logic *precedes* every experience--that something *is* so.

> It is before the How, not before the What. (*TLP* 145)

> 5.5521 And if this were not the case, how could we apply logic? We could say: if there were a logic, even if there no world, how then could there be a logic, since there is a world? (*TLP* 147)

> 5.6 *The limits of my language* mean the limits of my world.

> 5.61 Logic fills the world: the limits of the world are also its limits.

> We cannot therefore say in logic: This and this there is in the world, that there is not.

> For that would apparently presuppose that we exclude certain possibilities, and this cannot be the case since otherwise logic must get outside the limits of the world: that is, if it could consider these limits from the other side also.

> What we cannot think, that we cannot think: we cannot therefore *say* what we cannot think. (*TLP* 149-151)

> 6.12 The fact that the propositions of logic are tautologies *shows* the formal-- logical--properties of language, of the world.

6.124 The logical propositions describe the scaffolding of the world, or rather they exhibit (*stellen ... dar*) it. They "treat" of nothing. They presuppose that names have Meaning, and that elementary propositions have sense. And this is their connexion with the world. It is clear that it must show something about the world that certain combinations of symbols--which essentially have a definite character--are tautologies. Herein lies the decisive point. We said that in the symbols which we use something is arbitrary, something not. In logic only the latter expresses: but this means that in logic it is not *we* who express, by means of signs, what we want, but in logic the nature of the essentially necessary signs speaks for itself. That is to say, if we know the logical syntax of any sign language, then all the propositions of logic are already given. (*TLP* 165)

6.13 Logic is not a theory but a reflexion of the world. (*TLP* 169)

Obviously these Tractarian propositions are not immediately self-explanatory. But three basic theses emerge from them:

(1) Logic is the a priori essence of language.

(2) Logic is the a priori essence of thought.

(3) Logic is the a priori essence of the world.

As I pointed out earlier in this section, to say that something X is a priori is to say that X's representational content, truth, logico-linguistic meaning, or justification is strictly underdetermined by—i.e., neither identical with nor strongly supervenient on—any and all sensory experiences and/or contingent empirical facts. Moreover, to say that something X is *essential* is to say that X is intrinsic to, metaphysically necessary for, and perhaps also metaphysically sufficient for something else. So at the very least, early Wittgenstein is saying that

(i) The representational content, truth, logico-linguistic meaning, or justification of logic is strictly underdetermined by—i.e., neither identical with nor strongly supervenient on—any and all sensory experiences and/or contingent empirical facts.

and also that

(ii) Logic is intrinsic to and metaphysically necessary for language, thought, and the world.

In other words, early Wittgenstein is saying precisely that *logic is transcendental* in each of the five Kantian ways I argued for in sub-sections IV.2 to IV.6, and also that it is ultimately *that* thesis which finally (dis)solves The Logocentric Predicament.

V. The Critique of Self-Alienated Philosophy

> Human reason has this peculiar fate in one species of its cognitions which it cannot dismiss, since they are given to it as problems by the very nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason…. The battlefield of these endless controversies is called *metaphysics*. (*CPR* Avii-viii)

> He who has properly *learned* a system of philosophy …. has formed himself according to an alien reason, but the faculty of imitation is not that of generation, i.e., the cognition did not arise *from* reason for him, and although objectively it was certainly a rational cognition, subjectively it is still merely historical. He has grasped and preserved well, i.e., he has learned, and is a plaster cast of a living human being. (*CPR* A836/B864)

> Among all rational sciences (*a priori*) … only mathematics can be learned, never philosophy (except historically); rather, as far as reason is concerned, we can at best only learn *to philosophize*….[P]hilosophy is a mere idea of a possible science, which is nowhere given *in concreto*, but which one seeks to approach in various ways until the only footpath, much overgrown by sensibility, is discovered, and hitherto unsuccessful ectype, so far as it is has been granted to humans, is made equal to the archetype. Until then one cannot learn any philosophy, for where is it, who has possession of it, and by what can it be recognized? One can only learn to philosophize, i.e., to exercise the talent of reason is prosecuting its general principles in certaine experiments that come to hand, but always with the reservation of the right of reason to inevstigate the sources of these principles themselves and to confirm or reject them. (*CPR* A837-838/B865-866)

> Philosophy is the science of the relation of all cognition to the essential ends of human reason (*teleologia rationis humanae*), and the philosopher is not an artist of reason but the legislator of human reason…. The mathematician, the natural [scientist], the logician are only artists of reason, however eminent the [two] former may be in rational cognitions and however much progress the latter may have made in philosophical cognition. There is still a teacher in the ideal, who contrrols all of these and uses them as tools to advance the essential ends of human reason. Him alone we must call the philosopher; however, … he himself is still nowhere to be found, although the idea of his legislation is found in every human reason. (*CPR* A839-840/B867-868)

> For more than one reason what I publish here will have points of contact with what other people are writing today. –If my remarks do not bear a stamp which marks them as mine,—I do not wish to lay any further claim to them as my property. I make them public with doubtful feelings. It is not impossible that it should fall to the lot of this work, in its poverty and in the darkness of this time, to bring light into one brain or another—but, of course, it is not likely. I should not like my writing to spare other people the trouble of thinking. But, if possible, to stimulate someone to thoughts of his own. (*PI* x[e])

V.1 Kant's Critical Meta-*philosophy in the B Motto and A Preface*

The A edition of the *Critique of Pure Reason*, or the CPR for short, did not include a motto. But the 1787 or B edition of the CPR includes a Latin quotation from the preface of Francis Bacon's *Great Instauration* (*Instauratio Magna*) of 1620, which in turn contains the *Novum Organum* as its second part. Now 'instauration' means

"restoration" or "renewal." Bacon had explicitly and self-consciously initiated a restoration or renewal of what he took to be authentic or true philosophy in four steps:

(i) by rejecting what he took to be the "infinite errors" of Scholastic metaphysics,

(ii) by turning philosophy away from what could exist only beyond what is knowable by means of the human senses,

(iii) by turning philosophy *towards* the empirical facts, the finite, and the mortal, and

(iv) by proposing an essentially *experimental* method for philosophy and science.

That Bacon's philosophical and scientific method is aggressively and reductively inductive, and a mere assembly and categorization of classes of particular facts, is not what is important for Kant in this context. Kant's philosophical and scientific method not only includes Baconian induction and empirical evidence about particular facts, but also projective generalization from the particulars (i.e., classical Empiricist induction), logical deduction, abduction (i.e., inference to the best explanation), and also special forms of non-empirical, non-logical inference he dubs "transcendental deduction" and "transcendental proof." The point of the Motto is just to establish the following analogy:

As *The Great Instauration* is to Scholastic metaphysics, so the CPR is to classical Rationalist metaphysics.

Both are proposing a restorative, renewing, and indeed revolutionary *anthropocentric turn* in philosophy. Where Kant differs sharply from Bacon is in Kant's further thesis that, by means of this anthropocentric turn a new and *essentially mitigated* form of rationalist metaphysics is possible, which he dubs "Transcendental Idealism," and which, if true, both fully incorporates the Baconian, Lockean, Berkeleyan, and Humean empiricist critique of Scholastic or classical Rationalist metaphysics *and also* equally fully incorporates a counterpoint rationalist critique of classical Empiricism. This

fundamental dual theme of Kant's anthropocentric turn and his mitigated rationalism, i.e.,

his Transcendental Idealism—whether it is to be understood as Strong Transcendental

Idealism or STI, or as Weak or Counterfactual Transcendental Idealism or WCTI—is

developed and re-played in many subtle variations thoughout the CPR.

The same dual theme is made autobiographically significant by the important fact

that Kant *himself* was a fully committed classical Rationalist in the tradition of Leibniz

and Christian Wolff during Kant's Pre-Critical period. The Pre-Critical period runs from

the 1740s until at least the late 1760s or early 1770s, when, by his own retrospective

testimony in 1783, Kant was suddenly jolted out of his Leibnizian and Wolffian dreams

by a skeptical Humean Empiricist wake-up call:

> I openly confess that my remembering David Hume was the very thing which many years ago first interrupted my dogmatic slumber and gave my investigations in the field of speculative philosophy a quite new direction. I was far from following him in the conclusions at which he arrived … [But if] we start from a well-founded, but undeveloped, thought which another has bequeathed to us, we may well hope by continued reflection to advance further than the acute man to whom we owe the first spark of light. (*P* 4: 260)

It is reasonable to think that this wake-up call happened in July 1771, when Kant would

have read a German translation of the Conclusion of book I of Hume's *Treatise of*

Human Nature (1739-40), published in the *Königsberger gelehrte Zeitung*, entitled

"Nachtdenken eines Skeptikers" ("Night Thoughts of a Skeptic"). In the Conclusion,

Hume re-states the main claims of his skeptical Empiricist analysis of the concepts of

causation and causal necessity. Kant's reading knowledge of English was fairly limited,

so it seems very likely that he never read either the *Treatise* or the *Enquiry Concerning*

Human Understanding (1748) themselves, although he must have originally encountered

Hume's ideas in the 1750s or 60s in order to be able to "remember" them in 1771. It is

also quite possible that Kant's memory of Hume was further jogged by reading a 1772

German translation of the Scottish common sense philosopher James Beattie's highly influential 1770 *Essay on The Nature and Immutability of Truth*, in which Beattie carefully describes and then vigorously attacks Hume's skeptical Empiricist analysis of causation and causal necessity.

In any case, in the *Treatise* and again in the *Enquiry* Hume defends and develops three crucial theses, each of which importantly influenced Kant, whether positively or negatively, after 1771:

> (i) Human cognition is strictly limited as to its content, truth, and epistemic scope by sensory experience.

> (ii) The class of all judgments is exhaustively divided into those concerning "relations of ideas" (i.e., necessary a priori definitional or stipulative truths, e.g., truths of logic or mathematics) and those concerning "matters of fact" (i.e., contingent a posteriori experimental truths, e.g., truths of natural science).

and

> (iii) All our judgments concerning supposedly necessary causal relations in fact refer exclusively to experience and matters of fact, and that their content and justification is determined solely by non-rational "custom" or "habit," not reason.

As it turns out, Kant fully accepts a carefully qualified version of Hume's thesis (i), but also firmly rejects Hume's theses (ii) and (iii).

In another important and closely-related autobiographical remark in the *Reflexionen*, Kant says that "the year '69 gave me great light" (*R* 5037, 18: 69). By this, he means that in that particular year—falling exactly midway between his seminal 1768 essay "Concerning the Ultimate Ground of the Differentiation of Directions in Space" and his breakthrough 1770 Inaugural Dissertation, "On the Form and Principles of the Sensible and Intelligible World"—he proposed the revolutionary *idealistic* metaphysical doctrine that the ontic structure of physical spacetime *necessarily conforms* to the innate mentalistic structure of the rational human capacity for sensory intuition. In the B Preface

of the CPR, seventeen years later, Kant reports the revolutionary idealistic proposal of the

Inaugural Dissertation in the following way:

> Up to now it has been assumed that all our cognition must conform to the objects; but all attempts to find out something about them *a priori* through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the object must conform to our cognition, which would agree better with the requested possibility of an *a priori* cognition of them, which is to establish something about objects before they are given to us…. If intuition has to conform to the [physical] constitution of the objects, then I do not see how we can know anything of them *a priori*; but if the object (as an object of the senses) conforms to the [mentalistic] constitution of our faculty of intuition (*Anschauungsvermögens*), then I can very well represent the possibility to myself. (*CPR* Bxvi-xvii)

This means either:

> (i) There is a physical-to-mental *identity relation* between the ontic structure of physical spacetime and the innate mentalistic structure of rational human sensibility.

> (ii) There is a mental-to-physical *logical-supervenience-without-identity relation* between the innate mentalistic structure of rational human sensibility and the ontic structure of physical spacetime.

> (iii) There is a mental-to-physical *isomorphism-without-either-identity-or-logical-supervenience relation* between the innate mentalistic structure of rational human sensibility and the ontic structure of physical spacetime,

or most weakly of all:

> (iv) There is a physical-to-mental *strong modal actualist counterfactual dependency relation* between between the ontic structure of physical spacetime and the innate mentalistic structure of rational human sensibility, such that necessarily, if the physical world actually exists, then if rational human cognizers *were* actually to exist, then they *would* be able to know the ontic structure of physical spacetime directly through non-empirical "intuition" (*Anschauung*).

As I argued in section I, *the most philosophically defensible* version of this doctrine is the

conjunction of (iii) and (iv), which I call *Weak or Counterfactual Transcendental*

Idealism, or WCTI. As I also pointed out in section I, WCTI holds even if, and whenever,

no rational human minds actually *do* exist.

In a much later 1798 letter to Christian Garve, however, Kant *also* claims that it

was "the Antinomy of Pure Reason" that awoke him from his dogmatic slumber:

> It was not the investigation of the existence of God, immortality, and so on, but rather the Antinomy of Pure Reason …. that is what awoke me from my dogmatic slumber and drove me to the critique of reason itself, in order to resolve the scandal of ostensible contradiction with itself. (*PC* 12: 257-258).

What Kant wrote to Garve in 1798 may seem, initially, at odds with what he said in 1783 and in the *Reflexionen*, in part because, following what he wrote to Garve, it is then natural to think that Kant's awakening from his Leibnizian and Wolffian dreams actually originally occurred in his 1766 essay *Dreams of a Spirit Seer*. And of course that chronology fits the three-part "dogmatic slumber-Rationalist dreams-Critical awakening" metaphor to a T. But upon reflection we can see that it is in fact perfectly consistent with his earlier claims that it was Hume's skeptical Empiricism about content, truth, and justification of human cognition, especially as applied to the classical Rationalist metaphysical concepts of causation and causal necessity, together with Kant's own transcendentally idealistic thesis about the necessary conformity of the ontic structure of physical space to the innate mentalistic structure of human sensibility, that initiated the Critical Philosophy. This is because the Antinomy of Pure Reason, as discovered in 1766, *also* showed him the self-annihilating character of classical Rationalist metaphysical reasoning, and thereby the possibility of the critique of pure reason.

In effect, then, Kant was awakened and correspondingly enlightened *three times* during the immensely philosophically creative six-year period from 1766-1772. So his revolutionary anthropocentric turn to the mitigated rationalism of Transcendental Idealism in fact has *three* conjoined and equally important philosophical sources:

> (1) Hume's skeptical Empiricism about the content, truth, and justification of human cognition, especially as applied to the classical Rationalist metaphysical concepts of causation and causal necessity, remembered by Kant in 1771 or 1772.

> (2) Kant's own revolutionary idealistic thesis about the necessary conformity of the ontic structure of physical spacetime to the mentalistic structure of rational human sensibility,

discovered by him in 1769.

and

(3) The self-annihilating character of classical Rationalist metaphysical reasoning demonstrated by the Antinomy of Pure Reason, and thereby the possibility of the critique of pure reason, discovered by Kant in 1766.

This three-part source, in turn, generates the three-part critical meta-philosophy of the

CPR, which is

(i) the critical rejection of classical Rationalist metaphysics (including various versions of Dualism and Materialism—according to Kant, metaphysical Materialism must be sharply distinguished from Empiricism, in the sense that it is logically possible to accept Materialism and reject Empiricism, and conversely),

(ii) the critical rejection of the equal and opposite destructive and self-alienating radical skepticism which follows from the self-annihilating character of classical Rationality metaphysical reasoning,

and

(iii) the revolutionary replacement of classical Rationalist metaphysics by a new, inherently anthropocentric, and essentially mitigated kind of rationalist metaphysics: Transcendental Idealism.

As we have seen, The Motto in the B edition conveys this three-part critical meta-

philosophy epigramatically.

The A Preface, however, conveys this same two-part doctrine by means of a

rhetorically effective and semantically pregnant *framing metaphor* based on the tragic

plight of Hecuba. Hecuba was a mythic queen of Troy, wife of King Priam, and the

mother of Hector, Cassandra, and some forty-eight other children (some of them,

presumably, the offspring of Priam's concubines), many of whom died violent deaths.

Hecuba was enslaved by the Achaeans after the Fall of Troy. Kant cites her tragic

lament—

Greatest of all by race and birth, I am now cast out, powerless

—as movingly portrayed in Ovid's *Metamorphoses*. There are other equally moving portrayals in Euripides's *Hecuba* and *The Trojan Women*, and in Dante's 16th century *Inferno*, the first part of *The Divine Comedy*. Significantly, Dante's description of Hecuba has her ultimately descending into madness, and barking like a dog (*Inferno*, XXX, 13-20).

Kant's framing metaphor aligns Hecuba's plight with the tragic career of classical Rationalist metaphysics, the erstwhile "queen of all the sciences." The truly memorable opening line of the CPR in the A edition—

> Human reason has this peculiar fate in one species of its cognitions which it cannot dismiss, since they are given to it as problems by the very nature of reason itself, but which it also cannot answer, since they transcend every capacity of human reason. (*CPR* Avii)

—tells us that there is an innate cognitive capacity, faculty, or power (*Vermögen*), namely reason or *Vernunft*, which is partially *constitutive* of human cognition, but at the same time this very faculty of reason is the inherent source of insoluble, self-alienating, self-destructive difficulties for human cognition. More precisely, the faculty of reason in its human involvement naturally presents certain philosophical problems that lead inevitably to corresponding questions which in turn simply cannot be answered, even in principle, since the answers would require a kind of knowledge that inherently transcends the scope of human reason itself. So in other words, the very cognitive faculty that, according to classical Rationalism, is supposed to be, on its own, the strict determiner of the content, truth, and justification of necessary a priori judgments—namely, pure reason or *Vernunft*—by a wholly tragic reversal, turns out to be the primitive source of its own rational self-stultification, or cognitive suicide—in effect, a self-inflicted descent into barking madness, like poor Hecuba. How the philosophically mighty have fallen.

Kant then provides a psychologically-oriented *meta*-philosophical diagnosis of the logico-metaphysical and pragmatic self-stultification, or *cognitive suicide*, of pure reason. In an important sense, it is not reason's fault. For it is built into the very nature of innately-specified faculty of human reason that it cognizes and follows principles that are adequately warranted by their application to human experience. Nevertheless, the recursive application of these same principles—their consistent, constructive application to the results of previous applications of the very same principles—leads reason beyond the original data of human experience, and into applications that extend unrestrictedly beyond experience, and ultimately into "obscurity and contradictions" (*Dunkelheit und Widersprüche*). Alas.

What kind of obscurity and what kind of contradictions? Here there is a strong anticipation by Kant of the logico-metaphysical phenomenon of what contemporary philosophers of logic and mathematics call *impredicativity*, and even more precisely, of what can be called *vicious impredicativity*, both of which we have already briefly surveyed in the famous cases of Frege and Russell, in section IV.

Impredicativity is the construction or definition of sets or totalities of objects in terms of, or by reference to, those very totalities themselves. For example, the set or totality of all non-bachelors is itself a non-bachelor, and thus belongs to the membership of that very set or totality. On the other hand, the set or totality of bachelors is not a member of the set or totality of bachelors. So some sets or totalities are members of themselves, and some sets or totalities are not members of themselves. Let us call the self-membership of the set or totality of non-bachelors an instance of *benign impredicativity*, and let us call the rule which says that totalities may permissibly be

constructed or defined by impredicative operations, *The Principle of Benign Impredicativity.* Correspondingly, the construction or definition of sets or totalities according to The Principle of Benign Impredicativity seems perfectly logically and mathematically legitimate.

Vicious impredicativity, by sharp contrast, is the generation of logical paradoxes, hyper-contradictions, or instances of *dialetheia* (i.e., propositions such that, logically necessarily, they are true if and only if they are false, hence logically necessarily they are both true *and* false, a.k.a. "truth-value gluts") by means of impredicative reasoning. In order to make the notion of vicious impredicativity vivid, we need only consider, again, the famous Fregean example of vicious impredicativity in set theory, which, for the Frege of *Basic Laws of Arithmetic*, had not only logically catastrophic results, but also almost literally Hecuba-like personal consequences. As Frege wrote to Russell, who had discovered the contradiction in basic law V of Frege's logical system (a.k.a. The Unrestricted Comprehension Principle for naïve set theory):

> Your discovery of the contradiction has surprised me beyond words and, I should like to say, left me thunderstruck because it has rocked the ground on which I mean to build arithmetic…. It is all the more serious as the collapse of my law V seems to undermine not only the foundations of my arithmetic but the only possible foundations of arithmetic as such.[74]

Russell's discovery was not only logically catastrophic and personally tragic for Frege. It was also, from the standpoint of the CPR and its critical meta-philosophy, *meta-philosophically ironic.* This is because Frege's goal in *Basic Laws* was precisely to provide an explanatory and ontolological reduction of arithmetic to logic and thereby, by showing that arithmetic truths are analytic truths, refute *Kant's* thesis that arithmetic is synthetic a priori. Nevertheless, as we have just seen, Frege's attempt to resuscitate the project of classical Rationalist metaphysics against Kant in the guise of Logicism ended

in *Dunkelheit und Widersprüche*. Here is a brief Kantian diagnosis of how the Fregean disaster happened.

In order to reduce arithmetic to logic, Frege presupposed and used naïve set theory as the *reducing* theory, and characterized numbers as sets of all sets whose membership can be put into one-to-one correspondence with each other (equinumerosity). We have seen that by an application of The Principle of Benign Impredicativity to the objects of ordinary experience, there are some sets or totalities that *are* members of themselves, like the set or totality of all non-bachelors, and that there are other sets or totalities that are *not* members of themselves, like the set or totality of bachelors. But now, by a recursion that clearly extends beyond the scope of human experience, what about the set or totality consisting of *all* sets or totalities that are not members of themselves? Call this non-experiential set or totality K, and now let us try to apply The Principle of Benign Impredicativity to it. If K is a member of itself, then it is not a member of itself. But if K is not a member of itself, then it is a member of itself. So, by recursively applying beyond the limits of human experience what seemed to be the otherwise perfectly legitimate Principle of Benign Impredicativity, we discover to our rational dismay that K is a member of itself if and only if it is not a member of itself. In other words, *Dunkelheit und Widersprüche* galore, and we directly bear witness to the fact of pure reason's self-alienating, self-destructive, and self-inflicted descent into barking madness.

This is a particularly crisp and vivid example of what Kant calls *the Antinomy of Pure Reason*, the discovery of which in 1766, as we saw above, was the third basic source of Kant's revolutionary anthropocentric turn to the mitigated rationalism of

Transcendental Idealism. In 1903 and 1908 Russell called instances of the antinomy of pure reason *The Contradictions*, but *like* Frege and *unlike* Kant, Russell also refused to trace their generation to the innate constitution of human reason itself and its natural "human, all too human" psychological tendency to engage in what I will call *transcendent impredicative reasoning*.

Transcendent impredicative reasoning is the iterative self-including construction of higher-order set or totalities whose first-order membership does *not* consist exclusively of elements that are actual or possible objects of human experience—leaving aside *the empty set*, or "empty intuition without an object," which is a special representation of pure intuition that is also minimally contained in every non-empty first-order set or totalitity whatsoever (*CPR* A290-292/B346-349). Sets or totalities created by transcendent impredicative reasoning are what I will call *ill-founded* or *noumenal* sets or totalities. This in turn leads to the corresponding notion of a *well-founded* or *phenomenal* set or totality:

> A set or totality is well-founded or phenomenal if and only if either (i) every element of its membership (leaving aside the empty set) is an actual or possible object of human experience, or (ii) all its iterative self-including constructions necessarily presuppose that every element of its first-order membership (leaving aside the empty set) is an actual or possible object of human experience.

Kant ultimately argues in the Dialectic of Pure Reason that no well-founded or phenomenal set or totality is antinomous, whereas at least *some* of the ill-founded or noumenal sets or totalities, and possibly *all* of them, are antinomous.

Russell's refusal to trace the genesis of The Contradictions to the innate constitution of human reason and its natural psychological tendency to engage in transcendent impredicative reasoning, like Frege's similar refusal, stemmed directly from

the explicitly anti-Kantian metaphysical and epistemic commitments of Russell's

Logicism:

> [T]he Kantian view . . . asserted that mathematical reasoning is not strictly formal, but always uses intuitions, *i.e.* the *à priori* knowledge of space and time. Thanks to the progress of Symbolic Logic, especially as treated by Professor Peano, this part of the Kantian philosophy is now capable of a final and irrevocable refutation. (Russell, *Principles of Mathematics*, p. 4)

> Ever since I abandoned the philosophy of Kant . . . I have sought solutions of philosophical problems by means of analysis; and I remain firmly persuaded . . . that only by analysing is progress possible. (Russell, *My Philosophical Development*, pp. 14-15)

Just as Frege's Logicism had failed because of The Contradictions' emergence in

basic law V, so too Russell's Logicism failed for a somewhat different—although not so

very different—reason. In 1931, Kurt Gödel proved two seminal Incompletness theorems

which show that all classical second-order logical systems, like Russell and Whitehead's

system in *Principia Mathematica*, which contain enough of Peano's five axioms for

arithmetic, must also contain logically unprovable sentences (in effect, self-referring

versions of The Contradiction formally equivalent to the Liar Paradox) and are therefore

not only

> (i) incomplete (i.e., not all of their tautologies are theorems),

but also

> (ii) consistent if and only if the ground of truth for such systems is outside the system itself.

Gödel's seminal Incompleteness results are formally highly analogous to Kant's

dialectical logical analysis of the Antinomy of Pure Reason, which shows that classical

Rationalist metaphysics logically entails antinomies if and only if it is assumed that there

is no fundamental ontological difference between appearances or phenomena on the one

hand, and things-in-themselves or (positive) noumena on the other. Clearly, there is some

deep and essential logico-metaphysical connection, discovered by Kant, and later re-discovered by Gödel, between

(i) collapsing the fundamental ontological difference between appearances or phenomena on the one hand, and things-in-themselves or (positive) noumena on the other,

(ii) transcendent impredicative reasoning and its vicious impredicativity,

(iii) The Contradictions,

(iv) Incompleteness,

and

(v) anti-Logicism.

In any case, both Frege's and Russell's versions of Logicism ultimately failed because of their inability to control the logically explosive power of The Contradictions, and to constrain transcendent impredicative reasoning together with its vicious impredicativity. It is now known, however, that Peano arithmetic is logically provable in classical second-order logic together with a conceptual replacement for the Frege's attempted reduction of numbers to sets, known as *Hume's Principle*:

The number of Fs = the number of Gs if and only if there are exactly as many Fs as there are Gs.

This is the logico-metaphysical foundation of the late 20th century project of *Neo-Logicism*. What remains an importantly open logico-metaphysical question, however, is whether second-order logic and Hume's Principle are *analytic* or *synthetic a priori*. If it is the latter, then Kant was right after all, and as a consequence both Logicism and Neo-Logicism, just like classical Rationalist metaphysics, were and are tragically mistaken. But the whole debate concerning The Contradictions, Incompleteness, and the philosophical fate of Logicism and Neo-Logicism is made almost impossibly difficult to resolve by the sad historico-philosophical fact that even now, 230 years after the

publication of the CPR, there is neither a generally accepted theory of the analytic-synthetic distinction, nor a generally accepted theory of the nature and status of logic. Again, alas.

To summarize the philosophical plot up to here. Kant's first two main points in the A Preface, reprised and worked out in full detail in the the Antinomy of Pure Reason, are these:

> (1) Classical Rationalist metaphysics is inherently *capable* of generating an antinomy of pure reason precisely because the innate constitution of the faculty of human reason provides for a natural psychological tendency to engage in transcendent impredicative reasoning.

and

> (2) The faculty of human reason on its own, without appealing to any *other* basic human cognitive faculty—e.g., the faculty of sensibility or *Sinnlichkeit*—is also inherently *incapable* of comprehending what has gone wrong in its reasoning processes when it generates an antinomy of pure reason.

Indeed, Kant even goes so far here as to *identify* metaphysics with reasoning that satisfies these two conditions:

> The battlefield of these endless controversies is called metaphysics. (*CPR* Aviii)

Now it is bad enough, and certainly philosophically tragic, that not only classical Rationalist metaphysics but also its modern analogue, Logicism, both commit cognitive suicide by means of transcendent impredicative reasoning, its vicious impredicativity, and the antinomy of pure reason. But its inevitable self-ignorance about how and why this is happening is *also* a direct violation of the original Socratic philosophical imperative to "know thyself." Hence at this point, one might well think: "Well, so much the worse for human reason, and its supposedly innate faculty for infallible clear and distinct rational intuition!" And like Hume in the first *Enquiry*, one might well think that

as a consequence we should simply *burn* all classical Rationalist metaphysics and early

Analytic metaphysics *at the stake*:

> When we run over libraries, persuaded of these [Empiricist] principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.[75]

But although in 1771 or 1772 Kant was indeed philosophically awakened and enlightened

(for the third time in a six-year span) by Hume, nevertheless he was never *mesmerized* by

Hume's skeptical Empiricism. So the very fact that human reason has this "peculiar fate"

(*besondere Schicksal*) driving it towards cognitive self-annihilation, according to Kant,

also ultimately drives self-critical human reason through and beyond Hecuba's tragedy

towards a tragic rational *catharsis* in Aristotle's sense in the *Poetics*, i.e., towards a

purging of otherwise harmful emotions for the sake of self-knowledge. And by means of

this process of rational catharsis, as self-clarifying, self-critical, self-knowing

metaphysicians and knowers, we rid ourselves of the very *need* to transcend human

experience, and fully accept our "human, all too human" finitude. According to Kant we

thereby become *mitigated rationalists*, not Humean Empiricist *mitigated skeptics* about

human reason. I will have more to say about this fundamental contrast between Kant's

"mitigated rationalism" on the one hand, and Hume's Empiricist "mitigated skepticism"

on the other, directly below.

Back now to unfortunate Hecuba, fallen "queen of all the sciences," namely

classical Rationalist metaphysics. Within this metaphorical frame, now rhetorically

extended to an analogy with the political history of pre-Enlightenment despotism, Kant

sketches an apocalyptic history of the rise and fall of 17th and 18th century European

metaphysics. The basic details are these. The metaphysical theories of classical 17th or

early 18th century Rationalists like Descartes, Spinoza, Leibniz, Wolff, and Baumgarten were dogmatic and needlessly strict ("despotic," and retaining "traces of ancient barbarism"). As a consequence, classical Rationalism was wide open to highly effective critical attacks by a few Empiricist skeptics, especially Hume ("nomads"). But although the Rationalists did attempt to reply to these attacks, they did so only in a piecemeal way, and without any decisive reply to the skeptics ("continually attempting to rebuild, though never according to a plan unanimously accepted among themselves"). The result was philosophical confusion and endless controversy.

For a brief period in the late 17th century, however, it seemed that Locke's 1689 *Essay Concerning Human Understanding* might resolve the controversies, by the application of its positive Empiricist epistemology and philosophical psychology to the analysis of the human faculty for understanding or *Verstand* ("a certain physiology of the human understanding"). Yet although Locke's *Essay* plausibly argued that necessarily, all human cognition has its causal origins in human experience ("the birth of the purported queen was traced to the rabble of common experience"), his overall positive Empiricist line of argument ultimately failed because of some basic flaws ("this genealogy was attributed to her falsely"), and classical Rationalist metaphysics was able simply to re-assert its basic principles and stubbornly hold the line ("metaphysics fell back into the same old worm-eaten dogmatism"). The result of this unresolved Rationalist-Empiricist dialectic was that by the time of the CPR—i.e., the latter half of the 18th century—many or even most uncommitted philosophers were intellectually bored to death by the whole enterprise and simply refused to take sides, or to undertake any fruitful or new philosophical work ("what rules is tedium and complete **indifferentism"**), so that the

science of metaphysics was, in effect, going to hell in a handbasket ("the mother of chaos and night in the sciences").

At this point, before going on to look at Kant's extremely interesting philosophical riff on "indifferentism"—i.e., *agnosticism*, which we have encountered already in section III—I want to pull back for a moment and ask two critical questions:

> (1) Is Kant's apocalyptic history of the rise and fall of 17th and 18th century metaphysics in fact *accurate*?

> (2) What, according to Kant, precisely *were* the basic mistakes in Locke's "*physiology* of the human understanding"?

In answer to the *first* question, I do think that *yes*, Kant's story about the rise and fall of 17th and 18th century metaphysics *is* in fact quite accurate, assuming that one takes a suitably high-level point of view on the philosophical doctrines and texts he is talking about. It is of course possible to zoom in, study classical 17th and 18th century Rationalist and Empiricist metaphysics from a greatly resolved perspective, track sideways across the several metaphysical theories, note many non-trivial differences between them, deploy both interpretive charity and revisionist zeal, and then discover important tensions between what is actually written there, on the ground, and Kant's high-level philosophical picture of it—and, in fact, discover important tensions between what is actually written there, on the ground, and *any* high-level philosophical picture of it. *But the very same hermeneutic point holds of every piece of good and interesting philosophy that was ever or will be written*, and therefore the scholarly zoom-in and track-sideways strategy does not materially affect the high-level truth of Kant's claims. The science of metaphysics *was* going to hell in a handbasket in the latter half of the 18th century, and for *precisely* the reasons Kant gives.

In answer to the `second` and much more philosophically important question, according to Kant, Locke's positive Empiricist theory of human cognition makes two basic mistakes.

`First`, according to Kant, Locke's theory confuses a true strong modal claim about *the necessary causal origins* of the content, truth, and justification of human cognition with a false strong modal claim about *the strict determination* of the content, truth, and justification of human cognition. Even if it is true that necessarily, all human cognition *begins in* causally-triggered sensory experience, it simply does not follow that necessarily, the content, truth, and justification of all human cognition is *determined by* causally-triggered sensory experience. This is because it remains really possible that even though necessarily, all human cognition begins in causally-triggered sensory experience, nevertheless, necessarily, the content, truth, and justification of human cognition are all at least partially determined by *non*-sensory, *non*-empirical, and *non*-contingent factors. And that, as we have seen, is just what Kant wants to claim by asserting the truth of Transcendental Idealism. But Locke's Empiricist theory of human cognition asserts the denial of this claim. So if Kant is right, then Locke's theory of human cognition commits the serious fallacy of confusing the necessary causal origins of human cognition with the strict determination of human cognition as to its inherently normative content-based, truth-based, and justification-based features. In other words, Locke's theory commits the serious fallacy of what we would nowadays call *Psychologism*.

`Second`, according to Kant, Locke's theory of "primary qualities" and "secondary qualities," along with Locke's corresponding theory of *ideas* of primary qualities and *ideas* of secondary qualities, is false. The falsity of Locke's primary-secondary doctrine

is of course a standard thesis of Hume's skeptical Empiricism, and also of Berkeley's

skeptical metaphysical phenomenalism. But there are two essentially ways of denying

Locke's theory. It is one thing to deny Locke's theory by holding that

(i) All primary qualities are ultimately nothing but secondary qualities.

and

(ii) All ideas of primary qualities are ultimately nothing but ideas of secondary qualities.

That is exactly what Hume and Berkeley hold. Nevertheless, it is something

fundamentally different to deny Locke's theory by holding that

(i*) There are no really such things as either primary qualities or secondary qualities.

and in addition that

(ii*) There are really no such things as either ideas of primary qualities or ideas of secondary qualities.

precisely because

(iii) The one and only world that rational human beings do in fact objectively cognize, and can ever possibly objectively cognize, by means of ordinary sense perception, ordinary empirical judgment, and also by means of the exact sciences and their several different kinds of synthetic a priori judgments, is necessarily filled with real appearances or phenomena exclusively, and necessarily never contains any things-in-themselves or (positive) noumena.

And that is exactly what Kant holds in the CPR.

We left the fallen queen of all the sciences—classical Rationalist metaphysics in

the second half of the 18th century—on the verge of descending into chaos and night like

unfortunate Hecuba, due to the "indifferentism," or agnosticism, of most contemporary

philosophers. Kant's extremely interesting thought here is that the very same

indifferentism or agnosticism which threatens to drive philosophy down to cognitive

Hades, also contains at least the *anticipation* of a revolutionary philosophical turn

("transformation and enlightenment"). But in order to spell out Kant's extremely

interesting thought, we need to get a little clearer on the crucial cognitive notion of *agnosticism*, which as I mentioned, we have encountered already in section III.

Generally speaking, agnosticism is the negative doxic attitude of *comprehensive non-belief* or *suspension of judgment*. More specifically characterized,

(Agnosticism) A cognitive subject S is agnostic or indifferent that P if and only if S does not believe that P and S does not believe that not-P.

As characterized in this way, agnosticism is then to be contrasted with the slightly more limited negative doxic attitude of *doubt*. More specifically,

(Doubt) A cognitive subject S doubts that P if and only if either S does not believe that P or S believes that not-P.

Thus every form of agnosticism includes doubt, but not all forms of doubt are agnostic.Both Locke and Hume had fruitfully explored the negative doxic attitudes of agnosticism and doubt, but Kant pushes those explorations significantly beyond those of the classical Empiricists.

One form of agnosticism, which for clarity's sake I will dub *constructive agnosticism* or *constructive indifferentism*, consists in a doxic attitude of cognitive neutrality or open-mindedness, and in a cautious refusal to take a stand until all the relevant evidence is in. Constructive agnosticism or constructive indifferentism follows naturally from the act or process of reasonable doubt that Hume characterizes as *mitigated* skepticism. Mitigated skepticism leading to constructive agnosticism or constructive indifferentism is fully appropriate in everyday reasoning, exact science, and legal contexts, e.g., in an ordinary well-conducted court of law. In such contexts, the presiding judge should deploy reasonable doubt and remain in a doxic attitude or state that is constructively agnostic or constructively indifferent until it is appropriate to form a judgment, given all the relevant evidence.

This, in turn, is sharply opposed to what I will call *destructive agnosticism* or *destructive indifferentism*, which consists in a doxic attitude of cognitive paralysis or self-annihilating confusion (i.e., "obscurity and contradictions," and "chaos and night in the sciences"), or what the Stoic radical skeptics called "equipollence." The end result of destructive agnosticism or destructive agnosticism is, in effect, *doxic toxicity*, i.e., complete intellectual ambivalence and apathy, and the refusal to proceed in appropriate evidential circumstances from cognitive neutrality or open-mindedness to assertoric judgment. Destructive agnosticism or destructive indifferentism follows naturally from the corrosive and radical form of skepticism that Hume dubs *Pyrrhonian* skepticism, which seeks to undermine the grounds of all rational belief, and which in fact is *a direct attack on human rationality itself*. As we have already seen, the underlying cause of destructive agnosticism or destructive indifferentism in philosophy, and equally the pathological mechanism of Pyrrhonian radical equipollence skepticism in its application to classical Rationalist metaphysics, is *transcendent impredicative reasoning, its vicious impredicativity, and the antinomies of pure reason*. So Kant is saying that this is the cognitive hell where metaphysics as a science *could* end up forever if some radical metaphysical therapy is not carried out, fast, before it is too late.

According to Kant, this radical metaphysical therapy turns on a *third* form of indifferentism or agnosticism, i.e., a form of comprehensive non-belief or suspension of judgment that is in sharp contrast to *either* constructive *or* destructive agnosticism or indifferentism. This essentially more robust negative doxic attitude, which I called *radical agnosticism* in section III, and which we can also call *radical indifferentism* in the present context, is in fact a higher-order state of negative self-knowledge—i.e., of

knowing that it is impossible for me to know some first-order object or state of affairs one way or the other. Call any such first-order object, X, and call any such first-order state of affairs, the state of affairs that P. Then Radical Agnosticism or radical indifferentism about X and the state of affairs that P can be explicitly defined in the following way:

> (Radical Agnosticism) A cognitive subject S is radically agnostic about X or about the state of affairs that P if and only if S knows that it is impossible for S to know whether it is either true of X that P or false of X that P.

Kant holds that the mature rational human faculty or power of judgment, or *Urteilskraft*, is capable of advancing to Radical Agnosticism or radical indifferentism about the "illusory knowledge" (*Scheinwissen*) dogmatically postulated by classical Rationalist metaphysics. That is, our faculty of judgment can advance to the higher-order negative self-knowledge that it is impossible for us to know any objects or states of affairs *beyond the limits of human experience*, one way or the other. For example, as I argued in section III, we can have the higher-order negative self-knowledge that it is impossible for each one of us to know whether an omnipotent, omniscient, and all-good (a.k.a. "a 3-O") God exists or does not exist, whether noumenal agent-causal libertarian freedom exists or does not exist, and whether immortality of the soul exists or does not exist. Similarly, as I noted above—and not altogether coincidentally, in view of what we already know about the explicitly anti-Kantian project of Logicism—Gödel proved in 1931 that there are unprovable sentences in classical second-order logic plus enough of Peano's five axioms for arithmetic. Consider, now, any such unprovable sentence of Peano arithmetic. By means of Gödel's Incompleteness proofs, we have the higher-order logical self-knowledge that it is impossible for us to know, *by means of proof-theoretic means alone*, whether this sentence is true or false.

Given Gödel's radically agnostic 1931 result, it nevertheless remains possible that the truth of unprovable sentences of Peano arithmetic is knowable by *non*-proof-theoretic means, e.g., by *intuition*, perhaps in Descartes's classical sense of clear and distinct intuition. In the 1960s, Gödel asserted that the truth of unprovable sentences of Peano arithmetic, and also the truth of the seemingly unprovable Continuum Hypothesis, could be known by means of *conceptual intuition*, although he did not adequately explain *how* this conceptual intuition is possible. But Kantian intuition or *Anschauung*, by sharp contrast to Gödel's 1960s notion of intuition , is essentially *non*-conceptual. So, given Gödel's radically agnostic 1931 result, it *also* remains possible that the truth of unprovable sentences of Peano arithmetic, and also the truth of The Continuum Hypothesis, are knowable by *essentially non-conceptual a priori intuition* in Kant's sense.[76]

As we have seen, the inherently reasonable process of mitigated skepticism leads to constructive agnosticism or constructive indifferentism, and the inherently *anti*-reasonable process of Pyrrhonian radical skepticism leads to destructive agnosticism or constructive indifferentism. For Kant, the *paradigmatically philosophical* rational process of self-criticism that leads to the higher-order negative self-knowledge of radical agnosticism or radical indifferentism is nothing more and nothing less than *the critique of pure reason*. This rational process of self-criticism is *paradigmatically* philosophical precisely because it fully captures and also fully clarifies the epistemic force of *Socratic Ignorance*. The truly Critical philosopher knows only that she *cannot* know all and only those objects and states of affairs that transcend the limits of human experience, one way or the other.

Now stepping fully back from the rhetorical devices he has currently in play—his mythological metaphorical frame and his apocalyptic history—and addressing the philosophical reader directly, Kant states explicitly that the critique of pure reason is *not* merely a negatively self-knowing analysis of philosophical books and metaphysical systems. On the contrary, the critique of pure reason is a negatively self-knowing analysis of something far more fundamental than books and systems, namely the faculty of human reason itself, insofar as it purports to cognize "independently of all experience." If it can be known, according to the legitimate higher-order reflexive epistemic principles of the critique of pure reason, that it is impossible for us to know the sorts of objects and states of affairs that classical Rationalist metaphysics purports to know, one way or the other, then it is knowable that classical Rationalist metaphysics is impossible. Understood in this way, the critique of pure reason is *not* directly analogous to an ordinary or first-order court of law in which a human animal possessing the faculty of reason is the presiding judge about ordinary first-order experiential objects and matters of fact. On the contrary, the critique of pure reason is instead directly analogous to a *higher-order reflexive* court of law in which the faculty of human reason *itself* is being critically examined and judged *by the faculty of human reason itself*. This new metaphorical frame, which presents an analogy between the critique of pure reason and a higher-order reflexive court of law, re-appears at various points in the CPR.

In this particular connection, however, one important critical question needs to be noted:

> How can human reason critically examine and judge *itself* without there being a fundamental rational incoherence—that is, without there being, in effect, a fundamental epistemic "conflict of interest" whereby one and the same faculty is simultaneously in a state of *self-knowledge* and also in a state of *self-ignorance*?

The answer is that this *would* be a problem for Kant *only if* it were assumed by him that the faculty of human reason is structurally *flat* or linear. But according to Kant, human reason is inherently hierarchical or multi-level, and *not* flat or linear. In other words, human reason is inherently structured so as to be *both* first-order and *also* higher-order reflexive, up to any *nth* higher level. As long as we do not confuse the several distinct levels or orders within human reason's innate constitution, then there is no rational incoherence. Indeed rational incoherence results precisely from an attempt to *collapse* all the levels or orders downwards into one flat or linear structure, and thereby to *deny* that human reason has an essentially hierarchical reflexive architecture. Here Kant's Transcendental Idealist conception of the rational self as *inherently hierarchical, reflexive, and spontaneously dynamic* should also be sharply contrasted, on the one hand, with the classical Rationalist conception of the self as a *flat unchanging thinking substance*, and, on the other hand, with Hume's skeptical Empiricist conception of the self as an *inherently unstructured, epiphenomenal and causally inert, temporally successive bundle of impressions and ideas*.

At this point in the A Preface—*CPR* A xii to xiv—Kant makes several very strong claims on behalf of his revolutionary philosophical project, the critique of pure reason, and his big book, the CPR, alike: This is the *only* philosophical method by which classical Rationalist metaphysics can be adequately analyzed and criticized. And in this big book he has *completely* avoided the mistakes by which non-experiential human reason commits cognitive suicide. He has completely spelled out the questions, mentioned in the first paragraph of the Preface, which human reason puts to itself and cannot answer when it is being deployed non-experientially, according to principles. And

he has also answered these questions "to reason's full satisfaction." Indeed, his chief aim has been a *synoptic* understanding of human reason, and correspondingly he asserts that he has solved *all* the basic problems of classical Rationalist metaphysics.

This may appear to be nothing but sheer philosophical arrogance, and an unconscious lapsing back into the very same sort of classical Rationalist metaphysical dogmatism that Kant intends to be criticizing. But he is fully aware of this worry, and has two pre-emptive responses to it.

First, the structure of human reason in its non-experiential use (a.k.a. "pure reason") is not only inherently hierarchical, or multi-level: it is also inherently *holistic*: "a perfect unity." Its internal structure is therefore such that no one of its principles could be correct without *all* of its principles also being correct, and on the other hand, no one of its basic problems can be solved without also solving *all* of them. So assuming that the critique of pure reason yields even *one* positive result, then it also suffices to yield a *complete set* of such results. Again, as Gödel showed formally and rigorously in 1931, but as Kant importantly anticipates in 1781, an inference from systematicity to completeness is sound when and only when a system's ground of truth is explicitly *outside* that system and also constrains the ontic and semantic scope of that system in a *well-founded* way that inherently heeds the fundamental ontological difference between phenomena and noumena.

Second, as we have already seen, all the self-inflicted problems of classical Rationalist metaphysics stem from a single source—namely, *transcendent impredicative reasoning*, and the viciously impredicative attempt, leading to antinomies, to apply rational principles that are adequately legitimated by human experience, *beyond the limits*

of human experience. But the critique of pure reason is based entirely and explicitly on the *rejection* of transcendent impredicative reasoning, and also on psycho-meta-philosophically diagnosing and studying, step-by-step, the way in which an uncritical human reason can thereby tragically self-destruct. So it cannot possibly be the case that the critique of pure reason has committed the very same error that it was originally set up to detect and rectify.

On the basis of these two responses, Kant says that he is sufficiently warranted in concluding that he has established the primitive content of the critique of pure reason, not only as to its basic ends or purposes and the exhaustive articulation of *each* member of this set of ends, but also as to the comprehensiveness of the *total* set of ends. These, he thinks, are a mirror of the internal constitution of the human cognitive faculty.

Now in the course of formulating these responses and then drawing this very strong conclusion, Kant provides two other reasons for holding that the critique of pure reason is inherently more reliable than and essentially distinct from the sort of classical Rationalist metaphysical reasoning that purports to prove, e.g., the simplicity of the soul, or the thesis that the world necessarily has a first cause:

> I have merely to do with reason itself and its pure thinking; to gain exhaustive acquaintance with them I need not seek far beyond myself, because it is in myself that I encounter them, and general logic (*gemeine Logik*) also already gives me an example of how the simple acts of reason may be fully and systematically enumerated; only here the question is raised how much I may hope to settle with these simple acts if all the material and assistance of experience are taken away from me. (*CPR* Axiv)

In other words,

> (1) The critique of pure reason bases its claims on rational *self-reflection* or rational *introspection*, not on claims about anything outside the self.

and

(2) The critique of pure reason presupposes, as a fundamental guide, the complete analysis of reason's "simple acts" that can be found in what Kant later calls *pure general logic*.

It is clear enough, then, that Kant is making two crucial assumptions here:

(1*) The Reflexive Transparency Assumption: The representational content, formal structure, and intentional activity of reason is veridically accessible to reason itself by means of self-reflection or introspection.

and

(2*) The Pure General Logic Assumption: Pure general logic lays bare the complete formal structure of human theoretical rationality itself

The second assumption, in order to be understood, requires that we remind ourselves of some points I already covered in section IV.

For Kant, *logic* is the categorically normative, strictly universal, law-governed a priori science of *the laws of thought*—the complete set of rules or principles specifying how rational human animals *ought to think*, in order to conform to the highest or supreme standards of human reason itself. *Analytic logic* is the categorically normative, strict-universal, law-governed a priori science of consistent thinking, conceptually necessary truth, and valid inference (i.e., consequence). *Dialectical logic* is the categorically normative, strictly universal, law-governed a priori science of inconsistent thinking, falsity or illusion, and fallacy (i.e., invalidity or non-consequence). *Pure* logic is absolutely non-empirical logic, whether analytic or dialectical. *Applied* logic is logic, whether analytic or dialectical, under actual, empirical conditions of thinking, i.e., the empirical psychology of logic. *General* logic is logic, whether analytic or dialectical, whose consistency/inconsistency, conceptually necessary truth/falsity or illusion, and validity/fallacy does not metaphysically depend on, and therefore is not necessarily determined by, the comprehensions or *Umfangen* of objects or states of affairs designated

by propositions, singular terms, concept-terms, or discourse more generally, but which, at the same time, necessarily comprehends, or is synoptic over, all actual or possible topics of discourse. And finally, *particular* or *special* logic is logic, whether analytic or dialectical, whose consistency/inconsistency, conceptual necessary truth/falsity or illusion, and validity/fallacy *does* metaphysically depend on, and therefore *is* necessarily determined by, the comprehensions of objects or states of affairs designated by propositions, singular terms, concept-terms, or discourse more generally, and therefore is necessarily *non*-comprehensive, or *non*-synoptic, over all actual or possible topics of discourse (*CPR* A52/B76).

Here, I think it would be a serious interpretive error to hold that Kant's general solution to the Antinomies of Pure Reason consists in "showing the <u>apparent conflicts</u> involved in the Antinomies to to be illusory."[77] On the contrary, Kant holds that these are *real conflicts* of human reason: the dialectic of human reason is a *natural* dialectic. To be sure, all four of the Antinomies depend on the false assumption that there is no fundamental ontological distinction between appearances or phenomena, and things-in-themselves or noumena. Or in other words, all four of the Antinomies depend on the failure to recognize the truth of Transcendental Idealism. This shows that the Thesis and Antithesis of each antinomy are both false, and thus logical *contraries*, not logical *contradictories*. Kant also argues that the third and fourth Antinomies can be re-interpreted in such a way as to make the Thesis and Antithesis both come out true, and thus consistent, but this requires making the Transcendental Idealist assumption that there *is* a fundamental ontological distinction between phenomena and noumena. The crucial points here for our purposes are that

(1) For Kant, *pure general logic* includes both *analytic* logic (the logic of truth) and also *dialectical* logic (the logic of fallacy and illusion) (*CPR* A60-62/B84-86),

and that

(2) *Insofar as* classical Rational metaphysics naturally falls into antinomy, then the hyper-contradictions of pure reason *really do exist*.

Now pure general logic is presupposed by *transcendental* logic, which adds ontological commitments to pure logic, and is a "science of pure understanding and of the pure cognition of reason, by means of which we think objects completely *a priori*" (*CPR* A57/B81). Hence Kant's transcendental *dialectical* logic is a (deviant, non-classical) *dialetheic* logic (*CPR* A63-64/B88-89).

At the same time, however, Kant's transcendental logic is also clearly and explicitly a *paraconsistent logic*. This is shown by his clear and explicit commitment in the 1785 *Groundwork of the Metaphysics of Morals* to the Universal Law Formula of the Categorical Imperative, i.e.,

Act only on that maxim by which you can at the same time will that it should become a universal law. (*GMM* 4: 421)

which entails that by the time of the 1787 or B edition of the CPR, at the very latest, Kant also accepts what in section IV I called *The Minimal Meta-Logical Principle of Non-Contradiction*, i.e.,

~ (□S) (S & ~ S)

Not every statement in every language or logical system is both true and false. This is because The Formula of Universal Law clearly does *not* say that it is impossible to believe a maxim which, when generalized, yields an inconsistent principle, since such inconsistency is precisely the criterion of its being an *impermissible* maxim, which of course we believe all the time. This obvious fact allows for *implicit* belief in contradictions, and further supports Kant's commitment to dialetheism. Instead, the

Formula of Universal Law clearly *does* say only that any maxim which can be permissibly adopted as a principle of willing, must itself not yield self-inconsistent principles. And I think we can assume that at least one maxim is permissible, namely the maxim that I hereby commit myself to obeying the Categorical Imperative right now (e.g., in the Formula of Humanity as an End-in-Itself formulation). So there must be at least one self-consistent principle, i.e., the universal generalization of at least that one permissible maxim. Hence it cannot be true that *every* sentence or statement in every language or logical system is both true and false, i.e., necessarily, *not every* sentence or statement in every language or logical system is both true and false, i.e., The Minimal Meta-Logical Principle of Non-Contradiction is true. Therefore the logic of Kant's transcendental dialectic is *a dialetheic paraconsistent logic*.

  This brings us back to Pyrrhonian radical skepticism. If there are any good reasons for *seriously doubting* either either The Reflexive Transparency Assumption or the The Pure General Logic Assumption, then Kant's critical meta-philosophical project is correspondingly in serious trouble. So we need to raise the following fundamental hard question explicitly: *Can* Pyrrhonian radical skepticism be effectively applied to either of these two crucial assumptions? If so, then on the one hand, the critique of pure reason goes down, and so does human rationality itself. But on the other hand, if Kant is able to provide sufficiently good reasons for *making* these assumptions, and if he is able to respond effectively to Pyrrhonian radical skepticism about either rational self-reflection or general logic, then the the critical meta-philosophy expressed by *the critique of pure reason* is commensurately in good shape, and so is human rationality.

And I do think that Kant has *adequate* responses to Pyrrhonian radical skepticism about *The Reflexive Transparency Assumption* and *The Pure General Logic Assumption*. Even the Pyrrhonian radical skeptic, as a minimally rational and *moral* animal, presupposes and uses The Minimal Meta-Logical Principle of Non-Contradiction, and thereby *both logically and pragmatically contradicts himself* when he either asserts or even supposes its denial. More precisely, then, the Pyrrhonian radical skeptical worry about *The Reflexive Transparency Assumption* and *The Pure General Logic Assumption* is ruled out by showing that the Pyrrhonian radical skeptic *himself* necessarily presupposes *both*; hence in denying or in even merely supposing the denial of *either*, he both logically and pragmatically contradicts himself and thereby commits cognitive suicide. As Kant puts it in the *Vienna Logic*:

> Proceeding skeptically <u>nullifies all our effort</u>, and it is an antilogical principle… For if I bring cognition to the point where it nullifies itself, then it is <u>as if we were to regard all human cognition as nothing.</u> (*VL* 24: 884, underlining added)

In this way, as Kant clearly sees, the fundamental problem with Pyrrhonian radical skepticism, no matter how sophisticated it might be, is that it implies *self-annihilating logical nihilism about human reason*.

The great length of CPR necessitated deleting many "examples and illustrations" that were included in the first draft. Still, Kant thinks that such examples and illustrations are necessary really only in books intended for a "popular" (i.e., relatively well-educated, and middle class or better) audience. And philosophical experts will not need the extra intellectual labor-saving devices, which, as nice as they might be, may in fact impede their understanding of the CPR. But why would this happen? Kant then cites the French philosopher Jean Terasson's witticism to the effect that if the true size of a book is to be gauged by the time needed to understand it, then many books would have been much

shorter if they had not so short (*CPR* Axix). Kant's spin on this, with regard to "the intelligibility of a whole of speculative cognition that is wide-ranging yet is connected in principle"—i.e., the intelligibility of *a metaphysical system*—is that many books would have been much clearer if they had not been presented so clearly. In other words, it is possible even for philosophical experts to fail to grasp the overall "articulation and structure" of a metaphysical system, through being distracted by examples and illustrations which are intended only to illuminate its proper parts. Indeed, attaining "a survey of the whole" (*Überschauung des Ganzen*) is essential when properly evaluating the coherence and truth of the metaphysical system. Clearly Kant is using an *architectural* analogy here. Grasping a metaphysical system is like looking at *the complete working drawings* of a house.

This, in turn, confers a special cognitive-semantic status on metaphysical treatises. For they are the only scientific works that hold out the prospect of actually *completing* the intellectual projects on which they are based, at least in as far as their essential content is concerned. This is because metaphysical treatises, if correct, specify *complete* systems of pure human reason. Here Kant again deploys, at least implicitly, The Reflexive Transparency Assumption and The Pure General Logic Assumption:

> Nothing here can escape us, because what reason brings forth entirely out of itself cannot be hidden, but is brought to light by reason itself [i.e., The Reflexive Transparency Assumption is true] as soon as reason's common principle (*gemeinschaftliche Prinzip*) has been discovered. The perfect unity of this kind of cognition, and the fact that it arises solely out of pure concepts without any influence that would extend or increase it from experience or even particular intuition, which would lead to a determinate experience, make this unconditioned completeness not only feasible but necessary. [I.e., The Pure General Logic Assumption is true.] (*CPR* A xx)

The quotation from Persius—*Tecum habita, et noris quam sit tibi curta supellex*, i.e., "Dwell in your own house, and you will know how simple your possessions are"—then nicely closes out the architectural metaphor. The reflexively transparent relation of

human reason to itself, via pure general logic, is just like an architect who lives in the very house he has designed and built with his own hands.

In this connection it is satisfyingly ironic and also historically apt that early Wittgenstein, in the Preface to the *Tractatus Logico-Philosophicus*, 140 years after the publication of the CPR, like Kant, *also* claimed to have finally brought metaphysics and pure logic to their completion, and again like Kant, *also* claimed that this was a reflexively transparent and even trivial enterprise:

> [T]he truth of the thoughts communicated here seems to me unassailable and definitive. I am, therefore, of the opinion that the problrms have in essentials been finally solved. And if I am not mistaken in this, then the value of this work secondly consists in the fact that it shows how little has been done when these problems have been solved. (*TLP* 29)

Then Wittgenstein gave up philosophy, and as his next major project in the mid-1920s, designed and built a house with his own hands, the *Haus Wittgenstein*. But it is equally satisfyingly ironic and also historically apt that Wittgenstein's eldest sister, Hermine, later said this about the *Haus*:

> Even though I admired the house very much, I always knew that <u>I neither wanted to, nor could, live in it myself</u>. It seemed indeed to be <u>much more a dwelling for the gods than for a small mortal like me</u>.[78]

The underlined phrases ironically underline the very same "anthropocentric turn" point that Kant makes in the B edition CPR's Baconian Motto. And it is also my segue for turning now to Wittgenstein's critical meta-philosophy in the *Tractatus* and the *Investigations*.

V.2 Wittgenstein and Philosophical Analysis

According to the conception of philosophical analysis dominant in the writings of Frege, Moore, Russell, and early Wittgenstein himself from the late 1870s to the mid-1920s, which I will call *The Logical-Decompositional Theory of Analysis*,

(1) Analytic propositions are necessary a priori logical truths.

and

(2) Analysis is the process of

(2.1) logically decomposing analytic propositions[79] into conceptual or metaphysical simples which are mind-independently real yet immediately and infallibly apprehended with self-evidence,

and then
(2.2) rigorously logically reconstructing those propositions by formal deduction from

(a) general logical laws

and

(b) premises that express logical definitional knowledge in terms of the simple constituents.

But in the *Philosophical Investigations*, the later Wittgenstein's devastating critique of the semantic and logical doctrines of his own earlier philosophical self in the *Tractatus* motivates a radically wider and more open-textured conception of analysis. At the same time, his self-critique of Tractarian solipsism—commonly known as *The Private Language Argument*—further radicalizes his conception of analysis by rejecting several of the fundamental assumptions of classical Rationalist Cartesian epistemology and metaphysics that had been explicitly or implicitly retained by Frege, Moore, Russell, and the author of the *Tractatus*. Indeed, Wittgenstein's radical transformation of philosophical analysis goes significantly and seriously beyond the Analytic tradition, and also returns us full-circle to Kantian *transcendental* critical meta-philosophy.

V.2.1 Philosophical Analysis in the *Tractatus*

In the *Tractatus*, early Wittgenstein importantly extends the Frege-Moore-Russell conception of logical-decompositional analysis. According to the Tractarian account, the proper targets of logical-decompositional analysis are propositions. Logical analysis

consists in completely and uniquely decomposing propositional symbols into their constituent simple symbols, whether names of objects or logical constants (*TLP* 3.23-3.261, 49). Objects are known by direct cognitive acquaintance (*TLP* 2.0123-2.01231, 33), and logical constants are known "transcendentally," or by means of a priori showing (*TLP* 4.12-4.1213, 79). Every proposition has a unique and complete decomposition (*TLP* 3.25, 49). The way in which those names are configured into a propositional structure is made manifest through the process of analysis itself. Logical analysis is thus essentially a series of logical "elucidations" (*Erläuterungen*). More than that, logical analysis is essentially the *activity* (*Tätigkeit)* but not the *theory* (*Lehre*) of decomposing a proposition into its simple constituent symbols (*TLP* 4.112, 77).

This activist conception of logical analysis has the significant virtue of avoiding The Paradox of Analysis. According to The Paradox of Analysis:

> (1) If an analysis is true then it must be uninformative and trivial, because it is merely definitional and based on the identity of concepts.

But:

> (2) If an analysis were non-trivial and informative, then it would also be non-definitional and entail the non-identity of concepts, hence false.

Therefore:

> (3) Every analysis is either trivial or false.

Nevertheless if it is true, as early Wittgenstein holds, that analysis is essentially a logical *activity* and not a logical *theory*, then strictly speaking an analysis is never true or false, so the dilemma is avoided. Of course I am going very quickly here, and there is *much* more to say about The Paradox of Analysis, its origins in Moore's conception of analysis, and the many different attempts that have been made to solve it by appealing to various epistemic or semantic considerations. But we should note that early Wittgenstein's

solution is striking precisely because it is *non*-cognitivist and *non*-semantic. By sharp

contrast to both early Wittgenstein's solution and also to the other classical epistemic or

semantic solutions, however, Kant's own theory of conceptual-decompositional analysis

contains the elements of an interesting *cognitive-semantic* solution to The Paradox.[80]

In any case, and more precisely, the Tractarian activist conception of logical

analysis has two basic parts, and correspondingly two basic aims.

First, the activity of analysis is a "critique of language" (*TLP* 4.0031, 63) in that

it displays the fact that most propositions and questions that have been written about

philosophical matters are not false but nonsensical (*unsinnig*) (*TLP* 4.003, 63), recognizes

that the analytic truths of logic are tautologous and non-pictorial, hence "say nothing"

(*sagen ... Nichts*) (*TLP* 6.11, 155), then asserts as fully significant only the propositions

of natural science (*TLP* 6.53, 187-189), then recognizes its own propositions as

nonsensical, and finally ends in mystical silence (*TLP* 6.54, 189). Thus the first basic aim

of Tractarian logical analysis is to articulate the difference between sense (factual

meaningfulness) and nonsense. Here we must remember that "nonsense" for early

Wittgenstein is literally *non-sense*, i.e., everything of a cognitive or semantic nature that

is other than *what* is described or pictured or "said" by atomic propositions, and thus can

be either sheer absurdity , or *meaninglessness*, in roughly the sense of Lewis Carroll's

Jabberwocky, or else it can be *meaningful non-sense* in some other non-atomic-fact-

representing, but still logically, semantically, aesthetically, or ethically important

discursive way.

Second, the activity of logical analysis is the process of logically clarifying

thoughts, consisting in a series of propositional elucidations which "make clear and

delimit sharply the thoughts which otherwise are … opaque and blurred" (*TLP* 4.112, 77). Thus the second basic aim of Tractarian logical analysis is to reveal the deep or *logico*-grammatical structure of natural language and thought, as opposed to its merely surface or *psychologico*-grammatical structure. In order to reveal the deep structure of language, Tractarian philosophers must construct and study symbolic logical systems like those developed in the *Begriffsschrift* and *Principia Mathematica*. Such symbolic systems are "ideal" in the sense that the syntax of a *Begriffsschrift*-type notational system itself displays, encodes, or mirrors the deep structure of natural language and thought, and thereby also the deep structure of the world of facts that language and thought represent. Even so, Tractarian analysis does not aim at the *prescriptive reform* of natural language or thought. On the contrary, everything in natural language and thought is perfectly in order, just as it is (*TLP* 5.5563, 149).

V.2.2 From the *Tractatus* to the *Investigations*

The *Tractatus* brings a definitive closure to Logical Atomism by pushing the reductive project of logical-decompositional analysis to its limits and beyond. The *Tractatus* clearly indicates or shows (even if it does not explicitly describe or say) the inadequacy of logical atomism and the very idea of Logical-Decompositional Analysis. Or at least this is how Wittgstenstein himself came to regard the *Tractatus* by the time of the *Philosophical Investigations*. Indeed, in the *Investigations* the later Wittgenstein explicitly rejects and radically re-thinks his own Tractarian theory of logical analysis:

> Four years ago I had occasion to re-read my first book (the *Tractatus Logico-Philosophicus*) and to explain its ideas to someone. It suddenly seemed to me that I should publish these old thoughts and the new ones together: that the latter could be seen in the right light only by contrast with and against the background of my old way of thinking. For since beginning to occupy myself with philosophy again, sixteen years ago [in 1929], I have been forced to recognize grave mistakes in what I wrote in that first book. (*PI* x^e)[81]

It would be quite false and misleading, however, to say that there are no continuities between the *Tractatus* and the *Investigations*. Sharply on the contrary, not only is almost every doctrine of the latter is anticipated somewhere in the former, but also the basic topics of both books are the same: language, meaning, mind, and logic. Furthermore, both books take philosophy to be a "<u>Critique</u> of language" (*TLP* 4.0031, 63, underlining added) and this in turn ultimately determines a single Kantian transcendentalist line of critical meta-philosophical argument running right through them both.

Nevertheless, there is also a very definite sense in which the *Investigations* is intended by Wittgenstein to be the antithesis of the *Tractatus*. Whereas the *Tractatus* had proposed an essentialist a priori reduction of language, meaning, and even the world itself to classical logic and to a solipsistic Cartesian and Schopenhauerian version of the Kantian transcendental subject, the *Investigations* fully sinks language, meaning, mind, and logic into the everyday actions and practices of natural-language-using human animals in their commonsense or ordinary world. The basic results of this radical move are:

> (1) Under the slogan that *logic is grammar* the pure classical logic of propositions is replaced by a strongly non-classical logic of natural language, that is at once intuitionistic (i.e., it entails the rejection of the universal law of excluded middle), non-bivalent (i.e., it permits vagueness in its predicates), non-monotonic (i.e., it permits multiple conclusions), and dialetheic (i.e., permits "truth-value gluts" or some sentences or statements that are both true and false, and thereby entails a rejection of the universal law of non-contradiction).

> (2) Under the slogan *meaning is use* the nature of linguistic meaning becomes fully embedded in human action and human life.

> (3) Under the rubric of *language games* the scope of meaning is radically widened to include direct and indirect speech-acts (implicature), indexicality or context-dependency, emotive expression, metaphor, and more generally linguistic actions of all sorts.

> (4) Under the slogan *mind is human behavior* the Cartesian and Schopenhauerian

solipsistic version of the Kantian transcendental mind of the Tractatus becomes the living form of embodied human comportment.

And finally,

(5) Under the slogan "what has to be accepted, the given, is—so one could say, *forms of life*" (*PI* 226e), Tractarian essentialism and also the Tractarian *solipsistic* version of Kantian Transcendental Idealism are both sharply criticized, and then replaced by an *anthropocentric-communitarian* version of Kantian Transcendental Idealism,[82] namely a transcendental metaphysics of the commonsensical or the ordinary, relative to actual or possible rational human practices, in which essences and structures are all manifest ("nothing is hidden"), although almost always unseen by us because of philosophical confusion.

There is also a radical turn in Wittgenstein's conception of philosophy from *logical analysis* or the "logical clarification of thoughts" (*TLP* 4.112, 77) to *logical psychoanalysis* or "a battle against the bewitchment of our intelligence by means of language" (*PI* §109, 47ᵉ). The most obvious historical parallels here are with Freudian psychoanalysis and with Kant's critical meta-philosophy in the CPR, especially as it is formally developed in the Transcendental Dialectic. One crucial quasi-technical notion deployed by Wittgenstein in this connection is that of a "picture," that is, a simple philosophical model or analogy or stereotype or reification—often highly diagrammatic or imagistic—that we presuppose without argument, that narrowly constrains and limits our thinking, and that typically leads to antinomies, paradoxes and insoluble "problems" in philosophy: e.g., "the mind-body problem." These radical methodological features of the *Investigations* are perfectly reflected in its title (specifically *not* a treatise or systematic work), organization (a series of numbered remarks without any attempt at dividing them into topics or sections, without headings, etc.), argument-style (entirely non-linear and dialectical, with thought-experiments, epigrammatic pronouncements, constant use of metaphors, jokes, etc.), and prose style (highly conversational and

elliptical, notice the constant use of dashes, also of quotation-marks to indicate different voices and interlocutors, etc).

This raises a more fundamental point about the nature of philosophical explanations in the *Investigations*. In the *Tractatus*, the explanation of some fact or phenomenon typically took the form of a systematic decomposition to simple entities plus classical logic, against the necessary a priori presuppositional backdrop of a logic which is fully transcendental. The crucial features there are

(i) the idea of deeper and more basic levels of reality,

and

(ii) classical logic as a priori, universal, and essential for thought, language, and the world.

In the *Investigations*, by sharp contrast,

(i*) Explanations always appeal to factors *at the same level as what is being explained* .

and

(ii*) Logic is *neither classical nor essentially separate from the original phenomenon of meaningful natural language itself*, that is, that logic is essentially embedded in meaningful natural language, which is why Wittgenstein calls it "grammar."

Against that novel post-Tractarian philosophical backdrop then, the basic explanation for the fact that meaning is use is twofold: first, that language is essentially embedded in basic rational human linguistic practices called *language-games*, and second, that language-games in turn are essentially embedded in actual historical networks of rational human activity and human culture called *forms of life*. And that is where philosophical explanation stops.

One crucial consequence of this is that for Wittgenstein in the *Investigations*, meaningful language is ultimately a kind of rational human action, indeed the

characteristic kind of rational human action. To twist Goethe's famous line from *Faust* ("In the beginning was the Deed"), which in turn famously twists the famous line from the Bible ("In the beginning was the Word"), we can say that for the later Wittgenstein meaningful words just *are* deeds:

> Words are deeds. (*CV* 46ᵉ)

> In this way, I should like to say the words "Oh, *let* him come!" are charged with my desire. And words can be wrung from us, —like a cry. Words can be *hard* to say: such, for example, as are used to effect a renunciation, or to confess a weakness. (Words are also deeds.) (*PI* §546, 146ᵉ)

Humans are essentially linguistic agents and use of language is essentially the mastery of a skill (*PI* §20, 9ᵉ). In turn, this opens up the very idea of meaning to every conceivable role that language can play in rational human activity (see *PI* §23, 11e – 12e). It also opens up the possibility that some actions are essentially linguistic, e.g., giving commands, promising, or legal actions. In the two decades immediately following the publication of the *Investigations* J.L. Austin, John Searle, and others developed this idea into the scientific-seeming theory of *speech acts*, according to there is a finite, generative set of universal a priori rules that strictly govern our ability to use words, especially including our ability *to utter φ* (e.g, 'I do') *and thereby ψ* (e.g., marry someone), that is, our ability to make *performative utterances*.[83]

Unlike the *Tractatus* however, the *Investigations* is emphatically not a *treatise*, that is, a systematic scientific work written down as a linear philosophical text governed by the deductive canons of classical logic. Still, even in non-classical logic there are valid and sound *arguments*. So I think that we can reconstruct Wittgenstein's argument for the use thesis as a two-step line of non-classical reasoning:

(1) Display the inadequacies of the classical theories of meaning (dialectical criticism). and then

(2) Offer the thesis that meaning is use as the best overall explanation of the phenomenon of meaning (inference to the best explanation).

He carries this out by considering simpler languages and simpler language practices than our own, which in his terminology is to say that he considers "language games" in a second sense of that term, in particular the "Augustinian" language games in *PI* §§1-21.

V.2.3 The Thesis that Meaning is Use

In the *Tractatus* the only admissible sort of meaning ("saying") is the meaning of (atomic[84]) propositions, and there are four different competing and partially overlapping theories about the nature of the meaning belonging to (atomic) propositions:

(i) All words are names, and the meaning of a word is nothing but the object it names. Furthermore all names are proper names, and the meaning of every basic proper name in a basic proposition (whether a basic singular term or a basic general term—a.k.a., a "concept-word) is nothing but the referent or bearer of the name, i.e., an absolutely simple individual concrete object or a definite abstract concept or universal. (*Pure Referentialism.*)

(ii) The sense of every meaningful proposition is nothing but how it either isomorphically pictures a state of affairs (*Sachverhalt*) or is a truth-functional compound of such picturing sentences. (*The Picture Theory.*)

(iii) The meaning of any non-referring or non-picturing linguistic sign in a proposition—e.g., a logical constant—is nothing but a rule for manipulating or operating with that sign in a strict logical or mathematical calculus or in some other formal language-system. (*Rule-Based Semantics.*)

and

(iv) The ultimate meaning of any name, sentence, or other linguistic sign is nothing but a conscious mental representation (or "idea") in the mind of an individual speaker of a language. (*Semantic Solipsism.*)

By sharp contrast to all of these, in the *Investigations*, Wittgenstein wants to defend the thesis that *for a great many but not all cases in which we employ the word 'meaning', the meaning of any part of a language—any word—is its use in that language, and the meaning of a name in that language is sometimes explained by pointing to its bearer*:

> For a large class of cases—though not for all—in which we employ the word 'meaning' ("Bedeutung"), the meaning of a word is its use in the language (*Gebrauch in der Sprache*). And the meaning of a name is sometimes explained (*erklärt*) by pointing to its *bearer* (*PI* §43, 20ᵉ)

It should be noticed that this thesis says that the meaning of a word is almost always its use, *including* a few cases in which the meaning of a name is *explained* by someone's pointing to an object that bears the name. As regards that second clause, nowadays we would say that just as the semantic *content* of a demonstrative such as 'this' or 'that' requires a rational human *demonstration-act* of pointing as a vehicle of reference in order to fix the application of its variable *character* for a given *speech-context*,[85] so too a name can have its meaning fixed by a rational human ostensive act, and thereby explained. Later Wittgenstein puts the same idea about the meaning of names and human demonstration-acts this way:

> The demonstrative 'this' can never be without a bearer. It might be said: "so long as there is a *this*, the word 'this' has a meaning too, whether *this* is simple or complex." —But that does not make the word into a name. On the contrary: for a name is not used with, but rather only explained by means of, the gesture of pointing. (*PI* §45, 21ᵉ)

But in any case, and for convenience, I will henceforth always refer to the two-clause carefully qualified thesis in §43 as "the meaning-is-use thesis" and correspondingly to the concept of "meaning-as-use."

The concept of meaning-as-use covers two distinct but intimately related sub-notions:

> (i) *semantic function*, according to which the meaning of a word is its specific role (i.e., a specific normative pattern of operations) in a living human language,

and

> (ii) *semantic application*, according to which the meaning of a word is determined by how it is applied by individual human speakers, in actual or possible communities of speakers, in actual or possible speech-contexts.

The semantic *function* of words is the fact that any meaningful part of a language is essentially a "tool" which can be used correctly or incorrectly in the context of a larger totality of linguistic equipment or technology, and the semantic *application* of words is the fact that the meaning of a word depends on its implementation in actual or possible rational human speech-contexts and speech-communities. Of course semantic function and application can come apart when language is either misused (which produces nonsense) or applied in new contexts (which produces new sense). But normally they are smoothly complementary:

> Think of the tools in a tool-box: there is a hammer, pliers, a saw, a screw-driver, a rule (*Maßtab*), a glue-pot, glue, nails, and screws. –The functions (*Funktionen*) of words are as diverse as the functions of these objects. (And in both cases there are similarities.) Of course what confuses us is the unform appearance of words when we hear them spoken or meet them in script and print. For their *application* (*Verwendung*) is not presented to us so clearly. Especially not, when we are doing philosophy! (*PI* §11, 6ᵉ)

> It is interesting to compare the multiplicity of the tools in language and of the ways they are used, the multiplicity of kinds of word and sentence, with what logicians have said about the structure of language. (Including the author of the *Tractatus Logico-Philosophicus*.) (*PI* §24, 12ᵉ)

Here, and more generally, it is crucial to note later Wittgenstein's appeal to semantic function and application (words as tools for doing things, or the action-embeddedness of language), *norms* (governing ideals or standards of language use), *context-dependency* (indexicality), and *actual or possible communities of rational human speakers* (transcendental-anthropocentric communitarianism).

So, combining the notions of semantic function and application, norms, context-dependency, and actual or possible communities of rational human speakers into one complex concept, later Wittgstenstein's overarching thesis in the *Investigations* to the effect that meaning-is-use, is the same as the thesis that the meaning of a word is its specific role in a living and complete human language *together with* how it is applied by individual rational human speakers in actual or possible communities of rational human

speakers in actual or possible contexts. Or, to sloganize: *meaning is the career of words in rational human action.*

It is very important to emphasize *all* of these factors and not merely the application factor in the meaning-is-use thesis, because the application factor alone gives the false impression of empiricism and relativism about meaning. For my present purposes, empiricism about meaning is the thesis that semantic content (or sense) and reference are both strictly determined by sensory experiences and other contingent empirical facts. And relativism about meaning is the thesis that semantic content (or sense) and reference are both strictly determined by either actual human individuals (solipsistic relativism) or actual human communities (communitarian relativism). But the meaning-is-use thesis in the *Investigations*, although it certainly has some empiricist and relativist features, is neither a form of semantic empiricism nor a form of semantic relativism, nor indeed some mere combination of the two. What resists such explanatory reductions of the concept of meaning-as-use are the *action-embeddedness, rational normativity, indexicality,* and *transcendental-anthropocentric communitarianism* of meaning.

Here is a proposal about how best to understand Wittgenstein's theory that meaning-is-use, especially including his direct appeal to the five non-empiricist and non-relativist facts about meaning. The problem of philosophical interpretation is made substantially more difficult when we remember that the logic of later Wittgenstein's arguments in the *Investigations* is dialectical. But even allowing for that, I think we can charitably and rationally reconstruct his dialectical arguments in the *Investigations* for the meaning-is-use thesis just by asking ourselves the following question: What are the

implications and critical limitations of the four competing theories of meaning relevant to the *Tractatus*—Pure Referentialism, the Picture Theory, the Rule-Based Semantics, and Semantic Solipsism? The thesis that meaning-is-use will then be established in and through the critical deconstruction of the classical theories, together with the claim that meaning-is-use is the best overall explanation of all the relevant linguistic facts.

Now the primitive language games are supposed to make the critical deconstruction of the classical theories of meaning directly evident to us in philosophical *dioramas*, or "living pictures," at least as far as Pure Referentialism and the Picture Theory are concerned. Further on in the *Investigations* later Wittgenstein also offers separate deconstructive arguments against Rule-Based Semantics by developing the Rule Following Paradox, and against Semantic Solipsism by developing the Private Language Argument.[86] Finally, the failures of all the competing theories of meaning indirectly and cumulatively establish the several basic elements of the thesis that meaning-is-use. Then on Wittgenstein's behalf we can conclude that the thesis that meaning-is-use is the best overall explanation of all the relevant meaning-facts or meaning-phenomena.

V.2.4 A Map of the *Investigations*

In the Preface to the *Investigations*, later Wittgenstein beautifully describes the non-linear, non-Tractarian, non-classical logical structure of his second book.

> The thoughts that I publish in what follows are the precipitate of philosophical investigations which have occupied me for the last sixteen years. They concern many subjects: the concepts of meaning, of understanding, of a proposition, of logic, the foundations of mathematics, states of consciousness, and other things…. It was my intention at first to bring all this together in a book whose structure I pictured differently at different times. But the essential thing was that the thoughts should proceed from one subject to another in a natural order and without breaks. After several unsuccessful attempts to weld my results together into such a whole, I realized that I should never succeed. The best that I could write would never be more than philosophical remarks; my thoughts were soon crippled if I tried to force them on in any single direction against their natural inclination.—And this was, of course, connected with the very nature of the investigation. For this compels us to travel over a wide field of thought criss-cross in every direction.—The philosophical remarks in this book are, as it were, a numebr of sketches of landscapes which were made in the course of these long and involved journeyings. The same or

almost the same points were always being approached afresh from different directions, and new sketches made. Very many of these were badly drawn or uncharacteristic, marked by all the defects of a weak draughtsman. And when they were rejected a number of tolerable ones were left, which now had to be arranged and sometimes cut down, so that if you looked at them you could get a picture of the landscape. Thus this book is really only an album. (*PI*, ix^e).

The logical structure of the *Investigations* is analogous to the structure of a landscape: it cannot be digitally computed and recursively generated, like a decidable theorem in classical truth-functional logic or the monadic fragment of first-order classical predicate logic. But at the same time it is not in any way *amorphous* or *chaotic*. On the contrary, it is replete with rich logical structure of a non-computable and indeed even unprovable kind. Its non-classical logical structure can still be *mapped*.

In light of that fact, and more explicitly now, I want to say that the basic argument-structure of the *Investigations* has seven partially overlapping, yet nevertheless conceptually distinguishable and therefore non-trivially distinct, non-classical-logical successive parts, or "regions," rather like the differently colored and irregularly-shaped continental areas in a Mercator projection map of the (roughly) spherical surface of the Earth. Here are the seven regions, in sequence.

(Region 1: Meaning-is-Use) The main thesis of the book is that *linguistic meaning is use*, where the concept of *use* is the conjunction of the sub-concepts of

(i) *word-function*, or the normatively rule-governed role of words in the whole language, and

(ii) *word-application*, or the actual deployment of words by the linguistic acts of individual users, in communities, in context.

(Region 2: Language-Games and Forms of Life) The fact of linguistic use is then held to be explained by two more primitive facts:

(i) *language-games*, or basic human linguistic practices,

and

> (ii) *forms of life*, or actual living human beings in their actual human communities and their historically-embedded social practices, considered as unified normatively rule-governed bearers of meaning and purpose, hence as *transcendental-anthropocentric communitarian a priori conditions for the possibility of rational human experience in the manifest natural world.*

(Region 3: The Critiques of Referentialism, The Picture Theory, Rule-Based Semantics, and Solipsistic Semantics) The use theory is then indirectly demonstrated by criticizing and then rejecting four inadequate semantic theories:

> (i) *Referentialism or 'Fido'-Fido Semantics*. Referentialism says that the meaning of a word is nothing but its reference.Later Wittgenstein's rejection of Referentialism primarily appeals to critical arguments based on negative existential propositions and family resemblance concepts.

> (ii) *The Picture Theory*. The Picture Theory says that the meaning of a sentence is nothing but how it isomorphically model an atomic fact or else truth-functional compoundings of such sentences. Later Wittgenstein's rejection of the Picture Theory primarily appeals to an argument against absolute simples from the impossibility of unique decompositions of macrophysical objects.

> (iii) *Rule-Based Semantics.* Rule-Based Semantics says that the meaning of any linguistic sign is nothing but a rule for manipulating or operating with that sign in a logical or mathematical calculus, or other non-formalized language-system. Later Wittgenstein's rejection of Rule-Based Semantics primarily appeals to *The Rule Following Paradox.*

> (iv) *Solipsistic Semantics.* Solipsistic Semantics says that the meaning of a name, sentence, or other linguistic sign is nothing but a conscious mental representation or "idea" in the mind of an individual speaker of a language. Later Wittgenstein's rejection of Solipsistic Semantics primarily appeals to *The Private Language Argument*.

(Region 4: Five Positive Theses about Meaning) The rejections of the four inadequate semantic theories then lead correspondingly to five positive Wittgensteinian theses about linguistic meaning:

> (i) The meaning of a singular term is a partial function—or a specific contingently-determined set of mappings or "routes"—from language-games employing singular terms and forms of life onto individual objects, and each of these "routes" is literally part of the meaning itself.

> (ii) Concepts, the meanings of predicate expressions, are family-resemblance networks.

(iii) Propositions are pictures of facts *only* internally to propositional language games and under a relativized ontology of object-samples.

(iv) Rule-following is externally normatively justified by communal rule-following practices to which the rule-follower belongs non-cognitively by an "agreement" or *Übereinstimmung* with other participants in that language-game, which in turn supervenes on the deeper fact that human speakers are necessarily practically and vitally embedded in some or another form of life.

(v) Semantic anti-individualism and semantic externalism both hold for sensation-language.

(Region 5: Four Positive Theses about the Mind) The two positive theses

under (4) (v) then lead to four positive theses about the rational human mind:

(i) the token privacy of sensations,

(ii) human capacity behaviorism,

(iii) sensation personalism,

and

(iv) an activist phenomenology of mental states and processes.

(Region 6: Four Positive Theses about the Linguistic Phenomenology of

Seeing) These four theses, in turn, conjointly lead to the linguistic phenomenology of

seeing (or visual experience), which also has four positive theses:

(i) There is a basic distinction between *direct seeing* (seeing-this) and *interpretive seeing* (seeing-as).

(ii) Interpretive seeing requires direct seeing.

(iii) Interpretive seeing requires conceptual abilities.

(iv) The phenomenon of aspect-blindness entails that direct seeing can occur without any sort of interpretive seeing, hence direct seeing is *non-conceptual*.

(Region 7: The Linguistic Phenomenology of Experiencing the Meaning of a

Word) Finally, these four theses are then extended to the linguistic phenomenology of

experiencing the meaning of a word, which completes the whole account by returning

full-circle—as it were, all the way around the (roughly) spherical surface of the Earth—to

the meaning-is-use thesis, thereby also demonstrating some further positive theses about the concept of use.

V.2.5 The Critique of Pure Reference: What the Builders Did

Referentialism holds that all words are names, and that the meaning of a name is nothing but the referent or bearer of that name. Referentialism, as its name obviously implies, identifies linguistic meaning with reference. Thus according to Referentialism 'Fido' means Fido and 'dog' means the concept DOG. Furthermore, according to *Pure Referentialism*, all names are proper names, and the meaning of every basic proper name in a basic proposition (whether a basic singular term or a basic general term—a.k.a., a "concept-word") is nothing but the referent or bearer of the name, i.e., an absolutely simple individual concrete object or a definite abstract concept or universal. In turn, according to later Wittgenstein in the *Investigations*, there are two main problems with Pure Referentialism. First, identifying meaning with reference to individual objects in the case of singular terms does not account for systematic variations in the use-based meanings of ostensive terms having the same referent (*PI* §§28-38). Second, identifying meaning with reference in the case of general terms fails because there are no uniquely identifiable concepts or universals (*PI* §§66-71, and 75-78).

What follows now is a three-step, A-B-C style rational reconstruction of later Wittgenstein's opening argument against Pure Referentialism in the first twenty or so sections of the *Investigations*, as a paradigmatic case study in how he argues for the thesis that meaning-is-use. In turn, the meaning-is-use thesis is to be understood, as I have said, as the thesis that linguistic meaning is the career of words in human action, with its two distinct sub-notions of semantic function and semantic application, taken together with

the five non-empiricistic and non-relativistic facts about meaning. For each of the steps in

the reconstruction I will provide a detailed critical commentary on that step's rationale

and implications.

> (Step A) Referentialism holds that all words are names, and that the meaning of any word is nothing but the object it names. Furthermore, according to *Pure Referentialism*, all names are proper names, and every basic proper name in a basic proposition (whether a basic singular term or a basic general term) is nothing but the referent or bearer of the name, i.e., an absolutely simple individual concrete object or a definite abstract concept or universal.

Commentary on Step A.

Referentialism as a philosophical thesis goes at least as far back as Plato's *Theaetetus*.

Indeed, later Wittgenstein explicitly quotes the *Theaetetus* in §46, in support of the

particular version of Referentialism he is focusing on:

> What lies behind the idea that names really signify simples? Socrates says in the Theatetus: "If I make no mistake, I have heard some people say this: there is no definition of the primary elements—so to speak—out of which we we and everything else are composed; for everything that exists in its own right can only be *named*, no other determination is possible, neither that it *is* nor that it *is not*.… But what exists in its own right has to be … named without any other determination. In consequence it is impossible to give an account of any primary element; for it, nothing is possible but the bare name; its name is all it has. But just as what consists of these primary elements is itself complex, so the names of the elements become descriptive language by being compounded together. For the essence of speech is the composition of names." Both Russell's 'individuals' and my 'objects' (*Tractatus Logico-Philosophicus*) were such primary elements. (*TLP* §46, 21ᵉ)

This particular version of Referentialism thus includes two sub-theses to the effect that

> (i) The basic referring terms in basic propositions are all *proper names* (as opposed to, say, definite descriptions), including both *basic singular terms* in grammatical or logical subject position and also *basic general terms* (a.k.a. "concept-words") in grammatical or logical predicate position.

and

> (ii) The objects for which these basic singular terms and basic general terms stand are *absolutely simple concrete individuals* and *definite abstract concepts or universals*.

But not all Referentialists are as semantically puristic as Plato's Socrates was in

the *Theaetetus*. So this raises the critical question of whether *every* possible version of

Referentialism need be committed to the thesis that the basic referring terms in basic

proposition are all *proper* names. Indeed, it seems clear that a less puristic Referentialist might instead take the basic referring terms in basic propositions to be *demonstratives* or some other kind of essential indexical.[87] It also raises the critical question of whether every possible version of Referentialism need be committed to the puristic thesis that the objects for the basic singular terms stand are absolutely simple concrete individuals and that the objects for which the basic general terms stand are definite abstract universals. And indeed, it seems that a non-puristic or impuristic Referentialist might instead hold that the objects picked out by basic singular terms are only *relatively* simple concrete individuals, that is, simple relative to some particular way of humanly conceptualizing a decomposition of a complex perceivable object, and also that the objects for which basic general terms stand are just *real-world manifest properties,* corresponding to *rational human concepts* in all their varying degrees of vagueness and variety.[88]

I have been calling the special version of Referentialism that is committed to the puristic thesis that the basic referring terms are *proper* names (including both basic singular terms and basic general terms), and also to the further two-part purist thesis that the simple concrete individual objects for which basic singular terms stand are *absolute* simples and that the objects for which basic general terms stand are *definite abstract* concepts or universals, *Pure Referentialism.* This is a crucial interpretive move. It is Pure Referentialism, and not Referentialism as such, that is the philosophical target of later Wittgenstein's deconstructive critique of Referentialism in the *Investigations.*

Moreover, the bull's-eye of the philosophical target of Pure Referentialism is not in fact the semantics of names proposed by Plato's Socrates in the *Theaetetus,* but instead a dual bull's-eye consisting of *Russell's* semantics of names circa 1912 and *early*

Wittgenstein's semantics of names in the *Tractatus*. Russell's semantics of names circa 1912, in turn, depends heavily on two doctrines:

> (1) the *multiple relation* theory of judgment, which says that a proposition is nothing but an ordered set of absolute simples, definite abstract concepts or universals, and abstract logical constants organized by the mind of a subject who stands in multiple acquaintance relations to these objects in the act of judging,

and

> (2) the distinction between *knowledge by acquaintance* and *knowledge by description*.

Here is the gravamen of what Russell says about these doctrines.

> When we judge that Charles I died on the scaffold, we have before us not one object but several objects, namely, Charles I, dying, and the scaffold. Similarly, when we judge that Charles I died in his bed, we have before us the the objects Charles I, dying, and his bed. These objects are not fictions: they are just as good as the objects of the true judgment. We therefore escape the necessity of admitting objective falsehoods, or of admitting that in judging falsely we have nothing before the mind. Thus in this view judgment is a relation of the mind to several other terms: when these other terms have *inter se* a 'corresponding' relation, the judgment is true; when not, it is false.[89]

> Knowledge of things, when it is of the kind we call knowledge by *acquaintance*, is essentially simpler than knowledge of truths, and logically independent of knowledge of truths, though it would be rash to assume that human beings ever, in fact, have acquaintance with things without knowing some truth about them. Knowledge of things by *description*, on the contrary, always involves … some knowledge of truths as its source or ground…. We shall say that we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths.[90]

> All our knowledge, both knowledge of things and knowledge of truths, rests upon acquaintance as its foundation. It is therefore important to consider what kinds of things there are with which we have acquaintance…. We have acquaintance in sensation with the data of the outer senses, and in introspection with the data of what may be called the inner sense—thoughts, feelings, desires, etc.; we have acquaintance in memory with things what have been data either of the outer senses or inner sense… In addition to our acquaintance with particular existing things, we also have acquaintance with what we shall call *universals*, that is to say, general ideas, such as *whiteness*, *diversity*, *brotherhood*, and so on. Every complete sentence must contain at least one word which stands for a universal, since all verbs have a meaning which is universal… [We must] guard against the supposition that whatever we can be acquainted with must be something particular and existent. Awareness of universals is called *conceiving*, and a universal of which we are aware is called a *concept*.[91]

In the *Tractatus*, early Wittgenstein explicitly rejects Russell's multiple relation theory of judgment because it is psychologistic (*TLP* 4.1121, 77ᵉ) and also because he denies that logical constants stand for any sort of object—"my fundamental thought is that the 'logical constants' do not stand for something (*nicht vertreten*)" (*TLP* 4.0312,

69^e). But in the *Tractatus* he also explicitly *accepts* Russell's distinction between

knowledge by acquaintance and knowledge by description (in German, the distinction

between *Kennen* and *Beschreibung*). As a consequence of explicitly accepting Russell's

acquaintance-description distinction, in the *Tractatus* early Wittgenstein also explicitly

accepts Pure Referentialism. Here are the relevant texts.

> If I know (*kenne*) an object, then I also know (*kennen*) all the possibilities of its occurrence in atomic facts. (*TLP* 2.0123, 33^e)

> In order to know (*kennen*) an object, I must know (*kennen*) not its external but all its internal qualities. (*TLP* 2.01231, 33^e)

> States of affairs can be described (*beschreiben*), but not *named*. (Names resemble points; propositions resemble arrows, they have sense.) (*TLP* 3.144, 47^e)

> The simple signs employed in propositions are called names. (*TLP* 3.202, 47^e)

> The name means (*bedeutet*) the object, The object is its meaning (*Bedeutung*). (*TLP* 3.203, 47^e)

> In the proposition the name stands for (*vertritt*) the object. (*TLP* 3.22, 49^e)

> Objects I can only *name.* Signs stand for (*vertreten*) them. I can only speak of them. A proposition can only say *how* a thing is, not *what* it is. (*TLP* 3.2221, 49^e)

> Reality must be completely described (*beschreiben*) by the proposition. A proposition is a description (*Beschreibung*) of a fact. (*TLP* 4.023, 67^e)

Now in the *Investigations*, later Wittgenstein wants to establish the meaning-is-

use thesis. He therefore rejects the distinction that "the author of the *Tractatus Logico-*

Philosophicus" had accepted between acquaintance and description. Nevertheless the

rational human *act* of acquaintance, in the form of a demonstration-act of ostensive

pointing, plays an important yet subsidiary role in the concept of meaning-as-use. For as

we have seen, later Wittgenstein argues that although *generally* the meaning of a word is

its use, *sometimes* the meaning of a name is explained by pointing to its bearer. This two-

part claim is what he wants to prove in two steps, first by means of his deconstructive

critique of the philosophical living picture of Pure Referentialism that is etched into the

Augustinian theory of language, and then second by appealing to the meaning-is-use

thesis as the best overall explanation of all the relevant linguistic facts.

> (Step B) The Augustinian language game of the Builders in *PI* §§2, 6, and 8 is a living picture or diorama of a Pure Referentialist language.

Commentary on Step B.

The *Investigations* begins with a text from Augustine's *Confessions*, I, 8:

> When they (my elders) named some object , and some accordingly moced towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shewn by their bodily movements, as it were the natural language of all peoples: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice which expresseses our state of mind in seeking, having, rejecting, or avoding something. Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified; and after I had trained my mouth to form these signs, I used them to express my own desires. (*PI* n.1, 2ᵉ)

In order to understand later Wittgenstein's use of this text from the *Confessions*, we must

recall that the notion of a "language game" in this context means "the idea of a language

more primitive than ours" (*PI* §2, 3ᵉ). As the self-appointed Grand Inquisitor of his own

earlier Pure Referentialist conception of meaning in the *Tractatus*, later Wittgenstein is

deeply interested in *the Augustinian language game* in particular, precisely because "that

[Pure Referentialist] philosophical concept of meaning has its place in a primitive idea of

the way language functions" (*PI* §2, 3ᵉ). Here is what later Wittgenstein specifically says

about the Augustinian language game:

> Let us imagine a language for which the description given by Augustine is right. The language is meant to serve for communication between a builder A and an assistant B. A is building with building stones : there are blocks, pillars, slabs, and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words "block," "pillar," "slab," and "beam." A calls them out; —B brings the stone which he has learnt to bring at such-and-such a call. —Conceive this as a complete language game. (*PI* §2, 3ᵉ)

> We could even imagine that the language of §2 was the *whole* language of A and B; even the whole language of a tribe. The children are brought up to perform *these* actions, to use *these* words as they do, and to react in *this* way to the words of others. (*PI* §6, 4ᵉ)

> We can also think of the whole process of using words in (2) as one of those games by means of which children learn their native language. I will call these games "language-games" and will sometimes speak of a primitive language as a language-game. And the processes of naming the

stones and of repeating words after someone might also be called language-games. Think of much of the use of words in games like ring-a-ring-a-roses. I shall also call the whole, consisting of language and the actions (*Tätigkeiten*) into which it is woven, the "language-game." (*PI* §7, 5ᵉ)

Let us now look at an expansion of language (2). Besides the four words "block," "pillar," etc., let it contain a series of words used as the shopkeeper in (1) used the numerals [to stand for finite cardinal numbers and counting out groups of objects] (it can be the series of letters of the alphabet); further, let there be two words, which may as well as be "there" and "this" (because this roughly indicates their purpose), that are used in connexion with a pointing gesture; and finally a number of colour samples. A gives an order like: "d—slab—there". At the same time he shews the assistant a colour sample, and when he says "there" he points to a place on the building site. From the stock of slabs B takes one for each letter of the alphabet up to "d", of the same colour as the sample, and brings them to to the place indicated by A. —On other occasions A gives the order "this—there". At "this" he points to a building stone. And so on. (*PI* §8, 5ᵉ)

Later Wittgenstein never does tell us what the Builders are building. But it is not too fanciful, and indeed it even makes very good instructive philosophical sense, to imagine that the Builders described in §§2, 6, and 8 are trying to build either the Tower of Babel, as described in Genesis 11: 1-9—

And the LORD came down to see the city and the tower, which the children of men builded. And the LORD said, Behold, the people is one, and they have all one language; and this they begin to do: and now nothing will be restrained from them, which they have imagined to do. Go to, let us go down, and there confound their language, that they may not understand one another's speech. So the LORD scattered them abroad from thence upon the face of all the earth: and they left off to build the city. Therefore is the name of it called Babel; because the LORD did there confound the language of all the earth: and from thence did the LORD scatter them abroad upon the face of all the earth.

—or perhaps the wall upon which Humpty Dumpty sat in *Through the Looking Glass*:

Don't stand chattering to yourself like that," Humpty Dumpty said, looking at her for the first time, "but tell me your name and your business."
"My name is Alice, but—"
"It's a stupid name enough!" Humpty Dumpty interrupted impatiently. "What does it mean?"
"*Must* a name mean something?" Alice asked doubtfully.
"Of course it must," Humpty Dumpty said with a short laugh: "*my* name means the shape I am— and a good handsome shape it is, too. With a name like yours, you might be any shape, almost."[92]

It makes sense that the Builders are trying to build the Tower of Babel. This is because we can think of the Logical Atomists, including both the author of the *Tractatus* and Russell circa 1912, as attempting to build a logico-semantic tower, called *the Ideal Language*, all the way up to Platonic heaven. But this project led inevitably to a logico-semantic Fall into the irreducible and sometimes almost incommensurable plurality of

different natural languages and language games scattered abroad upon the face of all the earth. The ultimately abandoned, half-built tower of Logical Atomism—and here we can think of that amazing painting by Peter Breughel the Elder—is then rightly called *Babel*. It also makes sense that the Builders are trying to build Humpty Dumpty's wall. This is because, at least as Lewis Carroll presents that bumptious egghead H.D., he is quite explicitly a Pure Referentialist who holds that his own proper name uniquely means his own shape, and that he can use his own name to point directly to his shape.

Whatever the possible subterranean philosophical influences of the Book of Genesis and *Through the Looking Glass* on the *Investigations*, however, later Wittgenstein's own explicit two-part gloss on the text from Augustine's *Confessions* says this:

> These words, it seems to me, give us a particular picture of the essence of human language. It is this: the individual words in language name objects—sentences are combinations of such names. —In this picture of language we find the roots of the following idea: Every word has a meaning (*Bedeutung*). This meaning is correlated with the word. It is the object for which the word stands. Augustine does not speak of there being any difference between kinds of word. If you describe the learning of language in this way you are, I believe, thinking primarily of nouns like "table", "chair", "bread", and of people's names, and only secondarily of the names of certain actions and properties; and of the remaining kinds of word as something that will take care of itself. (*PI* §1, 2ᵉ).

So the Augustinian language game is a diorama of Pure Referentialism. But why did later Wittgenstein use the passage from *Confessions* as his starting text, and not other very similar passages from the *Theaetetus* or, indeed, from *Through the Looking Glass*? One obvious answer is that in addittition to being a Pure Referentialist *avant la lettre*, Augustine also strongly anticipates Cartesian epistemology and metaphysics of mind, not to mention Edmund Husserl's semantic phenomenology in the *Logical Investigations* I, chapter 1, section 10, provocatively entitled "Expressions in Solitary Life," as well as Rudolf Carnap's equally provocative notion of "methodological solipsism" in the *Logical*

Structure of the World. Augustine's *Confessions*, in short, strongly anticipates *semantic solipsism*. Augustine tellingly says this about his confessions:

> Why then does it matter to me whether men should hear what I have to confess, as though it were they who were to cure all the evil that is in me? They are an inquisitive race, always anxious to pry into other men's lives, but never ready to correct their own. Why do they wish to hear from me what sort of man I am, though they will not listen to you when you tell them what they are? When they hear me speak about myself, how do they know I am telling the truth, since no one *knows a man's thoughts, except the man's own spirit that is within him*?[93]

Augustine's *Cartesian* semantic solipsism, in turn, captures the core of early Wittgenstein's own *transcendental* semantic solipsism in the *Tractatus*. And in the early stages of his critical discussion of the Augustinian theory of language in the *Investigations*, later Wittgenstein directly addresses Cartesian and transcendental semantic solipsism alike:

> If you do not keep the multiplicity of language-games in view, you will perhaps be inclined to ask questions like: "What is a question?" –Is it the statement that I do not know such-and-such, or the statement that I wish the other person would tell me ….? Or is it the description of my mental state of uncertainty? –And is the cry "Help!" such a description? … The significance of such possibilities of transformation, for example of turning all sentences into sentences beginning "I think" or "I believe" (and thus, as it were, into descriptions of *my* inner life) will become clearer in another place. (Solipsism.) (*PI* §24, 12ᵉ)

Of course, what he means here by things becoming clearer in "another place" is The Private Language Argument.[94]

> (Step C) It is manifest that not everything that is language has meaning in this way (*PI* §3), **if only because the referring terms of the Builders' language also function as** orders (*PI* §18). In fact it is more correct to think of words as tools embedded in language-games and in forms of life, and as playing any number of roles relative to different games and forms of life, than to think of them as playing a single de-contextualized semantic role in the language, such as naming objects (*PI* §§19-23, 26-27).

Commentary on Step C.

This last step in the opening three-step argument is the philosophically seminal one, insofar as it presents later Wittgenstein's thesis that meaning-is-use. I have already

discussed the nature and philosophical implications of this thesis in sub-section V.2.3, so will not repeat that here.

As far as *establishing* the meaning-is-use thesis, however, the crucial move here is to get us to see how Pure Referentialism turns out to be either a completely tautologous and trivial thesis, or else a significant thesis that is clearly false. As to the former, later Wittgenstein says:

> When we say: "Every word in language signifies (*bezeichnet*) something" we have so far said *nothing* whatever; unless we have explained exactly *what* distinction we want to make. (It might be, of course, that we wanted to distinguish the words of language (8) from words 'without meaning' such as occur in Lewis Carroll's poems, or words like "Lilliburlero" in songs. (*PI* §13, 7ᵉ)

In other words, there are as many different uses of meaningful language as there are different kinds of human intentional action, and the only *real* point of asserting a thesis like Pure Referentialism would be to distinguish meaningful language in general from nonsense. This is not to say, however, that nonsense cannot have its *own* uses or career in human action—as, e.g., in Carroll's *The Walrus and the Carpenter*[95]—but instead just that this specifically nonsensical kind of linguistic career is distinct from that of meaningful words. The human career of nonsensical language is distinguished fundamentally from the human career of natural or ordinary language by the manifest *playfulness* of nonsense as opposed to the manifest *everday seriousness* of natural or ordinary language, despite the deep fact that both are language-*games* embedded in *forms of life*, or normatively rule-governed episodes in the total career of words in human action. So ultimately the difference in language use lies in different human act-intentions, as also of course do all real differences in meanings. If Pure Referentialism is taken in this way, then it simply re-states the meaning-is-use thesis.

But if, on the other hand, Pure Referentialism is taken to stand for the substantive three-part thesis that all words are proper names, that the meaning of word is nothing but the object it names, and that every basic proper name in a basic proposition (whether a basic singular term or a basic general term or concept-word) denotes either an absolutely simple individual concrete object or a definite abstract concept or universal, then it is simply false, by the following argument, which leads us deeper into the *Investigations*, and well beyond the simple constructions of the Builders, in this step-by-step way:

(1) In a Pure Referentialist semantics, there are two distinct types of basic proper names: basic singular terms and basic general terms or concept-words. Absolutely simple individual concrete objects are assigned to basic singular terms, and definite abstract concepts or universals are assigned to basic general terms.

(2) Absolutely simple individual concrete objects are assigned to basic singular terms by ostension (*PI* §6). Singular reference is then best understood as ostensively attaching a name-label to an absolutely simple individual thing, i.e., by dubbing it (*PI* §37).

(3) But every ostension is open to many distinct possible interpretations (*PI* §§28-38), and only actual use will uniquely fix an interpretation.

(4) Morever, if the meaning of a basic singular term were just the bearer of the name, then whenever the bearer was destroyed, the meaning would be destroyed, which is absurd because it would make true negative existentials with singular terms into nonsense (*PI* §40). But true negative existentials with singular terms, such as "Moses did not exist," are in fact perfectly meaningful, although such sentences do also allow of irreducibly different meanings, depending on their use (*PI* §79).

(5) Furthermore, there are no such things as absolutely simple individual concrete objects, because every object we can perceive is complex in various ways, and allows of no unique decomposition into ultimate simple parts (*PI* §§ 46-64).

(6) So Pure Referentialism about basic singular terms is false, and the thesis that meaning-is-use is the best overall explanation of how even basic singular terms have meaning.

(7) Pure Referentialism as applied to basic general terms requires that every concept-word stand for a definite abstract concept or universal.

(8) But not all concept-words, as actually used, mean definite abstract concepts or universals: on the contrary, at least some of them mean only family resemblances or clusters of partially overlapping human concepts, at least some of which have blurred or vague boundaries—see, e.g., the concept GAME (*PI* §§66-71). Only the actual use of the general term will adequately disambiguate its meaning as a concept-word. Indeed there

are no analytic definitions of general terms, only our actual patterns of application of them (*PI* §§75-78).

(9) So Pure Referentialism about general terms is false, and the thesis that meaning-is-use is the best overall explanation of how even basic general terms have meaning.

(10) So Pure Referentialism more generally is false, and the thesis that meaning-is-use is the best overall explanation of how words have meaning.

It then follows from Step A, Step B, and Step C that the meaning-is-use thesis is true, including the important qualification that sometimes the human act of ostending an object that bears a name also explains the meaning of that name. In this way, the Augustinian theory of language leads directly from *Referentialism* to *rational human action*.

V.2.6 The Picture Theory and the Vices of Simplicity

The Picture Theory identifies the meaning of a sentence with how sentences are isomorphic models of atomic facts or else truth-functional compoundings of these. According to later Wittgenstein, as in the case of Pure Referentialism, there are also two basic problems with The Picture Theory. First, The Picture Theory is committed to the existence of absolutely simple objects, but there is no sufficient reason to think that there are anything but only relatively simple objects relativized to language games. This in turn raises the important issue of the role of "samples" in language games. Second, The Picture Theory is committed to an isomorphism between propositional structures and the structure of facts, but there is no way to establish the existence of such an isomorphism short of either some sort of mysterious externally pre-established harmony or else transcendental idealism. This raises the equally important issue of semantic realism vs. anti-realism. Here, now, is a reconstruction of later Wittgenstein's argument against The Picture Theory.

V.2.6.1 Wittgenstein's Argument against The Picture Theory Reconstructed

(1) The Picture Theory says that the meaning of a sentence is nothing but how sentences isomorphically model atomic facts or else truth-functional compoundings of such sentences.

(2) Atomic facts are composed of configurations of absolutely simple objects in isomorphic correspondence with the parts of the atomi cproposition, which is a configuration of "real names" (Frege's *Eigennamen*).

(3) So The Picture Theory presupposes that "real names" in atomic propositions stand for absolutely simple objects (*PI* §39). In this respect, Russell's early semantics, the *Tractatus*, and Plato's *Theaetetus* have all captured the same basic idea (*PI* §46).

(4) But what is an absolutely simple object? The problem is that macroscopic objects apparently have no unique decomposition into simple parts (*PI* §47). And if we try to imagine a primitive language game that models the Tractarian Picture Theory, we find the same lack of unique decomposition into simple parts (*PI* §48).

(5) So there are no absolutely simple objects, and The Picture Theory is therefore false.

(6) But the language game of using factual propositions implies the constant semantic availability of simple objects of some sort, even across the difference between existence and non-existence (*PI* §§50, 55).

(7) Contrary to The Picture Theory, then, it seems to be a much better overall explanation of the semantic of factual propositions to say that the "simple" objects are in fact systems of paradigms or samples—hence only *relatively* simple objects—that belong strictly to the "instruments" or technology of the particular language-game (say, of factual propositions about colours) that is in play (*PI* §§50-51).

(8) Therefore even though The Picture Theory is false, relativizing simple objects to language-games gives a better overall explanation of the semantics of factual propositions, and thus use is the best overall explanation of how sentences have meaning.

(9) This however implies the relativization of the ontology of atomic facts to language-games (*PI* §§59-60), which also undermines the semantic realism of The Picture Theory.

V.2.7 Understanding and Rule-Following

As a lead-up to The Rule Following Paradox and its solution, later Wittgenstein wants to establish two theses that he will be able to use in that argument. The first thesis (Thesis 1) follows from considerations concerning propositions and understanding; and the second thesis (Thesis 2) follows from considerations concerning understanding and

reading. In the text of the *Investigations* itself, these arguments run seamlessly together.

But for our analytical-critical purposes, it is useful to reconstruct them in two distinct

chunks, thereby constituting two successive prolegomena to The Rule Following

Paradox.

V.2.7.1 First Prolegomenon to The Rule Following Paradox: Propositions and Understanding

(1) In the *Tractatus*, it was assumed that concept of a proposition expressed
the *essence* of the proposition: necessarily and sufficiently, all propositions describe facts
("this is how things are")—and every proposition is bipolar ("a proposition is whatever
can be true or false") (*PI* §§134, 136). So necessarily, a part of language is a proposition
if and only if it satisfies these basic conditions.

(2) This, however, is a bad philosophical picture: it is more correct to say that
there is a language game about propositions and that a proposition is automatically
whatever is determined by the use of signs in that game (*PI* §137).

(3) But since you can always automatically add as a prefix the phrase "This is how things
are:" to any proposition, or add as a suffix the phrase "is true" to any proposition, it
seems that necessarily, any part of language is a proposition if and only if it satisfies this
condition (*PI* §137).

(4) Similarly, it seems that any part of language has meaning if and only if it
satisfies the sense of a sentence that I understand (and in this connection, see Frege's
famous remark: "a word has meaning only in the context of a whole proposition," which
is also known as *the context principle*). And understanding is "grasping" the meaning of a
word or other expression in a "flash." If a flash-grasping understanding is possible, then
this contradicts the thesis that the meaning of a word is its *use* (*PI* §138). But what is the
understanding of a word?

(5) The understanding of a word is neither a picture that comes before my mind
when I hear a word, nor a picture plus a method of projection from the picture, because

> (i) the same mental picture/projection method can be correlated with different
> applications of the word (*PI* §§139-140),

and

> (ii) the same application can occur without the occurrence of that mental picture
> or projection method (*PI* §141).

(6) Consider the example of understanding how to complete a series by writing down
signs representing the natural numbers (*PI* §§140-148). Here the understanding of a word
is neither a state of consciousness nor a mental process because

(i) mental states have temporal duration, whereas understanding does not (*PI* §59e),

and

(ii) to hold that understanding is a mental process is to confuse the characteristic *accompaniments* of understanding—which can vary widely across contexts—with understanding itself (*PI* §§149-152).

(7) This argument requires two implicit premises in order to be valid. Here is the first implicit premise: mental pictures, rules of projection, states of consciousness, and mental processes exhaust the possible inner determinants of understanding.

(8) And here is the second implicit premise: the determinants of understanding are either inner or outer.

(9) Therefore, since understanding is after all determined by *something*, it can only be determined by something *outer*: by the manifest or behavioral mastery of a linguistic technique (*PI* §150), and by the "particular circumstances" or context of displaying that mastery (*PI* §§154-155). (Thesis 1)

V.2.7.2 Second Prolegomenon to The Rule Following Paradox: Understanding and Reading

(1) Let us consider now another example, this time of a simplified form of mastery of a linguistic technique that does not itself involve understanding: reading, where this is the activity of rendering out loud what is written or printed, writing from dictation, writing out something printed, following a score, etc. (*PI* §156).

(2) There is no single set of necessary and sufficient conditions (a definition or criterion) for mastery of this linguistic technique. Consider, e.g.,attentive reading, human "reading machines," beginning readers, etc. (*PI* §§156-158). We are tempted to say that the criterion for reading is the conscious act of reading (*PI* §159), but even if the conscious act of reading were lacking—imagine a "reading zombie"—it is conceivable that such a creature could still be a reader (*PI* §160). So consciousness is not the criterion of mastery.

(3) What then about "deriving the reproduction from the original" as a criterion of mastery of this linguistic technique? The problem with this is that even if someone never sticks to a single method of derivation, we can still call him a reader (*PI* §163).

(4) So there is no single sort of mastery of a technique: even for reading, there is a *family* of criteria for what counts as reading (*PI* §164), and there is no single specific marker of what will count as a genuine reading (*PI* §§165-168), because reading can always possibly occur without any such candidate single specific marker.

(5) Even if there is no single specific marker, however, it is true that reading always involves some sort of causal influence between the letters and the reading (*PI* §169). More generally, in all cases of reading I let myself be guided by the letters (*PI* §170).

(6) Therefore mastery of a linguistic technique always involves "being guided" by the linguistic basis of the technique. This could also be equivalently described as the experience of having the sound of the word "intimated" to me by the letters, such that there is a manifest *unity* between word and sound (*PI* §171). (*Thesis 2.*)

V.2.8 The Rule Following Paradox

The central parts of *Investigations* are generally known as *the rule following considerations*, because they are mainly given over to developing and then resolving a deep skeptical worry about the notion of following a rule: namely, The Rule Following Paradox. The basic rationale behind The Rule Following Paradox has three elements.

First, The Rule Following Paradox exposes a fatal flaw in Rule-Based Semantics, according to which the meaning of a linguistic sign is nothing but a rule for manipulating or operating with that sign in some logical or mathematical calculus, or other non-formal language-system. This in turn exposes a fatal flaw in any *function-based, compositional* theory of meaning, according to which the meaning of a complex expression is nothing but a function of the meanings of its simple parts, since such functions are taken to provide rules for computing the meaning of any expression in the language-system, no matter how long and complex, thus explaining how infinitely large languages (whether natural or artificial, e.g., arithmetic) are learnable by finite cognizers like us from finite informational and behavioral inputs.

Second, the notion of following a rule is essential to later Wittgenstein's own positive conception of linguistic understanding as manifest or behavioral mastery of linguistic techniques in context. But given later Wittgenstein's conception of philosophy as the active achievement of clarity by stating descriptive truisms in the right way, the correct characterization of rule following can emerge and be philosophically illuminating

only indirectly, by revealing the inadequacy and incoherence of various characterizations of rule following constrained by bad philosophical pictures.

Third, the leading *in*adequate characterization of rule following is also a version of Solipsistic Semantics—according to which the meaning of a name, sentence, or other linguistic sign is nothing but conscious mental representation or idea in the mind of some individual speaker. Hence the rejection of this inadequate characterization is *also* a crucial part of later Wittgenstein's critique of Solipsistic Semantics. That is, the rejection of that conception of rule following is also a crucial part of The Private Language Argument. Here is a reconstruction of The Rule Following Paradox as later Wittgenstein *himself* understood it.

V.2.8.1 Wittgenstein's Very Own Rule Following Paradox Reconstructed

(1) The Rule Following Paradox begins with an implicit premise: Assume that the meaning of any linguistic sign is nothing but a rule for operating with that sign in some logical or mathematical calculus, or other non-formal language-system (i.e., Rule-Based Semantics).

(2) Therefore, understanding the meaning of any linguistic sign S is being able to follow the rule for operating with S, i.e., being "guided" by the rule for S (*PI* §§ 172-184).

(3) Every rule is expressible as a function-sign which determines a systematic mapping from inputs, or arguments of the function, to outputs, or values of the function (*PI* §§143-146, 151, 185).

(4) Moreover, the meaning of that function-sign—hence the complete set of its systematic mappings—is understood by grasping the rule in a flash (*PI* §§186-197).

(5) But every function-sign can be multiply differently interpreted, such that although the interpretations yield the same mappings to outputs/values for all existing inputs/arguments, they diverge on some future inputs (*PI* §185).

(6) And since every interpretation is in turn expressible as a higher-order function sign, then each interpretation itself stands in need of further interpretation, which itself in turn can be multiply differently interpreted, ad infinitum (*PI* §198).

(7) So anything the speaker does with S can, on some interpretation or another, be in accordance with the rule (*PI* §201).

(8) Correspondingly, anything the speaker does with S can, on some interpretation or another, be also in conflict with the rule (*PI* § 201).

(9) So the speaker's actions, no matter what they are, neither accord with the rule nor conflict with the rule (*PI* § 201).

(10) Therefore it is impossible for a speaker to follow a rule.

(11) Therefore it is impossible for a speaker to understand the meaning of an expression.

(12) So Rule-Based Semantics is false, by reductio.

So much for Rule-Based Semantics. But here is an important complication for Wittgenstein's Very Own Rule Following Paradox. Even if Rule-Based Semantics is outright rejected by the argument we just surveyed, there is still a serious leftover problem, precisely because later Wittgenstein *himself* is also committed to a version of step (2) in the argument, i.e.,

> understanding the meaning of any linguistic sign S is being able to follow the rule for operating with S, i.e., being "guided" by the rule for S,

This commitment is determined by later Wittgenstein's antecedent commitment to Thesis 1 and Thesis 2, which as we saw followed respectively from the considerations on *understanding* and *reading* that prefaced the rule following considerations. So Wittgenstein's Very Own Rule Following Paradox requires a more adequate and deeper solution.

V.2.9 Kripkenstein's Rule Following Paradox: Why Read Kripke Too?

In the early 1980s, in *Wittgenstein, Rules, and Private Language*, Saul Kripke worked out a creative interpretation of the *Investigations* that in the intervening 30 years has come to take on a philosophical life of its own, both in the secondary literature on later Wittgenstein and also in the primary literature on the rule following considerations. Kripke's interpretation focuses on The Rule Following Paradox and its solution in *PI* §§134-242 and argues

(1) that this constitutes the essence of The Private Language Argument, which other commentators have almost always placed in *PI* §§ 243-315,

and

(2) that the rule following considerations introduce a radically new form of philosophical skepticism that should be taken every bit as as seriously as Cartesian evil-demon or dream skepticism in the *Meditations on First Philosophy* and as Hume's skepticism about induction and the concept of necessity in the *Treatise* and *Enquiry concerning Human Understanding*.[96]

The result is strictly speaking neither later Wittgenstein's very own argument nor Kripke's own original argument but in effect a philosophical hybrid also known as "Kripkenstein's Rule Following Paradox." Whatever its merits as a faithful interpretation of the *Investigations*, Kripkenstein's Rule Following Paradox is nevertheless a perfect example of Wittgensteinian philosophy that actually obeys Wittgenstein's own dictum in the preface to the *Investigations*:

I should not like my writing to spare other people the trouble of thinking. But, if possible, to stimulate someone to thoughts of his own. (x^e)

So Kripkenstein's Rule Following Paradox is well worth looking at both for its own sake and also for the light it indirectly casts on the doctrines of the *Investigations* itself. Here now is a reconstruction of the Kripkenstein Paradox.

V.2.9.1 Kripkenstein's Rule Following Paradox Reconstructed

(1) Consider any meaningful use of language but more specifically any meaningful mathematical use of language and in particular our everyday use of the word 'plus' and the symbol '+'. It is just a given fact that by means of my external symbolic representation and also my internal mental representation I *grasp* the rule for addition.

(2) Although I have computed only finitely many sums in the past, the rule for addition determines my answer for the indefinitely many sums that I have never considered. Indeed, the arithmetic function corresponding to the rule for addition determines a complete collection of infinitely many values/outputs for infinitely many arguments/inputs to that function.

(3) Suppose however that I compute '68+57' for the first time. I am confident that the correct answer is '125' and it is also true that the plus function when applied to the inputs 68 and 57 yield 125 as the output and that 'plus' as I intended to use it in the past denoted a function which when applied to the numbers I called '68' and '57' yields the value 125.

(4) But now a "bizarre skeptic" challenges my answer on the following grounds: that I might have intended (and indeed might now be intending) to use 'plus' such that the correct answer is in fact '5' and that the correct value of the function I intended is 5! That is because

> (i) in the past I computed only finitely many sums & by hypothesis had never encountered '68+57' (and let us assume for simplicity also that I had always referred to natural numbers less than 57),

and

> (ii) it is therefore possible that the rule I followed (and am following) corresponded in fact to the function *quus*:

> > If either x or y is less than 57, then x quus y = x + y, but if either x or y is greater than or equal to 57, then x quus y = 5.

So the rule following skeptic claims that I am misinterpreting my own previous (and present) usage: by 'plus' or '+' I always meant (and am currently meaning) quus, *not* plus.

(5) Any adequate reply to the rule following skeptic must satisfy two conditions:

> (i) it must give an account of what fact it is about my mental state that constitutes my meaning plus, not quus,

and

> (ii) it must show how I am normatively justified in giving the answer '125' to '68+57', that is, it must show that I *should* give the answer '125'.

(6) But there is *no mental fact about me*, whether it is an occurrent mental representation such as a mental image or an image together with a projection that interprets it, a mental disposition, a mental state or process, or even a unique phenomenal *quale* uniformly associated with my use of 'plus' and '+', that uniquely determines what I meant (and currently mean) by the use of those symbols, and therefore no mental fact about me that determines that I meant (and currently mean) plus and not quus, precisely because the existence of each of those mental facts can be interpreted consistently with the hypothesis that I actually meant (and am currently meaning) quus and not plus, or that (mutatis mutandis) I am "quounting" and not counting, etc. Indeed there is no mental fact about me that determines that I meant (and am currently meaning) *any definite function whatsoever* by 'plus' or '+'. Thus I might have meant (and currently mean) nothing definite at all! Not even *God* could tell, by inspecting my mental facts, whether I am following the rule for plus or quus.

(7) So I have no normative justification for my claim that the correct answer to '68+57' is '125' and that the corresponding value of the function is 125.

V.2.10 How to Solve The Rule Following **Paradox(es): Wittgenstein's Way and Kripke's Way**

There are two crucial differences between Wittgenstein's Very Own Rule Following Paradox and Kripkenstein's Rule Following Paradox.

The first difference is that later Wittgenstein's very own version directly concerns *semantic theory* and is aimed against Rule-Based Semantics, whereas Kripkenstein's version directly concerns the *epistemology* of normative justification and is aimed against at Solipsistic Semantics. To be sure, both later Wittgenstein's version and Kripkenstein's version agree in the idea that the Private Language Argument substantially depends on the results of the Rule Following Paradox, so they converge on the philosophy of mind. But the fact remains that Kripkeinstein's Rule Following Paradox mainly concerns the epistemology of rule-following and *not* rule-theoretic approaches to semantics.

The second and even more important difference arises from the fact that Wittgenstein's very own version of The Rule Following Paradox takes essentially the form of a non-classical logical paradox about the very idea of a rule, just as the Liar is a classical logical paradox about the very idea of truth or as Russell's Paradox is a classical logical paradox about the very idea of a logical collection or set. By sharp contrast however, Kripkenstein's Rule Following Paradox is essentially concerned with *skepticism* about normative justification, not with logical paradoxes. But Wittgenstein himself does not seem to have had precisely this sort of skepticism, or indeed any sort of philosophical skepticism, primarily in mind.

Crucial as it is, this difference may have a largely contextual and historical origin. Epistemological and skepticism-oriented readings of Wittgenstein were very much in the air at Harvard in the 1960s, 70s, and 80s, and are directly reflected in important work in

epistemology from that period by Stanley Cavell, Thompson Clarke, and Barry Stroud.[97]

Perhaps not too surprisingly at this point in our examination of the fundamental

relationships between Kant's Critical philosophy, Wittgenstein's earlier and later

philosophy, and *transcendental* philosophy more generally, this Harvard-centered

epistemological work was as much *Kant*-influenced as it was *Wittgenstein*-influenced. No

doubt it followed on more or less directly from C.I. Lewis's epistemological neo-

Kantianism in the first half of the century, together with Quine's skeptical attack on the

analytic-synthetic distinction, and Strawson's transcendentalist response to Quine. Not

entirely coincidentally, Kripke himself did a BA in mathematics at Harvard from 1958-

62, and was subsequently both a Junior Fellow in the Harvard Society of Fellows and

taught in the Department of Philosophy until he moved to Rockefeller University in 1967.

In any case, Kripke wants to insist that any solution to The Rule Following

Paradox, as he understands it, can therefore only be a *skeptical* solution that respects the

skeptical result, as opposed to a *straight* solution that accepts the constraints in (5) and

also shows that—contrary to (6) and (7)—there is some mental fact about me which

determines that I meant plus and not quus, and that I therefore have a normative

mentalistic justification (i.e., a good reason directly citing the relevant mental fact) that

the correct answer to '68+57' is '125' and that the corresponding value is of the function

is 125. I will look closely at Kripke's skeptical solution shortly, but just to anticipate

now, it consists in claiming

> (i) that I do indeed mean plus and not quus without *any* inner justification or reasons and
> thus follow the rule "blindly" (see *PI* §219)

and

> (ii) that no rule can be obeyed only "privately" (see *PI* §202): on the contrary rules are
> followed by individual speakers only in the context of social practices (language-games

and forms of life) & are legitimated non-cognitively solely in terms of those social practices alone: *this is simply what I do* (*PI* §217), because this is simply what *we* do.

As we will see, this conclusion brings Kripke's solution to Kripkenstein's Rule Following Paradox very close to Hume's solution to his own skepticism about causation and induction, and also to Goodman's solution to the "new riddle of induction." Later Wittgenstein himself, by striking contrast, proposes a solution to his very own Rule Following Paradox that is essentially *Kantian*, not Humean. Here is a reconstruction of later Wittgenstein's solution to his very own Rule Following Paradox, in the key of Kant.

V.2.10.1 Later Wittgenstein's Solution to his Very Own Rule Following Paradox Reconstructed

(1) We start with the following assumptions, justified by earlier arguments:

(i) that linguistic understanding actually occurs,

(ii) that all linguistic understanding is determined by something outer, not by something inner, and involves manifest or behavioral mastery of a linguistic technique, together with the particular circumstances or context of displaying that mastery,

(iii) that mastery of a linguistic technique always involves "being guided" by the linguistic basis of the technique (as, e.g., in reading),

(iv) that understanding the meaning of any linguistic sign S is being able to follow the rule for using S, i.e., being guided by S,

(v) that every rule is expressible as a function-sign determining a systematic mapping from inputs (arguments to the function) to outputs (values of the function),

(vi) that the meaning of a function-sign is understood by grasping the rule in a flash (i.e., mentally and instantaneously),

(vii) that a function-sign is given a meaning by virtue of assigning an interpretation to the function-sign, which in turn is itself expressible as a higher-order function-sign,

and

(viii) that all such function-signs can be multiply differently interpreted.

(2) From these assumptions, it follows that:

> No course of action could be determined by a rule because every course of action can be determined by the rule…. [I]f everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here. (*PI* § 201, 81ᵉ)

In other words: rule following is impossible.

(3) Therefore rule following both actually occurs and is also impossible. Paradox!

(4) Here is an implicit premise drawn from classical logic: given a contradictory conclusion, at least one of its premises must be false.

(5) It is false that a function-sign is given a meaning by virtue of assigning an interpretation to the function-sign, which in turn is itself expressible as a higher-order function-sign:

> It can be seen that there is a misunderstanding here from the mere fact that in the course of our argument we gave one interpretation after another; as if each one contented us for a moment, until we thought of yet another standing behind it. What this shows is that there is a way of grasping a rule which *is not an interpretation*, but which is exhibited in what we call "obeying the rule" and "going against it" in particular cases. (*PI* §201, 81ᵉ)

In other words: assumption (vii) above is false, and must be replaced by the new thesis that how a rule is given a meaning is fully displayed by *the speaker's actual acts of obeying the rule or going against the rule*.

(6) Given the truth of (5), it is also false that the meaning of a function-sign is understood by grasping the rule in a flash (i.e., mentally and instantaneously):

> And hence 'obeying a rule' is a practice. And to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule 'privately': otherwise thinking one was obeying the rule would be the same thing as obeying it. (*PI* § 202, 81ᵉ)

In other words: assumption (vi) above is also false, and must be replaced by the new thesis that the speaker's actual acts of obeying the rule or going against the rule are *practices*, or essentially *social* enterprises of manifest or behavioral mastery of a linguistic technique, together with the particular circumstances or context of displaying that mastery.

(7) The practice of following a particular rule is a normative activity involving both some type of imperative constraint on the rule-follower and also training in the linguistic techniques involved in the practice:

> Following a rule is analogous to obeying an order. We are trained to do so: we react to an order in a certain way (*PI* § 206, 82ᵉ)

(8) But rule following is *not* a practice that also requires *my* giving reasons for acting in the way I do, hence it does not require *my* providing an *internal* justification:

How can he *know* how he is to continue a pattern by himself—whatever instructions you give him?—Well, how do I know?—If that means. "Have I reasons?" the answer is: my reasons will soon give out. And then I shall act, without reasons. (*PI* § 211, 84ᵉ)

(9) On the contrary, the practice of following a particular rule itself *externally* justifies what I do, insofar as I merely *engage* in that practice:

"How am I to obey a rule?" —If this is not a question about causes, then it is about the justification for my following the rule in the way I do. If I have exhausted the justifications I reach bedrock, and my spade is turned. Then I am inclined to say: "This is simply what I do." (*PI* §217, 85ᵉ)

(10) And this external justification also provides the basis for projecting the rule infinitely into the future:

Whence comes the idea that the beginning of a series is a visible section of rails invisibly laid to infinity? Well, we might imagine rails instead of a rule. And infinitely long rails correspond to the unlimited application of the rule. (*PI* §218, 85ᵉ)

"All the steps are really already taken" means: I no longer have any choice. The rule, once stamped with a particular meaning, traces the lines along which it is to be followed through the whole of space. —But if something of this sort were really the case, how would it help? No; my description only made sense if it was to be understood symbolically. —I should have said: *This is how it strikes me.* When I obey the rule I do not choose. I follow the rule *blindly.* (*PI* § 219ᵉ)

(11) My following a rule therefore depends on my non-cognitively entering into an agreement (and here we are to think comparatively and contrastively about the very idea of a *social contract*) with other people who belong to the same practice:

The word 'agreement' ('*Übereinstimmung*') and the word 'rule' ('*Regel*') are related to one another, they are cousins. If I teach anyone the use of the one word, he learns the use of the other with it. (*PI* §224, 86ᵉ)

(12) But this agreement does *not* imply a non-truth-conditional theory of truth: on the contrary, truth is still correspondence to the facts and the agreement consists in the more basic sharing of a language [-game] and of a form of life:

"So you are saying that human agreement decides what is true and what is false?—It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life. (*PI* §241, 88ᵉ)

In other words: Wittgenstein is a *radical externalist* about justification and understanding (both as to its semantic content and its representational vehicle) but also an *internal realist* about truth.

(13) This agreement in language-game and form of life, in turn, is given essentially in the *activity* of making judgments:

If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments. This seems to abolish logic, but does not do so. (*PI* §242, 88ᵉ)

(14) So the fundamental agreement that is in play both in language-games and in forms of life *consists in our intersubjectively shared capacity for judgment*.

In this way, later Wittgenstein's own solution to his Very Own Rule Following Paradox, bounded in a nutshell, is that we can follow rules just because talking rational human animals like us are nothing more and nothing less than *judging animals*.

Correspondingly, and for the purposes of critical comparison and contrast, here is a reconstruction of Kripke's solution to Kripkenstein's Rule Following Paradox, in the key of Hume.

V.2.10.2 Kripke's Solution to Kripkenstein's Rule *Following Paradox Reconstructed*

(1) By virtue of (what Kripke takes to be) The Rule Following Paradox, Wittgenstein is committed to a radical skepticism about the determination of future linguistic usage by the past contents of my mind. This is fundamentally analogous to Hume's skepticism about the determination of the future by the past, where we understand this skepticism both inferentially (skepticism about induction) and also causally (skepticism about natural necessity).

(2) Kripkenstein's Rule Following Paradox can therefore be resolved only by a "skeptical solution" which accepts both

> (i) that there is no mental fact about me that determines whether I am following the rule for plus or the rule for quus,

and also

> (ii) that I have no internal justification for my claim that the correct answer to '68+57' is '125', and then turns instead to look purely descriptively at the actual circumstances under which I can be correctly said to be following plus rather than quus and in which it can be asserted that the correct answer to '68+57' should be '125'.

(3) If we consider a single individual in isolation, then although it is an empirical fact that the individual does confidently assert, or at least has the disposition to confidently assert, that the correct answer to '68+57' is '125', nevertheless (by step (2) (ii)) there is no internal justification for this assertion.

(4) But if we take into account the fact that the individual is in a community, then the philosophical picture radically changes and we must adopt an assertibility-conditions semantics (according to which a statement is true if and only if it is legitimately

assertible) and reject a truth-conditional semantics (according to which a statement is true if and only if it corresponds to the facts).

(5) The empirical fact of our successful rule-following practices (see step (3)) depends essentially on the further brute empirical fact that we agree with one another in our responses to questions like "What is 68+57?"

(6) Hence the relevant assertibility-condition for the answer '125' is simply whether the individual's response agrees with everyone else's response to the same question, and this external judgment is determined simply by observing the individual's behavior and surrounding circumstances. This solution to the Rule Following Paradox in turn is fundamentally analogous to Hume's claim to have shown that the only way to make sense of a causal relation between two phenomenal events is simply to subsume it under a customary or habitual regularity of constant conjunctions of instances of the relevant event-types.

(7) Therefore (according to Kripke) the Wittgensteinian thesis that there is no private language is necessarily equivalent to the Wittgensteinian thesis that there is no private rule-following.

V.2.11 What is a Private Language?

All readers of the *Investigations* agree that one of its central achievements is The Private Language Argument: i.e., later Wittgenstein's demonstration of the impossibility of (or: the incoherence of the concept of) a private language. But unfortunately very few readers agree *either* about

(1) what precisely a "private language" is,

or about

(2) what The Private Language Argument actually is.

But obviously since no progress can be made on the second question unless the first question has been adequately answered, that is where I will start.

First and foremost, a private language is a *solipsistic language* in the sense that it is a language whose *meanings* are nothing but mental representations (or "ideas") in the mind of an individual speaking subject. A solipsistic language of this sort is such that *only one person can understand it, because its meanings or semantic contents are*

determined wholly and solely by what is inside that person's head (or alternatively: *inside that person's Cartesian soul*, if you are a substance dualist). I will call any language that is solipsistic in this way *a language that is solipsistic with respect to its semantic content*.

But unfortunately even the notion of a language that is solipsistic with respect to its semantic content is ambiguous, because there are at least two disjointly different classes of subjective mental representations that might be identified with linguistic meanings:

(i) sensations (i.e., phenomenal qualia or phenomenally conscious mental states),

and

(ii) other mental items (i.e., other sorts of mental states, mental processes, mental images, mentalistic concepts, rule-following impressions, etc.).

As a consequence there are at least two different kinds of language that are solipsistic with respect to their semantic contents:

(i) *sensation languages*, i.e., languages in which words have meaning by standing for an individual speaker's sensations,

and

(ii) *non-sensational mentalistic languages*, i.e., languages in which words have meaning by standing for mental states, mental processes, mental images, mental concepts, rule-following impressions, etc., of an individual speaker *other than* her sensations.

And in fact later Wittgenstein wants to argue against the possibility of *both* sensation languages and non-sensational mentalistic languages.

In a *second* sense, however, a private language can *also* be a solipsistic language in the quite different sense of a "mental language" or *lingua mentis*, that is, a language whose *words* (types and tokens alike) are nothing but mental representations in the mind of an individual speaking subject. A solipsistic language of this sort is such that only one

person can understand it because its *grammatically-structured signs or symbols* are determined wholly and solely by what is inside the head (or Cartesian soul) of a single speaking subject. I will call languages of this sort *languages that are solipsistic with respect to their syntactic vehicles*.

From the standpoint of clearly understanding The Private Language Argument, the unfortunate thing about private languages that are solipsistic with respect to their syntactic vehicles is that they are *not* necessarily equivalent with private languages that are that are solipsistic with respect to their semantic contents. And that is because of the following two facts:

> (I) It is possible for there to be languages that *are* solipsistic with respect to their *semantic contents*, but also are *not* solipsistic with respect to their *syntactic vehicles*. These languages would include sensation languages that are also public natural languages. For example, according to Phenomenalists (say, the early Logical Empiricists or Logical Positivists), the ordinary English sensation-word 'pain' would mean *this painy sensation now*.

> (II) It is possible for there to be private languages that *are* solipsistic with respect to their *syntactic vehicles*, but also are *not* solipsistic with respect to their *semantic contents*. These languages would include any mental language or *lingua mentis* which has a direct translation into a public natural language. For example, my mental word '##' could mean the same as 'beetle' in English.

And if this were not bad enough, there are even also private languages that are actually public with respect to their semantic vehicles but also *trivially solipsistic with respect to their syntactic vehicles*, e.g., Robinson Crusoe's monologues on his island before encountering his man Friday, or the text of Lewis Carroll's *Jabberwocky* before he actually showed it to anyone else (or perhaps even before he actually wrote it down).

Now what is the point of drawing all these distinctions? The answer is that the private languages which are the target of The Private Language Argument are *just these*:

> All and only languages that are solipsistic with respect with their semantic contents, whether or not they are also solipsistic with respect to their syntactic vehicles, including all sensation languages and non-sensational mentalistic languages.

And this excludes *many* languages that are solipsistic with respect to their syntactic vehicles, as well as *most* languages that are trivially solipsistic with respect to their syntactic vehicles.

Now in the light of all that philosophical stage-setting, here is a reconstruction of The Private Language Argument.

V.2.11.1 The Private Language Argument Reconstructed

(1) Let us consider the possibility of languages that are solipsistic with respect to their semantic contents (as opposed to languages that are solipsistic with respect to their syntactic vehicles, whether trivially or not), and consider the possibility in particular of *sensation languages*.

> But we could also imagine a language in which a person could write down or give vocal expression to his inner experiences--his feelings, moods, and the rest—for his private use? —Well, can't we do so in our ordinary language? —But that is not what I mean. The individual words of this language are to refer to what can only be known to the person speaking; to his immediate private sensations. So another person cannot understand the language. (*PI* §243, 88^e-89^e)

(2) Now such sensation languages are solipsistic with respect to their semantic contents *by virtue of the fact* that the sensations for which the words stand are knowable by the individual speaker alone.

> In what sense are my sensations *private*? —Well, only I can know whether I am really in pain; another person can only surmise it. (*PI* §246, 89^e)

(3) If sensations are to be knowable *in any way* by the individual speaker, then it must also be possible for the speaker to identify and re-identify her sensations over time and across individual persons.

> "Another person can't have my pains." —Which are *my* pains? What counts as a criterion of identity here? Consider what makes it possible in the case of physical objects to speak of "two exactly the same," for example, to say "This chair is not the one you saw here yesterday, but is exactly the same as it." In so far as it makes *sense* to say that my pain is the same as his, it is also possible for us both to have the same pain. (*PI* §253, 91^e)

(4) But if sensations are knowable by the individual speaker alone, then that speaker's identification and re-identification of those sensations over time will lack any intersubjectively valid criteria for correctness.

> Let us imagine the following case. I want to keep a diary about the recurrence of a certain sensation. To this end I associate it with the sign "E" and write this sign in a calendar for every day on which I have the sensation. —I will remark first of all that a definition of the sign cannot be formulated. —But I can still give myself a kind of ostensive definition. —How? Can I point to the sensation? Not in the ordinary sense. But I speak, or write the

sign down, and at the same time I concentrate my attention on the sensation—and so, as it were, point to it inwardly. —But what is this ceremony for? for that is all it seems to be! A definition surely serves to establish the meaning of a sign. —Well, that is done precisely by the concentration of my attention; for in this way I impress on myself the connection between the sign and the sensation. —But "I impress it on myself" can only mean: this process brings it about that I remember the connection *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. (*PI* §258, 92ᵉ)

(5) But if an individual speaker's identification and re-identification of sensations over time lack any intersubjectively valid criteria of correctness, then it is also possible for everyone to believe that they are sharing the same sensation yet still have different sensations.

> The essential thing about private experience is really not that each person possesses his own examplar, but that nobody knows whether other people also have *this* or something else. The assumption would thus be possible—though unverifiable—that one section of mankind had one sensation of red and another section another. (*PI* §272, 95ᵉ)

(6) And if it is possible for everyone to believe that they are sharing the same sensation yet have different sensations, then it is also possible for everyone to have *no* sensations at all: in which case it is impossible to determine whether the sensation-word has any meaning at all.

> If I say of myself that it is only from my own case that I know what the word 'pain' means—must not I say the same thing of other people too? And how can I generalize the *one* case so irresponsibly? Now someone tells me that *he* knws what pain is only from his own case! —Suppose that everyone had a box with something in it: call it a "beetle." No one can look into anyone else's box, and everyone says that he knows what a beetle is only by looking at *his* beetle. —Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. —But suppose the word "beetle" had a use in these people's language? —If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*; for the box might even be empty. —No, one can "divide through" by the thing in the box; it cancels out, whatever it is. (*PI* §293, 100ᵉ)

(7) So sensation languages are impossible.

(8) And by a simple generalization of the same argument, non-sensational mentalistic languages—especially those in which words have meaning by standing for *rule-following impressions*—are also impossible.

> And to *think* one is obeying a rule is not to obey a rule. Hence it is not possible to obey a rule "privately": otherwise thinking one was obeying a rule would be the same thing as obeying it. (*PI* §202, 81ᵉ)

> Are the rules of the private language *impressions* of rules?—The balance on which impressions are weighed is not the *impression* of a balance. (*PI* §259, 92ᵉ)

(9) So private languages are impossible, and it follows that linguistic meanings or semantic contents are *not* determined wholly and solely by what is inside individual speakers' heads (or their Cartesian souls).

The conclusion clearly states later Wittgenstein's *anti-individualism about semantic content.* So, otherwise put, The Private Language Argument is ultimately an argument for *externalism about semantic content*—the thesis that linguistic meanings or semantic contents are determined at least partially by what is outside individual speakers' heads (or Cartesian souls). The Private Language Argument *also* indirectly shows that later Wittgenstein is a radical syntactic vehicle externalist. For it is not signs per se, and especially not signs in a *lingua mentis* or mental language, but instead only *public uses of signs by judging animals like us*, that have linguistic meaning or semantic content.

V.2.12 Is Later Wittgenstein a Behaviorist?

As we have seen The Private Language argument, if sound, shows that private languages—i.e., languages that are solipsistic with respect to their semantic content, i.e., languages such that their meanings or semantic contents are strictly determined by what is inside an individual speaker's head (or Cartesian soul)—are impossible. So it linguistic meanings or semantic contents are determined at least partially by what is outside the speaker's head (or Cartesian soul), and externalism about semantic content is true. Perhaps even more controversially, however, later Wittgenstein also believes that he can advance from this negative conclusion to a positive doctrine about the nature of mental states, and in particular about the nature of sensations, i.e., states of phenomenal consciousness.

One outstanding interpretive question about later Wittgenstein's doctrine is whether it is form of metaphysical behaviorism. Metaphysical behaviorism says that

mental properties are identical to behavioral properties, where behavioral properties are second-order physical properties consisting in either

> (i) a set of occurrent causal mappings from stimulus inputs to living organisms, to response outputs from those organisms, i.e., various natural or unlearned bodily movements, orientations, positionings, or sounds—and this is "animal behavior" in the broadest possible sense,

or

> (ii) a set of dispositions to animal behavior.

By contrast, *methodological* behaviorism says that scientific psychology should be conducted *as if* mental properties are identical to behavioral properties. And *semantic* behaviorism says that the linguistic meanings or semantics contents of mentalistic terms are nothing but rules for verifying or falsifying judgments about animal behavior.[98]

The Logical Empiricists or Logical Positivists were metaphysical, methodological, and semantic behaviorists. One big problem with metaphysical and semantic behaviorism alike, as Putnam famously pointed out,[99] is that it is a priori conceivable and therefore logically possible that there is a race of humanoids who have completely suppressed natural human pain-behavior and yet still feel pain ("super-spartans"). Correspondingly, it is also a priori conceivable and therefore logically possible that there is a race of humanoids who have fully developed natural human pain-behaviors and yet completely lack pain experiences accompanying those behaviors ("super-fakers"). Indeed, the truth of metaphysical behaviorism is a priori conceivably and logically consistent with all human beings lacking all phenomenal consiousness whatsoever. So for all that metaphysical behaviorism tells us, we might be *zombies* in the philosophical sense, i.e., perfect microphysical and behavioral duplicates of us as we actually are, only *without* our mental lives—all the lights are on, but no one is ever home.

If so, then metaphysical behaviorism cannot possibly be correct. For if zombies are possible, then mental properties are *not* identical with physical properties, and therefore the human mind *cannot* be reductively explained in terms of behavioral properties.[100]

My own view is that later Wittgenstein is *neither* a metaphysical behaviorist, *nor* a methodological behaviorist, *nor* a semantic behaviorist. On the contrary, he is what I will call a *human capacity behaviorist*, which is a uniquely non-reductive, transcendental-anthropocentric communitarian form of behaviorism that is fully consistent with the non-supervenient existence and fully finegrained character of rational human phenomenal consciousness and also with Weak or Counterfactual Transcendental Idealism in its transcendental-anthropocentric communitarian guise. So let us now see how this strikingly original view plays out, according to the following reconstruction .

V.2.12.1 Later Wittgenstein's Human Capacity Behaviorism Reconstructed

(1) One basic result of The Private Language Argument is that sensation-languages are impossible—which is to say that languages containing sensation-words which are solipsistic with respect to semantic content are impossible.

(2) From this result it follows that the meanings of sensation-words are *not* strictly determined by what is inside individual speakers' heads (or Cartesian souls).

(3) Here is an implicit premise: The meanings of words are strictly determined either by what is inside the heads (or Cartesian souls) of individual speakers or by what is outside the heads (or Cartesian souls) of individual speakers, and there are no other alternatives.

(4) So the meanings of sensation-words are strictly determined *at least partially* by what is outside the heads (or Cartesian souls) of individual speakers.

(5) Sensation-words refer to private sensory experiences. But sensations are private only in the entirely tautological sense that *only the person who actually has a given sensation, actually has that sensation*.

> Other people cannot be said to learn of my sensations *only* from my behavior,—for *I* cannot be said to learn of them. I *have* them. (*PI* §246, 89ᵉ)

> The proposition "Sensations are private" is comparable to: "One plays patience by oneself." (*PI* §248, 90ᵉ)

(6) The truistic or ordinary privacy of sensations, however, is consistent with the thesis

that the meaning of a sensation-word is at least *partially* strictly determined by human sensation-behavior, i.e., by the natural bodily expressions of sensation (e.g., grimacing or wincing when in pain, clutching the affected spot, etc.).

> How do words *refer* to sensations? … Here is one possibility: words are connected with the primitive, the natural expressions of the sensation and used in their place. A child has hurt himself and cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior. "So you are saying that the word 'pain' really means crying?" —On the contrary, the verbal expression of pain replaces (*ersetzt*) crying and does not describe it. (*PI* §244, 89ᵉ)

> *How* do I use words to stand for my sensations? —As we ordinarily do? Then are my words for sensations tied up with my natural expressions of sensation? In that case my language is not a 'private' one. Someone else might understand it as well as I. (*PI* §256, 91ᵉ)

> An "inner process" stands in need of outer criteria. (*PI* §580, 153ᵉ)

> The human body is the best picture of the human soul. (*PI* 178ᵉ)

(7) If (4), (5), and (6) are all true, then a necessary condition of having a sensation *is being the kind of creature that has the capacity to express sensation-behavior*—i.e., living (rational) human beings and other creatures that behave like living (rational) human beings.

> "But doesn't what you say come to this: that there is no pain, for example, without pain-behavior? –It comes to this: only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (*PI* §281, 97ᵉ)

> Only of what behaves like a human being can one say that it *has* pains. (*PI* §283, 98ᵉ)

(8) This however does *not* imply that the subject of sensation is the human body as such, or any of its proper parts: instead, the subject of sensation is the *whole individual rational human being or human person*, i.e., a rational human animal, and this is immediately manifest in our ordinary ascriptions of sensations to others.

> But isn't absurd to say of a *body* that it has pain? —And why does one feel an absurdity here? In what sense is it true that that my hand does not feel pain, but I in my my hand? What sort of issue is: Is it the *body* that feels pain? —How is it to be decided? What makes it plausible that it is *not* the body? —Well, something like this: if someone has a pain in his hand, then the hand does not say so (unless it writes it) and one does not comfort the hand, but the sufferer; one looks into his face. (*PI* §286, 98ᵉ)

> If one has to imagine the pain of someone else's pain on the model of one's own, this is none too easy a thing to do: for I have to imagine pain which I *do not feel* on the model of pain which I *do feel.* That is, what I have to do is not simply to make a transition in imagination from one place of pain to another. As, from pain in the hand to pain in the arm. For I am not to imagine that I feel pain ins ome region of his body. (Which would also be possible.) Pain-behavior can point to a painful place—but the subject of pain is the person who gives it expression. (*PI* §302, 101ᵉ)

> "I believe that he is suffering." —Do I also *believe* that he isn't an automaton? It would go against the grain to use the word ['believe', *Glaube*] in both connections. (Or is it like

this: I believe that he is suffering, but am certain that he is not an automaton. Nonsense!) … "I believe that he is not an automaton," just like that, so far makes no sense. My attitude towards him is an attitude towards a soul. I am not of the *opinion* that he has a soul. (*PI* 178ᵉ)

(9) And in turn the fact that the subject of sensation is the whole human person or rational human animal, and not merely the human body of that person also implies that sensation is *not* identical to sensation-behavior: for conceivably and in principle, pain-behaviors can be expressed by human persons without the corresponding pain-sensations (i.e., super-fakers). More generally however, mental states or mental processes are *activities of the whole human person* that must also exist as "inner states" or processes along with their behavior. So it is a priori necessarily false that there could be human behavioral sensation-automata (i.e., zombies in the philosophical sense).

"But you will surely admit that there is a difference between pain-behavior accompanied by pain and pain-behavior without any pain?—Admit it? What greater difference could there be? –"And yet you again and again reach the conclusion that the sensation is itself a *nothing*. --Not at all. It is not a *something*, but not a *nothing* either. The conclusion [of The PLA] was only that a nothing could serve just as well as a something about which nothing could be said. We have only rejected the grammar [of the private language of sensation] which tries to force itself on us here. (*PI* §304, 102ᵉ)

"But surely you cannot deny that, for example, in remembering, an inner process takes place." —What gives the impression that we want to deny anything? (*PI* §305ᵉ)

Why should I deny that there is a mental process? But "There has just taken place in me the mental process of remembering …," means nothing more than: "I have just remembered…." To deny the mental process would mean to deny the remembering; to deny that anyone ever remembers anything. (*PI* §306ᵉ)

(10) In this way, later Wittgenstein's thesis that sensations in particular, and mental states and processes more generally, are activities of the whole human person or rational human animal, requiring at least the capacity to behave, but allowing for the logical possibility of human sensation-behavior without corresponding sensations of that type and also denying the logical possibility of human sensation-behavior automata (i.e., zombies), is *not* a form of metaphysical behaviorism. More generally, metaphysical behaviorism is a *reductive materialist identity thesis* that is essentially based on a "grammatical fiction" or bad philosophical picture:

"Are you not really a behaviorist in disguise? Aren't you at bottom really saying that everything except human behavior is a fiction?" —If I do speak of a fiction, then it is of a *grammatical fiction*. (*PI*

So later Wittgenstein is not a metaphysical behaviorist. On the contrary, as a defender of Human Capacity Behaviorism in the light of his transcendental-anthropocentric communitarian idealism, he is in fact a *transcendental-anthropocentric communitarian idealist essentially embodied mind theorist*,[101] even despite its being rather a mouthful to say. More precisely, what later Wittgenstein is asserting is that synthetic a priori

necessarily, if X is a human person and X has a sensation, then that sensation occurs in and through the entire living human body of that human person or rational human animal.

V.2.13 Linguistic Phenomenology

Later Wittgenstein's philosophy of mind is, I think, most accurately described as *Linguistic Phenomenology*. In this sense, it is essentially a reversion to the basic themes of the early phenomenological tradition—especially Edmund Husserl's *Logical Investigations*—but without Husserl's later Cartesianism, most evident in *Cartesian Meditations*. Moreover, as we have just seen, via his Human Capacity Behaviorism, later Wittgenstein's Linguistic Phenomenology fully avoids the reductive metaphysical, methodological, and semantic behaviorism of the Logical Empiricists or Logical Positivists.

Later Wittgenstein's Linguistic Phenomenology is largely a direct consequence of The Private Language Argument, which entails that

> (i) the meanings of sensation-words are not strictly determined by what is inside individual speakers' heads (or Cartesian souls) = *semantic anti-individualism for sensation language*.

Assuming then that the meanings of words are strictly determined EITHER by what is inside the heads (or Cartesian souls) of individual speakers OR by what is outside the heads (or Cartesian souls) of individual speakers, and that there are no other alternatives, it follows that

> (ii) the meanings of sensation-words are strictly determined at least partially by what is outside the heads (or Cartesian souls) of individual speakers = *semantic externalism for sensation language*.

This is in turn becomes the basis of a positive claim about the meaning of sensation-words:

(iii) the meaning of a sensation-word is at least partially determined by human sensation-behavior types, which function in particular contexts as criteria for the application of sensation-words = *weak semantic behaviorism for sensation language*.

On the basis of these three claims, later Wittgenstein then asserts four positive theses about the nature of the rational human mind:

(Thesis I) *The token privacy of sensations*. Sensation-words refer to private sensory experiences, but sensations are "private" only in the entirely tautological sense that only the person who actually has a particular token of a sensation-type, actually has that particular token.

(Thesis II) *Human capacity behaviorism*. A necessary condition of X's having a token sensation is X's being the kind of creature that has the capacity to express sensation-behavior: living (rational) human beings and other creatures that behave like living (rational) human beings.

(Thesis III) *Sensation personalism*. The subject of a token sensation is the whole individual (rational) human being or human person.

(Thesis IV) *The activist phenomenology of mental states and processes*. Mental states and mental processes are inner activities of the whole human person—not any sort of static objects or things—that can exist in the absence of occurrent behavior.

V.2.14 Two Kinds of Seeing

In light of later Wittgenstein's weak semantic behaviorism and also in light of theses I and IV of his Linguistic Phenomenology, it is fully possible for him to conduct a phenomenological investigation of various types of inner mental states and mental processes just by describing various uses of sensation language. In particular, he examines the phenomenology of visual perception, and argues for the following four-part conclusion:

(i) There are two irreducibly different types of seeing: *direct seeing* ("seeing-this") and *interpretive seeing* ("seeing-as").

(ii) Interpretive seeing requires direct seeing.

(iii) Interpretive seeing requires high-grade cognitive (and in particular, conceptual) capacities, whereas direct seeing does not.

and

(iv) Direct seeing can occur in the absence of interpretive seeing.

For convenience's sake, I will call this four-part doctrine *Dual Seeing*. In turn, here is a reconstruction of later Wittgenstein's argument for Dual Seeing.

V.2.14.1 **Later Wittgenstein's Argument** for *Dual Seeing Reconstructed*

(1) It is one phenomenon of visual experience to see *this* directly, and another phenomenon to see this *as* having a certain visual aspect.

> Two uses of the word 'see'. The one: "What do you see there?" – "I see *this*" (and then a description, a drawing, a copy). The other: "I see a likeness between these two faces"—let the man I tell this to be seeing the faces as clearly as I do myself. The importance of this is the categorical difference between the two 'objects' of sight. The one man might make an accurate drawing of the two faces, and the other notice in the drawing the likeness which the former did not see. I contemplate a face, and then suddenly notice its likeness to another. I *see* that it has not changed; and yet I see it differently. I call this experience "noticing an aspect." (*PI* 193ᵉ)

(2) Interpretive seeing requires direct seeing.

> The description of the immediate experience, i.e., of the visual experience, by means of an interpretation—is an indirect description. "I see the figure as a box" means: I have a particular visual experience which I have found that I always have when I interpret the figure as a box or when I look at a box. But if it meant this I ought to know it. I ought to be able to refer to the experience directly, and not only indirectly. (*PI* 193ᵉ-194ᵉ)

(3) Interpretive seeing requires high-grade cognitive (and in particular, conceptual) capacities, including abilities for

> (i) multiple interpretations of the same direct visual object (see, e.g., the figures at *PI* 193ᵉ and 200ᵉ),

> (ii) multistability as between different aspects of the same direct visual object (see, e.g., the figures at *PI* 194ᵉ, 203ᵉ, and 207ᵉ—and see also the Necker Cube at *TLP* 5.5423),

> (iii) introduction of three dimensionality into 2D visual objects (see, e.g., the figures at *PI* 203ᵉ and at *TLP* 5.5423),

and

> (iv) organizations of directly-seen shapes into pictorial representations (see, e.g., the figure at *PI* 194ᵉ).

>> You could imagine the following illustration appearing in several places in a book, a text book for instance. In the relevant text something different is in question every time: here a glass cube, there an inverted open box, there a wire frame of that shape, there three boards forming a solid angle. Each time the text supplies the interpretation of the illustration. But we can also *see* the illustration

now as one thing now as another. –So we interpret it, and *see* it as we *interpret* it. (*PI* 193ᵉ)

I shall call the following figure, derived from Jastrow, the duck-rabbit. It can be seen as a rabbit's head or as a duck's. And I must distinguish between the "continuous seeing" of an aspect and the "dawning" of an aspect. (*PI* 194ᵉ)

The change of aspect. "But surely you would say that the picture is altogether different now!" But what is different: my impression? my point of view? –Can I say? I *describe* the alteration like a perception; quite as if the object had altered before my eyes. (*PI* 195ᵉ)

"Seeing as …." is not part of perception. And for that reason it is like seeing and again not like…. If you are looking at the object, you need not think of it; but if you are having the visual experience expressed by the exclamation ["a rabbit!"], you are also *thinking* of what you see. Hence the flashing of an aspect on us seems half visual experience, half thought. (*PI* 197ᵉ)

How does one tell that human beings *see* three-dimensionally? …. The only thing that is natural to us is to represent what we see three-dimensionally; special practice and training are needed for two-dimensional representation whether in drawing or words. (The queerness of children's drawings.) (*PI* 198ᵉ)

Hold the drawing of a face upside down and you can't recognize the expression of the face. Perhaps you can see that it is smiling, but not exactly what *kind* of smile it is. You cannot imitate the smile or describe it more exactly. (*PI* 198ᵉ)

Of course we can say: There are certain things which fall equally under the concept "picture-rabbit" and under the concept "picture duck." And a picture, a drawing, is such a thing. –But the *impression* is not simultaneously of a picture-duck and a picture rabbit. (*PI* 199ᵉ)

Take as an example the aspects of a triangle. This triangle can be seen as a triangular hole, as a geometrical drawing; as standing on its base; as hanging from its apex; as a mountain, as a wedge, as an arrow or a pointer, as an overturned object which is meant to stand on the shorter side of the right-angle, as a half-parallelogram, and as various other things. (*PI* 200ᵉ)

Certain drawings are always seen as flat figures, and others three-dimensionally…. And then it seems queer that with some drawings our impression should be a flat thing, and with some a three-dimensional thing. One asks oneself "Where is this going to end?" (*PI* 202ᵉ)

"Is it a *genuine* visual experience?" The question is: in what sense is it one? Here is it difficult to see that what is at issue is the fixing of concepts. A *concept* forces itself on one. (This is what you must not forget.) (*PI* 204ᵉ)

The aspects of the triangle: it is as if an image came into contact, and for a time remained in contact, with the visual impression. In this, however, these aspects differ from the concave and convex aspects of the step (for example). And also from the aspects of the figure (which I shall call a "double cross") as a white cross on a black ground and a black cross on a white ground. You must remember that the descriptions of the elternating aspects are of a different kind in each case. (*PI* 207ᵉ)

(4) But it is also possible for perceivers to lack the ability to see aspects: to have "aspect-blindness." Therefore direct seeing can occur in the absence of interpretive seeing.

> The question now arises: Could there be human beings lacking in the capacity to see something as *something*—and what would that be like? What sort of consequences would it have? …. We will call it "aspect-blindness"—and will next consider what might be meant by this. (A conceptual investigation.) (*PI* 213ᵉ)

> The [humans who are] "aspect-blind" will have an altogether different relationship to pictures from ours. (Anomalies of *this* kind are easy for us to imagine.) Aspect blindness will be *akin* to the lack of a "musical ear." (*PI* 214ᵉ)

Although later Wittgenstein does not explicitly say this, it also seems very easy to conceive of the possibility of creatures who are capable of direct seeing but also are either characteristically or else constitutionally aspect-blind, e.g., infant humans and various kinds of non-human animals. This in turn strongly suggests that the capacity for seeing or visual perception in normal adult humans is in fact made up of two distinct sorts of sub-abilities:

(i) essentially nonconceptual visual abilities

and

(ii) conceptual visual abilities,

the former of which is basic for the latter, and also shared with non-human animals.[102]

V.2.15 Experiencing the Meaning of a Word

We have just seen that later Wittgenstein's phenomenology of visual perception yields the doctrine of Dual Seeing as a four-membered set of substantive results in Linguistic Phenomenology:

(i) There are two irreducibly different types of seeing: *direct seeing* ("seeing-this") and *interpretive seeing* ("seeing-as").

(ii) Interpretive seeing requires direct seeing.

(iii) Interpretive seeing requires high-grade cognitive (and in particular, conceptual) capacities, whereas direct seeing does not.

and

(iv) Direct seeing can occur in the absence of interpretive seeing.

In the final phase of the *Investigations*, later Wittgenstein extends Dual Seeing to the linguistic phenomenology of *experiencing the meaning of a word*. This fascinating investigation then completes the whole argument of the *Investigations* by returning us full-circle, all the way around the globe, to the meaning-is-use thesis and by demonstrating some further positive theses about the concept of use. Again for convenience's sake, I will call this doctrine *Dual Semantic Phenomenology*. And in turn, here is a reconstruction of Dual Semantic Phenomenology.

V.2.15.1 Dual Semantic Phenomenology Reconstructed

(1) Just as there is a linguistic phenomenology of seeing, so too there is an analogously structured linguistic phenomenology of experiencing the meaning of a word.

> The importance of this concept [of aspect blindness] lies in the connection between the concepts of "seeing an aspect" and "experiencing the meaning of a word." For we want to ask "What would you be missing if you did not *experience* the meaning of a word?" What would you be missing if you did not understand the request to pronounce the word 'till' and to mean it as a verb, —or if you did not feel that a word lost its meaning and became a mere sound if it was repeated ten times over? (*PI* 214ᵉ) [Cf. claim (iv) of Dual Seeing.]

> When I pronounce this word while reading with expression it is completely filled with meaning. –"How can this be, if meaning is the use of the word?" Well, what I said was intended figuratively. Not that I chose the figure: it forced itself on me. (*PI* 215ᵉ) [Cf. claim (i) of Dual Seeing.]

> "But what is this queer experience?" –Of course it is not queerer than any other; it simply differs in kind from those experiences which we regard as the most fundamental ones, our sense impressions for instance. (*PI* 215ᵉ) [Cf. claim (ii) of Dual Seeing.]

> You can say the word "March" to yourself and mean it at one time as an imperative at another as the name of a month. And now say "March!"—and then March *no further*!" — Does the *same* experience accompany the word both times—are you sure? (p. 215ᵉ) [Cf. claim (iii) of Dual Seeing.]

(2) But the phenomenology of experiencing the meaning of a word also *differs from* the phenomenology of seeing, in that words can be sometimes used "transparently," that is, *without* any special experience of meaning.

> If a sensitive ear shows me, when I am playing this game [with 'March'], that I have now *this* now *that* experience of the word—doesn't it also show me that often I do not have any experience of it in the course of talking? —For the fact that I also mean it, intend it,

now like *this* now like *that*, and maybe also say so later is, of course, not in question. (*PI* 215ᵉ-216ᵉ)

(3) Therefore *the meaning of a word*—its use, which necessarily involves both its function and its application—is not the same as *experiencing the meaning of a word.*

> Someone tells me: "Wait for me by the bank." Question: Did you, as you were saying the word, mean this bank? –This question is of the same kind as "Did you intend to say such-and-such to him on your way to meet him?" It refers to a definite time (the time of walking, as the former question refers to the time of speaking)—but not to an *experience* during that time. Meaning is as little an experience as intending. (*PI* 216ᵉ –217ᵉ)

> Meaning is not a process which accompanies a word. For no *process* could have the consequences of meaning. (*PI* 217ᵉ)

> "At that word we both thought of him." Let us assume that each of us said the same words to himself—and how can it mean MORE that that?—But wouldn't even those words contain only a *germ*? The must surely belong to a language and to a context, in order really to be the expression of the thought *of* that man. If God had looked into our minds he would not have been able to see there whom we were speaking of. (*PI* 217ᵉ)

(4) Nevertheless the speaker's possession of a *human capacity* for experiencing the meaning of a word is at least a necessary (although not sufficient) condition of using a word meaningfully.

> Experiencing a meaning and experiencing a mental image. "In both cases," we should like to say, "we are experiencing something, only something different. A different content is proffered—is present—to consciousness." –What is the content of imagining? The answer is a picture or a description. And what is the content of the experience of meaning? I don't know what I am supposed to say to this.—If there is any sense in the above remark, it is that the two concepts are related like "red" and "blue" [i.e., two determinates under the same determinable concept CONSCIOUS EXPERIENCE]; and that is wrong. (*PI* 175ᵉ-176ᵉ)

> The importance of this concept [of aspect blindness] lies in the connection between the concepts of "seeing an aspect" and "experiencing the meaning of a word." For we want to ask "What would you be missing if you did not *experience* the meaning of a word?" What would you be missing if you did not understand the request to pronounce the word 'till' and to mean it as a verb, —or if you did not feel that a word lost its meaning and became a mere sound if it was repeated ten times over? (*PI* 214ᵉ)

> "Talking" (whether out loud or silently) and "thinking" are not concepts of the same kind; even though they are in closest connection. (*PI* 217ᵉ)

V.2.16 Later Wittgenstein's Critique of Philosophical Analysis

We are now finally in a position to return to the overarching main theme of this section, which is how later Wittgenstein's critical transformations of philosophical analysis in in the *Investigations* fundamentally re-connects his philosophy to Kant's

Transcendental Idealism in particular and to transcendental philosophy more generally.

This Kantian and transcendentalist re-connection occurs by means of later Wittgenstein's

philosophically liberating proposal *that logic is not anything "sublime" but is really

nothing but grammar*. I will call this *The Logic-is-Grammar Thesis*, and reconstruct his

argument for it in the following way.

V.2.16.1 Later Wittgenstein's Argument for The Logic-is-Grammar Thesis Reconstructed

(1) Frege, Moore, Russell, Carnap, the members of the Vienna Circle, and all other philosophers within the mainstream Analytic tradition after them explicitly or implicitly hold the thesis that logic is something "sublime": universal, a priori, necessary, and *noumenally essential.*

> These considerations bring us up to the problem: In what sense is logic sublime? For there seemed to pertain to logic a peculiar depth—a universal signficance. Logic lay, it seemed, at the bottom of all the sciences.— For logical investigation explores the nature of all things. <u>It seeks to see to the bottom of things and is not meant to concern itself whether what actually happens is this or that.</u>— It takes its rise, not from an interest in the facts of nature, nor from a need to grasp causal connections: <u>but from an urge to understand the basis, or essence, of everything empirical.</u> (*PI* §89, 42^e, underlining added)

(2) Furthermore, logic is required to carry out a complete decompositional analysis of our forms of language, propositions, and thoughts, which reveals their "hidden" "simple" structures and constituents, that is, their decomposable essences (*PI* §§ 91-92).

(3) This in turn implies that language, propositions, thought, and the world all *possess* decomposable noumenal essences (*PI* §§ 93-96).

(4) But in fact

> (i) every sentence in our language is in order just as it is,

> (ii) vagueness (via the pervasive family resemblance nature of all concepts) is a *constitutive* feature of meaning,

> (iii) language is essentially a spatiotemporal phenomenon, not something abstract,

and

> (iv) the essence of language, proposition, thought, and the world is something that "already lies open to view and that becomes surveyable by a rearrangement" (*PI* §§ 92, 98-100, 108-109).

(5) So neither language, nor propositions, nor thought, nor the world have hidden decomposable noumenal essences, and therefore the thesis that logic is sublime is false.

(6) Furthermore the thesis that logic is sublime turns out to be only a methodological assumption we have unintentionally imposed upon the phenomena, indeed nothing but an artifact of an idealized metaphysical "picture" that lay hidden in our language and held us captive (*PI* §§101-108, and 110-115).

(7) On the contrary, however, "the philosophy of logic speaks of sentences and words in exactly the sense in which we speak of them in ordinary life when we say, e.g., 'Here is a Chinese sentence,' or 'No, that only looks like writing; it is actually an ornament' and so on.'" (*PI* §108). That is: we can regard logic as purely descriptive or re-descriptive, not noumenally essential; and "what *we* do is to bring words back from their metaphysical use to their everyday use" by asking "is the word ever actually used in this way in the language which is its original home" (*PI* §116).

(8) Therefore we should adopt the thesis that logic is really nothing but "grammar," which "sheds light on our problem by clearing misunderstandings away…. misunderstandings concerning the use of words, caused, among other things by certain analogies between the forms of expression in different regions of language …. [and] some of them can be removed by substituting one form of expression for another; this may be called an 'analysis' of our forms of expression, for the process is sometimes like one of taking things apart" (*PI* § 90). For an example of this, see the discussion of negation at *PI* §§547-557.

(9) Furthermore, the goal of logic-as-grammar is to produce a "perspicuous representation" of language, propositions, thought, and the world, which produces "that understanding which consists in 'seeing connections'" (*PI* §122).

(10) So logic is not sublime, and logical-decompositional analysis is impossible, but logic-as-grammar is possible, and grammar in this sense is the descriptive logic of our language games, as transcendentally-anthropocentrically embedded in our communal forms of life. And to the extent that logic as a theory of valid reasoning still exists in the form of logic-as-grammar, this logic is *fully transcendental in the Kantian and also Tractarian sense*.

If later Wittgenstein's argument for The Logic-is-Grammar Thesis is sound, then philosophical analysis in the tradition that runs from Frege to Moore to Russell to Carnap to the Vienna Circle and beyond, all the way to Kripke, David Lewis, Kit Fine, David Chalmers, Frank Jackson, Timothy Williamson, John Hawthorne, Ted Sider, etc., and the contemporary movement of *Analytic Metaphysics* more generally, is *impossible*, precisely because it is fully committed to the false and rationally self-alienating thesis that logic is sublime or noumenally essential. Or in other words, if later Wittgenstein's

argument for The Logic-is-Grammar Thesis is sound, then *Analytic philosophy as we know it is impossible.*

Now suppose that this radical line of critical meta-philosophical reasoning is correct. What, according to later Wittgenstein, does philosophy become after the collapse of philosopical analysis, i.e., after the collapse of Analytic philosophy as we know it? The answer to this question has two parts.

First, later Wittgenstein's conception of philosophy in fact shares some fundamental features in common with his activist conception of logical analyis in the *Tractatus.* But this activist conception of logical analysis is now *minus the "sublimity" or "noumenal essentialism" of logic*, that is to say, minus the comprehensive *noumenal essentialist* metaphysical picture of logic, language, thought, and the world that would justify The Logical-Decompositional Theory of analysis, *but still accepting the transcendental character of logic, now understood to be logic-as-grammar.* Logic is *not* sublime, but logic *is* transcendental for early Wittgenstein in the *Tractatus* and for later Wittgenstein in the *Investigations* alike. Here are some relevant texts describing this radical turn in philosophy:

> The results of philosophy are the uncovering of one or another piece of plain nonsense and of bumps that the understanding has got by running its head up against the limits of language. (*PI* §119, 48ᵉ)
>
> A philosophical problem has the form: 'I don't know my way about'. (*PI* §123, 49ᵉ)
>
> Philosophy may in no way interfere with the actual use of language; it can in the end only describe it. For it cannot give it any foundation either. It leaves everything as it is. (*PI* §124, 49ᵉ)
>
> The civil status of a contradiction, or its status in civil life: there is the philosophical problem. (*PI* §125, 50ᵉ)
>
> Philosophy simply puts everything before us, and neither explains nor deduces anything. —S ince everything lies open to view, there is nothing to explain. For what is hidden, for example, is of no interest to us. One might give the name 'philosophy' to what is possible *before* all new discoveries and inventions. (*PI* §126, 50ᵉ)

The work of the philosopher consists in assembling reminders for a particular purpose. (*PI* §127, 50ᵉ)

If one tried to advance *theses* in philosophy, it would never be possible to question them, because everyone would agree to them. (*PI* §128, 50ᵉ)

It is not our aim to refine or complete the system of rules for the use of words in unheard-of ways. For the clarity that we are aiming at is indeed *complete* clarity. But that simply means that the philosophical problems should *completely* disappear. The real discovery is one that makes me capable of stopping doing philosophy when I want to. –The one that gives philosophy peace, so that it is no longer tormented by questions which bring *itself* into question…. There is not *a* philosophical method, though there are indeed methods, like different therapies. (*PI* §133, 51ᵉ)

In philosophy we do not draw conclusions. 'But it must be like this!' is not a philosophical proposition. Philosophy only states what everyone admits. (*PI* §599, 156ᵉ)

In short, the later Wittgenstein's radical *turn* in philosophy towards logic-as-grammar is simply a radical *re-turn* to Kant's critical meta-philosophy, that is, a radical return to transcendental logic understood as *transcendental dialectic*, whereby, as in Kant's sense, "transcendental dialectic" means *the non-classical dialetheic paraconsistnt logic-guided meta-philosophical critique of metaphysical illusion in philosophy, as a form of rational self-knowledge* (*CPR* A61-62/B85-86, A293-298/B349-354). The main idea is that by explicitly or implicitly deploying a dialetheic paraconsistent logic, the logical grammarian

(i) displays and diagnoses the dialectical structure of philosophical problems, i.e., displays and diagnoses "[t]he civil status of a contradiction, or its status in civil life" (*PI* §125, 50ᵉ),

(ii) describes, unpacks, compares, and contrasts the concepts implicit in our various ordinary uses of language and states a priori truisms about them,

and then

(iii) *stops doing philosophy when he wants to*, in order to make *changing one's life for the better* possible, and in order to achieve "insight into what lies in front of everyone's eyes."

Second, and as a direct consequence of this, the other crucial thing about later Wittgenstein's conception of philosophy is that it is *fundamentally non-cognitive*, that is, fundamentally desire-based, emotive, normative, practical, and moral. On this view,

philosophy is neither a natural science nor indeed in any sense a mere source of factual

knowledge but rather essentially a self-conscious and deliberate act—the act of "doing

philosophy"—whose final aim is achieving perspicuous insight into what already is

completely there already in front of us: i.e., human persons or rational human animals

and their linguistic activities in their rational human manifest natural world, i.e., *forms of*

rational human life:

> So you are saying that human agreement decides what is true and what is false?—It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in form of life. (*PI* § 241)

This linguistic agreement in forms of life, in turn, is given essentially in the activity of

making judgments:

> If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments. This seems to abolish logic, but does not do so. (*PI* § 242)

So in other words: our linguistic agreement in form of life *consists in our shared capacity*

for logical and practical reasoning, as human persons or rational human animals living

together in their shared world. In the light of this, we can now also say that the aim of

philosophy for the later Wittgenstein is precisely to achieve insight into what lies before

everyone's eyes, i.e., into *us and our manifest natural world*.

Here is the upshot of this sub-section. In *Philosophical Investigations*, later

Wittgenstein transforms philosophy from the reductive method of classical analytic

decomposition and logical reconstruction into the non-reductive, logically-guided study

of rational human animals inherently constrained in their individual intentional actions

and social practices by self-legislated and communally-constituted normative rules of

judgment and language-use. Here the logical theory, or logic-as-grammar, that guides the

later Wittgensteinian philosophy is not a classical logic but instead a strongly non-

classical dialeteic paraconsistent logic allowing for multiple conclusions, a denial of the principle of excluded middle, a denial of two-valuedness or bivalence, "true contradictions" or paradoxes, vagueness or borderline cases, irreducibly normative inferences, irreducible intensionality, and various irreducible intentional propositional attitudes. This does *not* mean that anything goes: logic-as-grammar is still strictly normatively guided by some conception or another of logical consequence; and not every proposition is both true and false.[103]

Even more importantly, later Wittgenstein's strongly non-classical dialetheic paraconsistent logic bears a fundamental affinity with and similarity to Kant's *transcendental logic*, which includes both *transcendental analytic* (the logic of truth) and *transcendental dialectic* (the dialetheic paraconsistent logic of illusion). The only salient difference between Kant and the later Wittgenstein is that Wittgenstein's logic-as-grammar explicitly incorporates the total range of facts encompassing human *linguistic competence and linguistic performance* within its scope, whereas this basic concern with language is at best implicit for Kant.[104] But the crucial point is that the later Wittgenstein's conception of philosophy has a fundamental affinity with and similarity to Kant's critical meta-philosophy as it is worked out in the *Critique of Pure Reason*. There, as we have seen, Kant tells us that philosophy is the study of rational human cognition. Rational human cognition, in turn, is cognition from principles, which are the fundamental normative necessary a priori laws of scientific knowing, thought, volition, action, and feeling. Thus philosophy for Kant and for Wittgenstein, both in the *Tractatus* and also in the *Investigations*, is *rational anthropology*: the study of rational human animals in their manifest natural world, insofar as their scientific knowing, thought,

volition, action, and feeling are inherently governed and evaluable by categorically normative principles. This does not of course imply that rational human animals ever actually manage to conform perfectly or even terribly adequately to these principles. Rationality is the *recognition* of principles and the *capacity* to conform to them freely, and does not itself entail that we ever actually (fully) *conform* to them. Indeed, only a rational animal who recognizes principles and and has the capacity to conform to them freely would ever be capable of, or even remotely interested in, trying to *rationalize* his way out of his responsibility for actually failing to match up to the principles that strictly obligate him. Rational human animals are the animals uniquely capable of *confusion, rationally self-alienating and self-annihilating contradictions, moral evil*, and *suffering*. This Kantian, early Wittgensteinian, later Wittgensteinian, and more generally *transcendental* philosophical recognition of these actual "human, all too human" facts on the ground fully captures a tragic sense of rational human life in a desperately nonideal world.

VI. Conclusion

As I previewed it in section I, my overall conclusion is that transcendental philosophy, as jointly done by Kant and Wittgenstein, not only constitutes a defensible, distinctive, and serious *alternative* to the other basic kinds of classical modern philosophy—i.e., Rationalism, Empiricism, and the more extreme forms of Idealism and/or Anti-Realism—but also provides a defensible, distinctive, and serious *successor-discipline* to Analytic philosophy. What, more precisely, do I mean by this?

As Quine, Hans Reichenbach, and Wilfrid Sellars so clearly saw in the 1950s, mainstream Analytic philosophy after the successive downfalls of Logicism and Logical Empiricism or Logical Positivism in the first half of the 20th century is nothing more and nothing less than a series of minor variations on the overarching theme of *scientific philosophy*:

> In the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not.[105]

But Kant and Wittgenstein systematically challenge and reject this essentially *scientistic* conception of philosophy:

> I cannot even assume God, freedom, or immortality for the sake of the necssary practical use of my reason unless I simultaneously deprive speculative reason of its pretensions to extravagant insights; because in order to attain to to such insights, speculative reason would have to help itself to principles that in fact reach only to objects of possible experience, and which, if they were to be applied to what cannot be an object of possible experience, then they would always transform it into an appearance and thus declare all practical extension of pure reason to be impossible. Thus I had to deny scientific knowing (*Wissen*) in order to make room for faith (*Glauben*). (*CPR* Bxxix-xxx, underlining added).
>
> The truly apocalyptic view of the world is that things do not repeat themselves. It isn't absurd, e.g., to believe that the age of science and technology is the beginning of the end for humanity; the idea of great progress is a delusion, along with the idea that the truth will ultimately be known; that there is nothing good or desirable about scientific knowledge and that mankind, in seeking it, is falling into a trap. It is by no means obvious that this is not how things are. (*CV*: 56ᵉ, underlining added)
>
> It was true to say that our considerations could not be scientific (*wissenschaftliche*) ones. It was not of any possible interest to us to to find out empirically 'that, contrary to our preconceived

ideas, it is possible think such-and-such'—whatever that may mean… <u>And we may not advance any kind of [scientific] theory</u>…. We must do away with all *explanation*, and description alone must take its place. <u>These are, of course, not empirical problems</u>; they are solved, rather, by looking into the workings of our language, and that in such a way as to make us recognize those workings: *in spite of* an urge to misunderstand them. The problems are solved, not by giving new information, but by arranging what we have always known. <u>Philosophy is a battle against the bewitchment of our intelligence by means of language.</u> (*PI* §109, 47ᵉ, underlining added)

Otherwise put: Analytic philosophy began with Logicism, fell into self-alienation and self-annihilation with the successive implosions of Logicism and Logical Empiricism or Logical Positivism, and ended in Scientism. But if Kant and Wittgenstein are correct, then natural science is *not* the measure of all things, for the following three reasons:

(i) Logic is the measure of natural science.

(ii) "Logic is transcendental," insofar as it is *pure general logic*.

and

(iii) Logic is also "grammar, "insofar as it is the dialetheic paraconsistent meta-philosophical logic of *transcendental dialectic*.

In this way, transcendental philosophy does not *either* seek a humanly impossible scientific, practical, theological, or philosophical insight into things-in-themselves *or* draw destructive skeptical conclusions from our inevitable and tragic failure to achieve noumenal insight and noumenal grounding. For *each* of these projects, whether noumenal or skeptical, is equally, although oppositely, inherently self-alienating, "inauthentic," and in "bad faith" in the Existential sense. As philosophers with a tragic sense of life, then, we recognize that we must forego, give up, let go of, purify ourselves of, renounce every variety of *the felt need* for noumenal insight and noumenal grounding; and this is the deepest lesson of transcendental philosophy. Therefore, transcendental philosophy, as jointly done by Kant and Wittgenstein, is the right successor discipline to Analytic philosophy precisely because *rational "human, all too human" animals and their manifest natural world*, not natural science, is the measure of all things, *including* natural

science, and precisely because *only* the full recognition of this transcendental fact and all its life-changing theoretical, practical, and Existential implications, finally enables the trans-Analytic transcendental philosopher to have "insight into what lies in front of everyone's eyes."

NOTES

[1] For convenience I refer to Kant's works and to Wittgenstein's works internally, that is, infratextually in parentheses.

The citations to Kant include both an abbreviation of the English title and the corresponding volume and page numbers in the standard "Akademie" edition of Kant's works: *Kants gesammelte Schriften*, edited by the Königlich Preussischen (now Deutschen) Akademie der Wissenschaften (Berlin: G. Reimer [now de Gruyter], 1902-). For references to the first *Critique*, I follow the common practice of giving page numbers from the A (1781) and B (1787) German editions only. For references to Kant's *Reflexionen*, i.e., entries in *Kants handschriftlicher Nachlaß*—which I abbreviate as '*R*'—I give the entry number in addition to the Akademie volume and page numbers. The translations from the *Reflexionen* are my own.

The citations to Wittgenstein include an abbreviation of the English title, the relevant proposition or paragraph number(s), and/or the corresponding page number(s) in the English translation.

I generally follow the standard English translations from the German texts, but have occasionally modified them where appropriate. Here is a list of the abbreviations and English translations of the works cited in the internal references:

Kant

CPJ	*Critique of the Power of Judgment*. Trans. P. Guyer and E. Matthews. Cambridge: Cambridge Univ. Press, 2000.
CPR	*Critique of Pure Reason*. Trans. P. Guyer and A. Wood. Cambridge: Cambridge Univ. Press, 1997.
CPrR	*Critique of Practical Reason*. Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy.* Cambridge: Cambridge Univ. Press, 1996. Pp. 133-272.
GMM	*Groundwork of the Metaphysics of Morals*. Trans. M. Gregor. In *Immanuel Kant: Practical Philosophy.* Pp. 37-108.
JL	"The Jäsche Logic," in *Immanuel Kant: Lectures on Logic*. Trans. J.M. Young. Cambridge: Cambridge Univ. Press, 1992. Pp. 519-640.
P	*Prolegomena to Any Future Metaphysics*. Trans. J. Ellington. Indianapolis: Hackett, 1977.
PC	*Immanuel Kant: Philosophical Correspondence, 1759-99*. Trans. A. Zweig. Chicago: Univ. of Chicago Press, 1967.
Rel	*Religion within the Boundaries of Mere Reason*. Trans. A. Wood and G. Di Giovanni. Cambridge: Cambridge Univ. Press, 1998.
VL	"The Vienna Logic," in *Immanuel Kant: Lectures on Logic*. Pp. 249-377.
WiE	"An Answer to the Question: 'What is Enlightenment?'" Trans. H.B. Nisbet. In H. Reiss (ed.), *Kant's Political Writings*. Cambridge: Cambridge Univ. Press, 1970. Pp. 54-60.

Wittgenstein

CV	*Culture and Value.* Trans. P. Winch. Chicago, IL: Univ. of Chicago Press, 1980.
NB	*Notebooks 1914-1916*, trans. G.E.M. Anscombe (2nd edn., Oxford: Blackwell, 1979).
PI	*Philosophical Investigations*. Trans. G.E.M. Anscombe. New York: Macmillan, 1953.
TLP	*Tractatus Logico-Philosophicus*. Trans. C.K. Ogden. London: Routledge, 1922/1992.

[2] R. Monk, *Wittgenstein: The Duty of Genius* (London: Jonathan Cape, 1990), pp. 136 and 158.

[3] See, e.g., M. de Unamuno, *The Tragic Sense of Life*, available online at URL = http://www.gutenberg.org/files/14636/14636-h/14636-h.htm; and also S. Cavell, *The Claim of Reason: Wittgenstein, Skepticism, Morality, and Tragedy* (Oxford: Oxford Univ. Press, 1979), esp. part 1 and part 4.

[4] See also R. Hanna, *The Rational Human Condition* (Unpublished MS, Summer 2011 version).

[5] See, e.g., URL = <http://en.wikipedia.org/wiki/Pikes_Peak>. I visited the summit of Pike's Peak in summer 2010, and confirmed this claim by direct perception.

[6] See R. Hanna, "The Inner and the Outer: Kant's 'Refutation' Reconstructed," *Ratio* 13 (2000): 146-174; and Hanna, *Kant, Science, and Human Nature*, ch. 1.

[7] See Monk, *Ludwig Wittgenstein: The Duty of Genius,* p. 206.

[8] See also R. Brockhaus, *Pulling Up the Ladder: The Metaphysical Roots of Wittgenstein's Tractatus Logico-Philosophicus* (Chiacgo, IL: Open Court, 1991).

[9] G.H. von Wright, "A Biographical Sketch," in N. Malcolm, *Ludwig Wittgenstein: A Memoir* (Oxford: Oxford Univ. Press, 1984), pp. 3-20, at p. 6.

[10] See also, e.g., P.M.S. Hacker, *Insight and Illusion: Wittgenstein on Philosophy and the Metaphysics of Experience* (Oxford: Oxford Univ. Press, 1972); J. Lear, "Leaving the World Alone," *Journal of Philosophy* 79 (1982): 382-403; and B. Williams, "Wittgenstein and Idealism," in B. Williams, *Moral Luck* (Cambridge: Cambridge Univ. Press, 1981), pp. 144-163. It should also be noted that Hacker changed his mind about the Kantian/Schopehnauerian influence on Wittgenstein in the second edition of *Insight and Illusion*, in my opinion mistakenly. I think that he was much closer to being correct the first time around.

[11] S. Kierkegaard, "Stages on Life's Way,"in *The Essential Kierkegaard*, trans. H. Hong and E. Hong (Princeton, NJ: Princeton Univ. Press, 1997), pp. 170-186, at p. 182 ; see also Kant (*Rel* 6: 57-202).

[12] F. Dostoyevsky, *The Brothers Karamazov*, 2 vols., trans. D. Magarshack (Harmondsworth, Middlesex: Penguin Books, 1958), vol. 2, p. 743.

[13] See, e.g., H. Barnes, *Humanistic Existentialism* (Lincoln, NE: Univ. of Nebraska Press, 1959).

[14] See R. Hanna, *Kant and the Foundations of Analytic Philosophy* (Oxford: Clarendon/Oxford Univ. Press, 2001), ch. 2, esp. section 2.2.

[15] See L. Stevenson, "Opinion, Belief or Faith, and Knowledge," *Kantian Review* 7 (2003): 72-101, at p. 85.

[16] R. Descartes, "Principles of Philosophy," in R. Descartes, *The Philosophical Writings of Descartes*, 3 vols., trans. J. Cottingham et al. (Cambridge: Cambridge Univ. Press, 1985), vol. 1, pp. 177-291, article 205, (first sentence of the French edition), p. 289 (AT 327).

[17] R. Descartes, "Discourse on the Method of Rightly Conducting One's Reason and Seeking the Truth in the Sciences," in Descartes, *The Philosophical Writings of Descartes*, pp. 109-151, at p. 130 (AT 37-38).

[18] M. Kuehn, *Kant: A Biography* (Cambridge: Cambridge Univ. Press, 2001), p. 3.

[19] See, e.g., R.Hanna, "How Do We Know Necessary Truths? Kant's Answer," *European Journal of Philosophy* 6 (1998): 115-145.

[20] See, e.g., G. Oppy, "Ontological Arguments," *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2009/entries/ontological-arguments/, esp. sections 3 and 6-9.

[21] See K. Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in J. Van Heijenoort (ed.), *From Frege to Gödel* (Cambridge, MA: Harvard Univ. Press, 1967), pp. 596-617.

[22] See Oppy, "Ontological Arguments," section 6.

[23] See, e.g., K. Gödel, "What in Cantor's Continuum Problem?," in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings* (2nd edn., Cambridge: Cambridge Univ. Press, 1984), pp. 470-485.

[24] See, e.g., M. Tooley, "The Problem of Evil," *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/spr2010/entries/evil/>.

[25] Kierkegaard, "Fear and Trembling," in *The Essential Kierkegaard*, pp. 93-101.

[26] See, e.g., R. Iemhoff, "Intuitionism in the Philosophy of Mathematics," *The Stanford Encyclopedia of Philosophy (Winter 2009 Edition)*, E. N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2009/entries/intuitionism/>; and M. van Atten, "The Development of Intuitionistic Logic," *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, E. N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2009/entries/intuitionistic-logic-development/.

[27] As quoted in R. Brockhaus, *Pulling Up the Ladder: The Metaphysical Roots of Wittgenstein's Tractatus Logico-Philosophicus*, p. 296.

[28] See S. Kierkegaard, "Purity of Heart is to Will One Thing," *The Essential Kierkegaard*, trans. H. Hong and E. Hong (Princeton, NJ: Princeton Univ. Press, 1980), pp. 271-276.

[29] A very similar point is made by Christine Korsgaard in *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford Univ. Press, 2009), ch. 5, pp. 84-89. Oddly enough, however, she concludes her highly insightful discussion by saying that "my point in bringing all this up is not to make a brief for Kant's philosophy of religion, or for the need for agency to be supported by faith" (p. 89). *But why not*?

[30] Monk, *Wittgenstein: The Duty of Genius,* p. 158.

[31] Kierkegaard, "Purity of Heart is to Will One Thing," p. 271.

[32] I am grateful to Robert Pasnau for a set of extremely helpful critical comments on an earlier version of this section.

[33] For more on the notion of basic authoritative philosophical rational intuition, see R. Hanna, *Objectivity Regained: Benacerraf's Dilemma and Intuitions in Mathematics, Logic, Morality, and Philosophy* (Unpublished MS, March 2011 version), complete working draft available online at URL = http://www.colorado.edu/philosophy/paper_hanna_objectivity_regained_march11.pdf.

[34] See, e.g., C.I. Lewis, *Survey of Symbolic Logic* (Berkeley, CA: Univ. of California Press, 1918), pp. 1-2; and B. Russell, *Introduction to Mathematical Philosophy* (London: Routledge, 1993).

[35] Wittgenstein, *Tractatus Logico-Philosophicus*, prop. 6.3751, p. 181.

[36] L. Wittgenstein, "Some Remarks on Logical Form," *Proceedings of the Aristotelian Society*, suppl. vol. 9 (1929): 162-171.

[37] F. Waismann, *Wittgenstein and the Vienna Circle*, trans. J. Schulte and B. McGuinness (New York: Harper and Row, 1979), pp. 67-68.

[38] See also Hanna, *Kant, Science, and Human Nature*, ch. 6.

[39] I am counting non-Euclidean geometry as a conservative extension of Euclidean geometry, on the two-part ground that (i) the parallel postulate is logically independent of the basic Euclidean postulates and (ii) substituting either of the classical Riemannian or Lobachevskian alternatives for the parallel postulate does not entail the denial of any other Euclidean postulates.

[40] See, e.g., B. Russell, "Mathematical Logic as Based on the the Theory of Types," in B. Russell, *Logic and Knowledge* (New York: G.P. Putnam's Sons, 1971), pp. 59-102, at p. 63.

[41] See, e.g., M. Potter, *Reason's Nearest Kin* (Oxford: Clarendon/Oxford Univ. Press, 2000), ch. 5.

[42] See, e.g., B. Mates, *Elementary Logic* (2nd edn., New York: Oxford Univ. Press, 1972).

[43] A. Tarski, "The Semantic Conception of Truth and the Foundations of Semantics," p. 371.

[44] See also, e.g., C. Parsons, "Kant's Philosophy of Arithmetic," in C. Parsons, *Mathematics in Philosophy* (New York: Cornell Univ. Press, 1983), pp. 110-149, at p. 131; and S. Shapiro, "Induction and Indefinite Extensibility: The Gödel Sentence is True, But Did Someone Change the Subject?," *Mind* 107 (1998): 597-624, at p. 604.

[45] W.V.O. Quine, *Philosophy of Logic* (2nd edn., Cambridge, MA: Harvard Univ. Press, 1986), p. 64.

[46] W.V.O. Quine, "Carnap and Logical Truth," in W.V.O. Quine, *The Ways of Paradox and Other Essays* (2nd edn., Cambridge, MA: Harvard Univ. Press, 1975), pp. 107-132, at p. 111.

[47] See also N. Denyer, "Pure Second-Order Logic," *Notre Dame Journal of Formal Logic* 33 (1992): 220-224; and A. Paseau, "Pure Second-Order Logic with Second-Order Identity," *Notre Dame Journal of Formal Logic* 51 (2010): 351-360. Pure second-order logic is second-order monadic logic without any functional or first-order variables, i.e., with systematic insensitivity as to whether domains are empty or non-empty. In this respect it is formally very similar to Kant's pure general logic, although pure general logic does contain first-order variables ranging over comprehensions (*Umfangen*) of actual and possible individuals.

[48] See, e.g., G. Boolos and R. Jeffrey, *Computability and Logic* (3rd edn., Cambridge: Cambridge Univ. Press, 1989), chs. 10, 22, and 25, and esp. pp. 250-255.

[49] Quine, *Philosophy of Logic*, p. 81.

[50] See G. Priest, *In Contradiction*. (Dordrecht: Martinus Nijhoff, 1987); and G. Priest, "What is So Bad About Contradictions?," *Journal of Philosophy* (1998): 410-426.

[51] See, e.g., Tarski, "The Semantic Conception of Truth and the Foundations of Semantics."

[52] See, e.g., K. Gödel, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems," in Van Heijenoort (ed.), *From Frege to Gödel*, pp. 596-617.

[53] See, e.g., Skolem, "The Foundations of Elementary Arithmetic Established by Means of the Recursive Mode of Thought, Wthout the Use of Appperent Variables Ranging Over Infinite Domains."
[54] Quine, *Philosophy of Logic*, p. 82.

[55] Quine, "Two Dogmas of Empiricism," pp. 22-23.

[56] Quine, *Philosophy of Logic,* p. 81.

[57] See, e.g., N. Chomsky, *Knowledge of Language* (Westport, CT: Praeger, 1986); and R. Hanna, *Rationality and Logic* (Cambridge, MA: MIT Press, 2006).

[58] See, e.g., S. Haack, *Deviant Logic* (Cambridge: Cambridge Univ. Press, 1974); and G. Priest, *An Introduction to Non-Classical Logic* (Cambridge: Cambridge Univ. Press, 2001).

[59] See G. Priest, *In Contradiction.* (Dordrecht: Martinus Nijhoff, 1987); and G. Priest, "What is So Bad About Contradictions?," *Journal of Philosophy* (1998): 410-426.

[60] In *Rationality and Logic*, ch. 3—see esp. p. 45—I did not adequately recognize the crucial difference between The Minimal Meta-Logical Principle of Non-Contradiction on the one hand, and other weak principles of classical logic on the other. Only The Minimal Metalogical Principle of Non-Contradiction is obeyed by *every possible* non-classical logic, e.g., by dialetheic paraconsistent logics. The other weak principles of classical logic, by contrast, are undermined by logics that are either not truth-preserving or not consistency-preserving. Many thanks to Richard Grandy and Jeffrey Rowlands for pointing this out to me.

[61] H. Putnam, "There is At Least One A Priori Truth," in H. Putnam, *Realism and Reason: Philosophical Papers, Vol. 3* (Cambridge: Cambridge Univ. Press, 1983), pp. 98-114, at pp. 100-101.

[62] See Hanna, *Kant and the Foundations of Analytic Philosophy*, esp. section 5.2; and Hanna, *Kant, Science, and Human Nature*, esp. sections 3.3 and 7.1.

[63] H.M. Sheffer, "Review of *Principia Mathematica*, Volume I, second edition," *Isis* 8 (1926): 226-231, at p. 228.

[64] L. Carroll, "What the Tortoise Said to Achilles," *Mind* 4 (1895): 278-280.

[65] Quine, "Truth by Convention," in Quine, *The Ways of Paradox and Other Essays*, pp. 77-106, at p. 104.

[66] S. Haack, "The Justification of Deduction," *Mind* 85 (1976): 112-119.

[67] See Hanna, *Rationality and Logic*, ch. 3.

[68] O. O'Neill, "Vindicating Reason," in P. Guyer (ed.), *The Cambridge Companion to Kant* (Cambridge: Cambridge Univ. Press, 1992), pp. 280-308, at p. 305.

[69] This is not, however, to say that pure general logic is a "transcendental logic" in Kant's technical sense of that term.

[70] See C.I. Lewis, "A Pragmatic Conception of the A Priori," *Journal of Philosophy* 20 (1923): 169-177; C.I. Lewis, *Mind and the World Order* (New York: Dover, 1956); and W.V.O. Quine, "Two Dogmas of Empiricism," in W.V.O. Quine, *From a Logical Point of View* (2nd edn., New York: Harper and Row, 1961), pp. 2-46. Quine is a holist about confirmation and disconfirmation, whereas Lewis was not. But otherwise it is obvious that Quine's conception of the a priori is essentially the same as Lewis's pragmatic conception. Nor is this fact at all surprising, since (i) Lewis was Quine's teacher at Harvard, and (ii) Lewis was the leading philosopher in the Harvard Department of Philosophy during Quine's graduate student days. See B. Kuklick, *The Rise of American Philosophy* (New Haven, CT: Yale Univ. Press, 1977).

[71] See, e.g., J. Kim, *Supervenience and Mind* (Cambridge: Cambridge Univ. Press, 1993), esp. part 1; D. Chalmers, *The Conscious Mind* (New York: Oxford Univ. Press, 1996), chs. 2-3; and T. Horgan, "From Supervenience to Superdupervenience: Meeting the Demands of a Material World," *Mind* 102 (1993): 555-586.

[72] See Hanna, *Rationality and Logic*, esp. sections 6.5-6.6; and Hanna, *Objectivity Regained*, esp. sections I-III, and VI-IX.

[73] See Hanna, *Objectivity Regained*, esp. sections IV, VII-VIII, X-XI, and XIII.

[74] R. Monk, *Bertrand Russell: The Spirit of Solitude* (London: Jonathan Cape, 1996), p. 153.

[75] Hume, *Enquiry Concerning Human Understanding*, section XII, part III, p. 165.

[76] See Hanna, *Objectivity Regained*, esp. sections IV and VII.

[77] M.N. Forster, *Kant and Skepticism* (Princeton, NJ: Princeton Univ. Press, 2008), p. 45, underlining added.

[78] S. Jeffries, "A Dwelling for the Gods," *The Guardian* [5 January 2002], underlining added.

[79] Non-analytic or synthetic propositions can, of course, *also* be analyzed. But the goal of a specifically *philosophical* analysis is a priori knowledge of analytically (logically) necessary truths.

[80] See Hanna, "How Do We Know Necessary Truths? Kant's Answer."

[81] See also, e.g., Hacker, *Wittgenstein's Place in Twentieth-Century Analytic Philosophy*, ch. 5.

[82] See note 10 above.

[83] See Austin, *How to Do Things with Words*; and Searle, *Speech Acts*.

[84] My view is that in the *Tractatus*, the atomic propositions all have sense, and then the molecular propositions *inherit* sense via the truth-functional operations that construct molecular propositions. It is possible, however, that *only* the atomic propositions have sense, and that all the molecular propositions are supposed by Wittgenstein to be senseless. Many thanks to Michael Potter for many conversations about this issue.

[85] See D. Kaplan, "Demonstratives: An Essay on the Semantics, Logic, and Epistemology of Demonstratives and Other Indexicals," in J. Almog et al. (eds.), *Themes from Kaplan* (New York: Oxford Univ. Press, 1989), pp. 481-614; G. Evans, *Varieties of Reference* (Oxford: Clarendon/Oxford Univ. Press, 1982); J. Perry, "The Problem of the Essential Indexical," *Noûs* 13 (1979): 3-21; and R. Hanna, "Direct Reference, Direct Perception, and the Cognitive Theory of Demonstratives," *Pacific Philosophical Quarterly* 74 (1993): 96-117.

[86] See S. Kripke, *Wittgenstein on Rules and Private Language* (Cambridge: Harvard Univ. Press, 1982); and Hanna, *Rationality and Logic*, ch. 6.

[87] See note 85 above.

[88] See, e.g., Hanna, *The Rational Human Condition*, part 2, esp. chs. 2.2-24.

[89] B. Russell, "On the Nature of Truth and Falsehood," in B. Russell, *Philosophical Essays* (London: Allen & Unwin, 1966), pp. 147-159, at 153.

[90] B. Russell, *The Problems of Philosophy* (Oxford: Oxford Univ. Press, 1912), p. 46.

[91] Russell, *The Problems of Philosophy*, pp. 48-52.

[92] L. Carroll, *Through the Looking-Glass* (New York: Dial, 1988), pp. 105-106.

[93] Augustine, *Confessions*, trans. R. Pine-Coffin (Harmondsworth, Middlesex: Penguin, 1961), X, 3, p. 208.

[94] See, e.g., P. Hacker, *Insight and Illusion* (2nd edn., Oxford: Clarendon/Oxford Univ. Press, 1986), ch. 8; Kripke, *Wittgenstein on Rules and Private Language*; and D. Pears, *The False Prison*, 2 vols. (Oxford: Clarendon/Oxford Univ. Press, 1987), vol. 2, chs. 13-15.

[95] See Carroll, *Through the Looking-Glass*, pp. 65-71.

[96] Kripke also explicitly and relevantly compares and contrasts his own interpretation of Wittgenstein's Rule Following Paradox—i.e., Kripkenstein's Rule Following Paradox—with Quine's famous "indeterminacy of translation" and "inscrutability of reference" arguments about the nature of meaning in *Word and Object*, and Goodman's equally famous "grue" paradox about the nature of induction in "The New Riddle of Induction."

[97] See Cavell, *The Claim of Reason*; T. Clarke, "The Legacy of Skepticism," *Journal of Philosophy* 69 (1972): 754-769; and B. Stroud, *The Significance of Philosophical Skepticism* (Oxford: Oxford Univ. Press, 1984).

[98] See Kim, *Philosophy of Mind*, ch. 2. What I am calling "metaphysical behaviorism" Kim calls "ontological behaviorism," and what I am calling "semantic behaviorism" Kim calls "logical behaviorism."

[99] Putnam, "Brains and Behavior."

[100] See Chalmers, *The Conscious Mind*, chs. 1-4.

[101] See also R. Hanna and M. Maiese, *Embodied Minds in Action* (Oxford: Oxford Univ. Press, 2009), ch. 5.

[102] See also R. Hanna, "Kantian Non-Conceptualism," *Philosophical Studies* 137 (2008): 41-64; Hanna, *Kant, Science, and Human Nature*, ch. 2; R. Hanna and M. Chadha, "Non-Conceptualism and the Problem of Perceptual Self-Knowledge," *European Journal of Philosophy*, pre-published URL = <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0378.2009.00377.x/abstract>, and forthcoming in print in vol. 19, 2 (June 2011); R. Hanna, "The Myth of the Given and the Grip of the Given," DIAMETROS 27 (March 2011), available online at URL = <http://www.diametros.iphils.uj.edu.pl/?l=2&p=anr25&m=25&if=0&ii=29&ik=27>; and R. Hanna, "Beyond the Myth of the Myth: A Kantian Theory of Non-Conceptual Content," forthcoming in *International Journal of Philosophical Studies*.

[103] See Putnam, "There is at Least One A Priori Truth."

[104] Kant has very little to say explicitly about the nature of language except for one provocative remark in the *Anthropology* (*A* 7: 192), where he seems to endorse the theses (1) that linguistic meanings are thoughts or parts of thoughts, and (2) that thinking is inner speech. This of course is similar to Wittgenstein's theory of language in the *Tractatus*

[105] W. Sellars, "Empiricism and the Philosophy of Mind," in W. Sellars, *Science, Perception, and Reality* (New York: Humanities Press, 1963), 127-196, at p. 173. See also H. Rechenbach, *The Rise of Scientific Philosophy* (Berkeley: Univ. of California Press, 1951).