# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**

> 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.
>
> --Immanuel Kant (*JL* 9: 24-25)

## *Course description:*

Kant's *Critique of Pure Reason* (CPR) is arguably the single most brilliant, important, and difficult book in modern philosophy. Its main topic is the nature, scope, and limits of human cognition and reason; and its main conclusion is that necessary truth, a priori knowledge, and freedom of the will in a deterministic natural world are possible if and only if transcendental idealism is true. The purpose of this course is to give a close, critical reading of the central line of argument in the CPR all the way from the Preface to the Ideal of Pure Reason.

***Text:*** Kant, I., *Critique of Pure Reason*, trans. P. Guyer and A. Wood (Cambridge: Cambridge Univ. Press, 1998).

## *Outline of lecture topics:*

LECTURE 1: Kant's Critical Project & Kant's Transcendental Project
LECTURE 2: The Introduction & Beyond: Basic Terms, Notions, & Distinctions
LECTURE 3: Space, Time, & Mathematics: The Transcendental Aesthetic
LECTURE 4: Transcendental Idealism
LECTURE 5: The Refutation of Idealism
LECTURE 6: Concepts, Logic, & Judgment: The Metaphysical Deduction of the Categories
LECTURE 7: The Transcendental Deduction of the Categories in the A Edition
LECTURE 8: The Transcendental Deduction of the Categories in the B Edition
LECTURE 9: The System of Principles I: Schematism, Axioms of Intuition, & Anticipations of Perception
LECTURE 10: The System of Principles II: Analogies of Experience
LECTURE 11: Transcendental Dialectic & Transcendental Ideas
LECTURE 12: The Third Antinomy, Freedom, & Determinism
LECTURE 13: The Ideal of Pure Reason, the Impossibility of Ontological Arguments, & How to Deal with the Unprovability of God's Existence (or Non-Existence)

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**

## LECTURE 1

Covers: CPR: 95-124 = Ai-xxiii/Bi-xliv (Preface, both editions).

### Kant's Critical Project and Kant's Transcendental Project

*(1.1) Why the CPR is arguably the most brilliant and important book in modern philosophy*.

Kant was born in 1724 and died in 1804. So his life spanned most of the 18th century. Kant's CPR (A edition = 1781, B edition = 1787) is arguably the most brilliant and important book in modern philosophy. Why do I think that? There are three answers.

**First**, the CPR offers a radically original resolution of the basic epistemological, semantic, & ontological problems of Rationalism and Empiricism (how is a priori knowledge possible? what kinds of truths are there? how do they have meaning? how do we cognize them? how do we know them? what really exists & what is its nature?).

**Second,** the CPR offers a radically original resolution of the basic metaphysical problems of Rationalism and Empiricism (what is God's nature, & does God exist? how can we be free in a deterministic natural world? what is the nature of the mind & can it exist independently of the body?, etc.).

**Third**, K's CPR in particular & his Critical Philosophy more generally have had a greater impact on the modern tradition than any other single book or philosophical theory.

I'll offer an argument for all of these claims, starting with the third one.

*(1.2) The place of the CPR in the history of modern philosophy*.

Let classical Rationalism (esp. Descartes & Leibniz) or **CR** be the thesis that : all fully meaningful cognition & knowledge begins in & is derived from (strictly determined by) reason, independently of sense experience

Let classical Empiricism (esp. Locke & Hume) or **CE** be the thesis that: all fully meaningful cognition & knowledge begins in & is derived from (strictly determined by) sense experience, independently of reason.

Then we can easily see that K's CPR in particular & the Critical Philosophy more generally are in fact the central nodes on the following history of modern philosophy timeline & line-of-influence:

**History of Modern Philosophy Timeline & Line-of-Influence 1600-2011**

**Classical Rationalism (Desc, Leibniz) {early 17th c. to early 18th}**
***vs***
**Classical Empiricism (Locke, Hume) {mid-17th c. to mid-18th }**

↓ ↓ ↓

**Kant's Pre-Critical period = 1746-1770: early commitment to Leibnizian/Wolffian philosophy → Three wake-up calls between 1766 & 1770:**

1. **"Dreams of a Ghost Seer" (1766) & Kant's 1798 letter to Garve: "the Antinomies of Pure Reason awoke me from my dogmatic slumber"**
2. **"the year 69 gave me great light" = space and time necessarily conform to our sensibility (perceptual capacities), and are not things-in-themselves**
3. **1770 or 1771: "remembering David Hume awoke me from my dogmatic slumber"**

↓ ↓ ↓

**Kant's Critical period = 1770-1787**

**ID + CPR + {PAFM+ GMM+ MFNS}**
**1770 1781/1787 1783 1785 1786**

**ID: *Inaugural Dissertation* (1770) → letter to Marcus Herz (1772)**
**CPR: *Critique of Pure Reason* (1781/1787)**
**PAFM: *Prolegomena to Any Future Metaphysics* (1783)**
**GMM: *Groundwork of the Metaphysics of Morals* (1785)**
**MFNS: *Metaphysics Foundations of Nature Science* (1786)**

↓ ↓ ↓

**Kant's Post-Critical period = 1788-1800**

**CPrR + CPJ + REL + MM + A + L + OP**
**1788 1790 1793 1797 1798 1800 1790s**

**CrPR: *Critique of Practical Reason* (1788)**
**CPJ: *Critique of the Power of Judgment* (1790)**
**REL: *Religion Within the Boundaries of Mere Reason* (1793)**

**MM:** ***Metaphysics of Morals*** **(1797)**
**A:** ***Anthropology from a Pragmatic Point of View*** **(1798)**
**L:** ***Logic*** **(1800)**
**OP:** ***Opus postumum*** **=** ***Transition from the MFNS to Physics*** **(late 1790s→ unfinished at Kant's death in 1804)**

↓ ↓ ↓

**Absolute Idealism (Fichte, Schelling, Hegel) {late 18th to mid-19th}**

↓ ↓ ↓

**Neo-Kantianism {mid-19th to early 20th}**

| | |
|---|---|
| **(<u>Trendelenburg</u>, Cohen, Natorp,** | **<u>Fischer</u>, Rickert, etc.)** |
| **Marburg** | **Heidelberg** |
| **Neo-Hegelianism**<br>**(Bradley, Bosanquet, Royce, etc.) {late 19th}** | **Early Phenomenology**<br>**(Brentano, Meinong, etc.)**<br>**{late 19th/early 20th}** |
| ↓ ↓ | ↓ |
| **Analytic Philosophy (20th/21st)** | **Phenomenology (20th)** |
| **(Frege, Russell, Moore, Wittgenstein, Carnap, Quine, Putnam/Kripke, etc.)** | **(Husserl, Heidegger, Sartre, Merleau-Ponty, etc.)** |

*(1.3) K's two projects in CPR*

'CPR' was not the original title of CPR; in fact CPR was originally called *The Limits of Sense & Reason*, as K tells us in his famous letter to his student Marcus Herz of 21 Feb 1772. This letter is extremely important for other reasons as well, so I'll come back to it later in this lecture, & also quote the relevant passages.

But for the moment the crucial point is that the 2 titles nicely pick out two distinct but also intimately-related philosophical projects that K is working on, which I'll name & correlate with their relevant titles as follows:

*The Critical Project* → *The Critique of Pure Reason*
*The Transcendental Project* → *The Limits of Sense & Reason*

This gets us back to my first two claims about CPR.

The Critical Project is how K offers a radically original resolution of the basic metaphysical problems of classical Rationalism, especially the problems of free will & natural determinism, & the existence & nature of God, by arguing that we must critically restrict the scope of scientific, theoretical, & pure rational knowledge, & then radically reconfigure our conception of cognition & knowledge to fit the human standpoint, in order to make room for free will & morality (so this captures the "Critique of Pure Reason" part).

And The Transcendental Project is how K offers a radically original resolution of the basic epistemological & semantic problems of Rationalism and Empiricism by showing how each is equally mistaken virtue of its basic commitments to certain false assumptions & theses, yet once we've radically reconfigured our conception of cognition & knowledge so that it fits the human standpoint, then we can adequately recover defensible elements of Rationalism & Empiricism alike, & also make progress beyond both (so this captures the "Limits of Sense & Reason" part).

In this lecture course, I will focus almost exclusively on K's Transcendental Project, except for the material from the Transcendental Dialectic, which covers some of the major highlights of Critical Project.

*(1.4) The Transcendental Project, or: How to Solve the Problem of Cognitive-Semantic Luck*

The CPR is, in one sense, a treatise on epistemology and ontology. But K's way of doing epistemology & ontology is sharply different not only from CR & CE, but also from contemporary epistemology & ontology.

This is because K grounds epistemology & ontology on the theory of human *cognition* (*Erkenntnis*) or conscious mental representation. This is explicitly stated in the letter to Herz, & is the key to understanding the Preface of CPR, not to mention the rest of the book as well.

A theory of human cognition focuses on the nature of the various *acts*, *objects*, and *representational contents* of conscious mental representation. A theory of content is also a theory of meaning. So K's theory of human cognition is also a semantics.

What I want to do now is to try, in two steps, to explain some of the basic doctrines of K's cognitive semantics, which he calls *Transcendental Idealism*, by comparing & contrasting it with CR and CE, & without using any Kantian technical terminology. If we can get a synoptic grasp of the basic notions that K is working with, & also of the basic philosophical moves that he makes, then the terminology can be fairly easily acquired against the backdrop of that synoptic grasp. Or otherwise put, once we have the Big Picture of what K is trying to argue, then all the little pictures will fall into place.

STEP 1: *Transcendental Idealism = (1) Transcendentalism + (2) Idealism = The Conformity Thesis*

**(1) Transcendentalism** = All the forms or structures of cognitions are imposed a priori by our innate spontaneous cognitive capacities (= cognitive faculties, cognitive powers)

**(2) Idealism** = All the proper objects of cognition are nothing but appearances or phenomena (i.e., mind-dependent, sensory, spatiotemporal, directly perceivable objects) and never things-in-themselves or noumena (i.e., mind-independent, non-sensible, non-spatiotemporal, real essences constituted by intrinsic non-relational properties).

Strong mind-dependence: X is strongly mind-dependent if & only if X exists only insofar as it is being consciously represented, & if all human minds went out of existence, X would go out of existence too.

Weak mind-dependence: X is weakly mind-dependent if and & only if necessarily, *were* human minds to exist, then they *would* be able to know X directly.

**(1) + (2) = K's "Copernican revolution"** = **The Conformity Thesis** =

It is not the case that human minds passively conform to the objects they cognize (as in Classical Rationalism & Classical Empiricism): on the contrary, necessarily all the proper objects of human cognition conform to (i.e., have the *same* form or structure as) our innate spontaneous cognitive capacities.

*STEP 2: The Problem of Cognitive-Semantic Luck*

What would justify Kant's asserting **Transcendentalism**, **Idealism Thesis** and **The Conformity Thesis**, i.e., what would justify his asserting the truth of Transcendental Idealism, or TI? I think that we can rationally reconstruct his fundamental or "master" argument for TI in the following way. Suppose that we accept, as initial suppositions,

> (i) the minimal Empiricist assumption that all human cognition begins in causally-triggered, direct, non-conceptual sense perception of contingent natural objects or facts,
>
> (ii) the minimal Rationalist assumption that we rational human animals actually cognitively possess some non-empirical or a priori mental representations, and that we also have non-empirical or a priori knowledge of some objectively necessary truths, e.g., in mathematics, logic, and metaphysics, and
>
> (iii) the minimal cognitive-semantic assumptions that (iiia) truth is the agreement (*Übereinstimmung*) of a belief with the object described by the propositional content of that belief, and (iiib) reference is the direct relation (*Beziehung*) between any cognition and its object.

For expository convenience, let us call all non-empirical or a priori mental representations, including a priori beliefs and a priori knowledge, "a priori cognitions."

What then rules out the possibility that the cognitive-semantic connection between our a priori cognitions on the one hand, and the truth-making objects or facts on the one hand, is nothing but a massive coincidence?

And if it *is* a massive coincidence, then the connection between our a priori cognitions and their truth-making objects or facts is merely *accidental* or *contingent*, and could just as easily have *failed* to obtain in at least some introspectively cognitively indistinguishable situations. If so, then a priori cognition is inherently unreliable and cannot constitute a priori knowledge. This deep skeptical worry is **The Problem of Cognitive-Semantic Luck.**

Now one possible solution to The Problem of Cognitive-Semantic Luck is that the truth-making objects or facts are all abstract, non-spatiotemporal, non-natural, non-empirical, and non-sensory in nature—say, they are constituted by Platonic Essences, Forms, or *Eide*—and that those truth-making objects or facts are directly encountered by our immortal souls in a previous condition of disembodied mindedness, and then in this embodied life, or perhaps in another later more fortunate embodied life of the same soul, we "remember" that earlier direct encounter, by means of philosophical dialectic. That is Plato's theory of *anamnesis*, and of course it is an early version of the *innate ideas* theory later held by Descartes, the Cambridge Platonists, and Leibniz.

But not only does the classical Platonic theory require the transmigration of immortal souls, it also provides no explanation whatsoever of either how immortal souls in a state of disembodied mindedness can ever directly *encounter* abstract, non-spatiotemporal, non-natural, non-empirical, non-sensory objects or facts, or how souls in their "human, all too human" embodied state can ever *re*-encounter them. In short, such encounters and re-encounters are a metaphysical mystery.

Another possible solution to The Problem of Cognitive-Semantic Luck is that the objects and facts are all concrete, spatiotemporal, natural, empirical, and sensory in nature, and that they *naturally cause* our a priori cognitions. That is the *classical Empiricis*t or Lockean-Humean solution. The basic problem with the classical Empiricist solution, however, is that it is incompatible with the initial assumption that the cognitions caused by these truth-making object or facts states of affairs are a priori, and not a posteriori. Otherwise put, how could these cognition be other than a posteriori, if their truth-making objects are strictly concrete, spatiotemporal, natural, empirical, and sensory natural causes of those cognitions?

And another pair of possible solutions to The Problem of Cognitive-Semantic Luck take the two-step strategy that, **first**, the truth-making objects or facts are all, again, abstract, non-spatiotemporal, non-natural, non-empirical, and non-sensory in nature, and **second**, an all-powerful, all-knowing, and all-good or non-deceiving God creates either

(i) a direct non-causal cognitive-semantic relation of *acquaintance* (*kennen*), or

(ii) an indirect non-relational cognitive-semantic *pre-established harmony*,

between the the a priori cognitions on the one hand, and the abstract, non-spatiotemporal, non-natural, non-empirical, and non-sensory truth-making objects or facts on the other.

Those, respectively, are the classical *Cartesian* and *Leibnizian* solutions. But given the fact that all the proper objects of a rational but also specifically human capacity for cognition are apparent, phenomenal, or manifest natural objects, and never noumena or things-in-themselves, the appeal to a non-deceiving God and to God's creation of humanly-inaccessible mysterious cognitive acquaintance relations or equally mysterious pre-established harmonies seems no better justified—in effect, no more than an arbitrary and question-begging appeal to a *deus ex machina* (= "a god from the machine," i.e., a standard theatrical device in classical Greek drama, when the gods arrive from the heavens & save the day)—than the skeptical hypothesis that the correspondence is nothing but a massive coincidence. Indeed, in the light of the implausibility of the Cartesian and Leibnizian *deus ex machina*-style solutions, what could decisively rule out the further skeptical possibility that the correspondence is *simply illusory* and has been created by an Evil Demon, i.e., by a God-like being who *is* a deceiver, given the introspective cognitive indistinguishability of at least some worlds in which this is possible?

In view of the failures of the classical Platonic, Empiricist, Cartesian, and Leibnizian solutions to The Problem of Cognitive-Semantic Luck, and assuming that these four possible solutions exhaust the logical space of all the most promising and relevant solutions to The Problem, then we can infer the truth of TI, by philosophical abduction (= inference-to-the-best-explanation) as the only adequate solution.

In the famous letter to Marcus Herz of 21 February 1772, and then again 15 years later in the B edition of the CPR, Kant formulates this basic argument for TI in the following ways:

> [I] was then making plans for a work that might perhaps have the title, "The Limits of Sense and Reason." I planned to have it consist of two parts, a theoretical and a practical. The first part would have two sections, (1) general phenomenology and (2) metaphysics, but this only with regard to its method. As I thought through the theoretical part [of *The Limits of Sense and Reason*], considering its whole scope and and the reciprocal relations of its parts, I noticed that I still lacked something essential, something that in my long metaphysical studies I, as well as others, had failed to pay attention to and that, in fact, constitutes the key to the whole secret of hitherto still obscure metaphysics. I asked myself: What is the ground of the relation of that in us which we call "representation" to the object? If a representation is only a way in which the subject is affected by the object, then it is easy to see how the representation is in conformity with this object, namely as an effect in accord with its cause, and it is easy to see this modification of our mind can *represent* something, that is, have an object. Thus the passive or sensuous representations have an understandable relationship to objects, and the principles that are deribed from the nature of our soul have an understandable validity for all things insofar as those things are supposed to be objects of the senses. In the same way, if that in us which we call "representation" were active with regard to the object, that is, if the object were created by the representation (as when divine cognitions are conceived as the archetypes of all things), the conformity of these representations to their objects could be understood. Thus the possibility of both an *intellectus archetypi* (on whose intuitions the things thmselves would be grounded) and an *intellectus ectypi* (which would derive the data for its logical procedure from the sensible intuition of things) is at least intelligible. However, our understanding, through its representations, is not

the cause of the object …. nor is the object the cause of the intellectual representations in the mind…. Therefore the pure concepts of the understanding must not be abstracted from sense perceptions, nor must they express the reception of representations through the senses; but though they must have their origin in the nature of the soul, they are neither caused by the object nor bring the object into being. In my dissertation I was content to explain the nature of intellectual representations in a mrerely negative way, namely, to state that they were not modifications of the soul brought about by the object.

However I silently passed over the further question of how a representation that refers to the object without being in any way affected by it can be possible. I had said: The sensuous representations present things as they appear, the intellectual representations present them as they are. But by what means are these things given to us, if not by the way in which they affect us? And if such intellectual representations depend on our inner activity, whence comes the agreement that that they are supposed to have with objects—objects that are nevertheless not possibly produced thereby? And the axioms of pure reason concerning these objects—how do they agree with these objects, since the agreement has not been reached with the aid of experience? In mathematics this is possible, because the objects before us are quantities and can be represented as quantities only because it is possible for us to produce their mathematical representations (by taking numerical units a given number of times). But in the case of relationships involving qualities—as to how my understanding may form for itself concepts of things completely a priori, with which concepts the things must necessarily agree, and as to how my understanding may formulate *real* principles concerning the possibility of such concepts, with which principles experience must be in exact agreement, and which nevertheless are independent of experience—this question, of how the faculty of understanding achieves this conformity with the things themselves, is still left in a state of obscurity.

Plato assumed a previous intuition of divinity as the primary source of the pure concepts of the understanding and of first principles. [Malebranche] believed in a still-continuing perennial intuition of this primary being. Various moralists have accepted precisely this view with respect to basic moral laws. Crusius believed in certain implanted rules for the purpose of forming judgments and ready-made concepts that God implanted in the human soul just as they had to be in order to harmonize with things. Of these systems, one may call the former the *influxum hyperphysicum* and the latter the *harmonium preastabilitatem intellectualem*. But the *deus ex machina* is the greatest absurdity one could hit on in the determination of the origin and validity of our knowledge. It has—beside its deceptive circle in the conclusion concerning our cognitions—also this additional disadvantage: it encourages all sorts of wild notions and every pious and speculative brainstorm. (*PC* 10: 129-135)

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 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 constitution of our faculty of intuition (*Anschauungsvermögens*), then I can very well represent the possibility to myself. (*CPR* Bxvi-xvii)

Now there are only two ways in which a **necessary** agreement of experience with the concepts of its objects can be thought: either the experience makes these concepts possible, or these concepts make the experience possible. The first is not the case with the categories (nor with pure sensible intuition); for they are *a priori* concepts, hence independent of experience (the assertion of an empirical origin would be a sort of *generatio aequivoca*). Consequently only the second way remains (as it were a system of the **epigenesis** of pure reason): namely, that the categories contain the grounds of the possibility of all experience in general from the side of the understanding…. If

> someone still wanted to propose a middle way between the only two, already named ways, namely, that the categories were neither **self-thought** *a priori* first principles of our cognition, nor drawn from experience, but were rather subjective predispositions of our thinking, implanted in us along with our existence by our author in such a ways that their use would agree exactly with the laws of nature along which experience runs (a kind of **preformation-system** of pure reason), then (besides the fact that on such a hypothesis no end can be seen to how far one might drive the presupposition of predetermined predispositions for future judgments) this would be decisive against the supposed middle way: that in such a case the categories would lack the **necessity** that is essential to their concept. For, e.g., the concept of cause, which asserts the necessity of a consequent under a presupposed condition, would be false if it rested only on a subjective necessity, arbitrarily implanted in us, of combining certain empirical representations according to a rule of relation. I would not be able to say that the effect is combined with the cause in the object (i.e., necessarily), but only that I am so constituted that I cannot think of this representation otherwise than as so connected; which is precisely what the skeptic wishes most, for then all our insight through the supposed objective validity of our judgments is nothing but sheer illusion, and there would ne no shortage of people who would not concede this subjective necessity (which must be felt) on their own; at least one would not be able to quarrel with anyone about that which merely depends on the way in which his subject is organized. (*CPR* B166-168)

*STEP 3: Classical Rationalism vs Classical Empiricism vs. Kant: shared assumptions basic disagreements*

| | CR | CE |
|---|---|---|
| 1) rep turn = primacy of consc mental reps = "the 'idea'-idea" | x | x |
| 2) kinds of truths | 2 | 2 |
| 3) basic objects of truths | SNs | EOs |
| 4) basic knowl of truths | a priori | a posteriori |
| 5) basic cognitive faculties | reason | sense |

rep turn = *the representational turn* = instead of focusing on objects in the world, focus on how we consciously represent or cognize that world (NB. CR & CE both share this particular assumption with each other & with Kant)
SNs = simple natures, real essences
EOs = empirical objects

K's doctrine of the limits of reason & the limits of sense can be understood as a dual critique of CR & CE which consists, first, in rejecting 3 shared false assumptions about cognition:

(a) the single source thesis = cognition is derived from a single source only

(b) 2 pronged forkism (a.k.a. Hume's Fork) = there are 2 & only 2 types of truths

(c) the passivity thesis = the mind passively conforms to its objects

What I will call *Kant's mitigated rationalism* then consists in developing a positive view that is neither CR nor CE yet also combines elements of both:

(i) the dual source thesis: cognitive faculty innateness (vs. cognitive content innateness) + sense experience as triggering input & raw data source (versus sense experience as determining source)

(ii) 3 pronged forkism (a.k.a. Kant's pitchfork) = 3 kinds of truth, including 2 irreducibly distinct kinds of a priori nec truth: purely conceptual nec truth (analytic truth = logically necessary truth) *versus* substantive or world-based nec truth (synthetic nec truth = non-logically necessary truth = "strongly" metaphysically necessary truth)

(iii) the activity thesis = the mind is essentially an innate capacity for spontaneously operating on & transforming sensory inputs, for rational agency, & for normativity

(iv) a new conception of the a priori = apriority as underdetermination by sense experience, not exclusion of sense experience: all our cognition *begins in* sensory experience but it is *not* the case that our cognition is wholly *derived from* sensory experience

Here is Kant's Transcendental Project in a nutshell, formulated as a leading question:

**How are synthetic a priori propositions possible?** (**Subsidiary issues: How are synthetic a priori propositions true? How is knowledge of synthetic a priori truths possible?)**

And here is K's answer to that leading question in a nutshell:

**Because transcendental idealism is true**. **Synthetic a priori propositions directly express the formal contributions of our innate spontaneous cognitive capacities to the content of our cognition, & transcendental idealism (=the necessary conformity of the basic structures of the world to the basic structures of our minds) guarantees the truth of these propositions, & we know these synthetic a priori truths because "reason has insight only into what it itself produces according to its own design" (**CPR: Bxiii).

# Table 1: Kant's Epigenetic Model of the Rational Human Mind

*Kant's Cognitive Psychology: A Flow Chart*
Bob Hanna

- Faculties are innate spontaneous capacities.
- X is either a noumenal object (thing-in-itself) or a phenomenal object (empirical thing).
- "S is P" corresponds to a real fact in the empirical world.

⇔⇳ mean 'delivers information to'

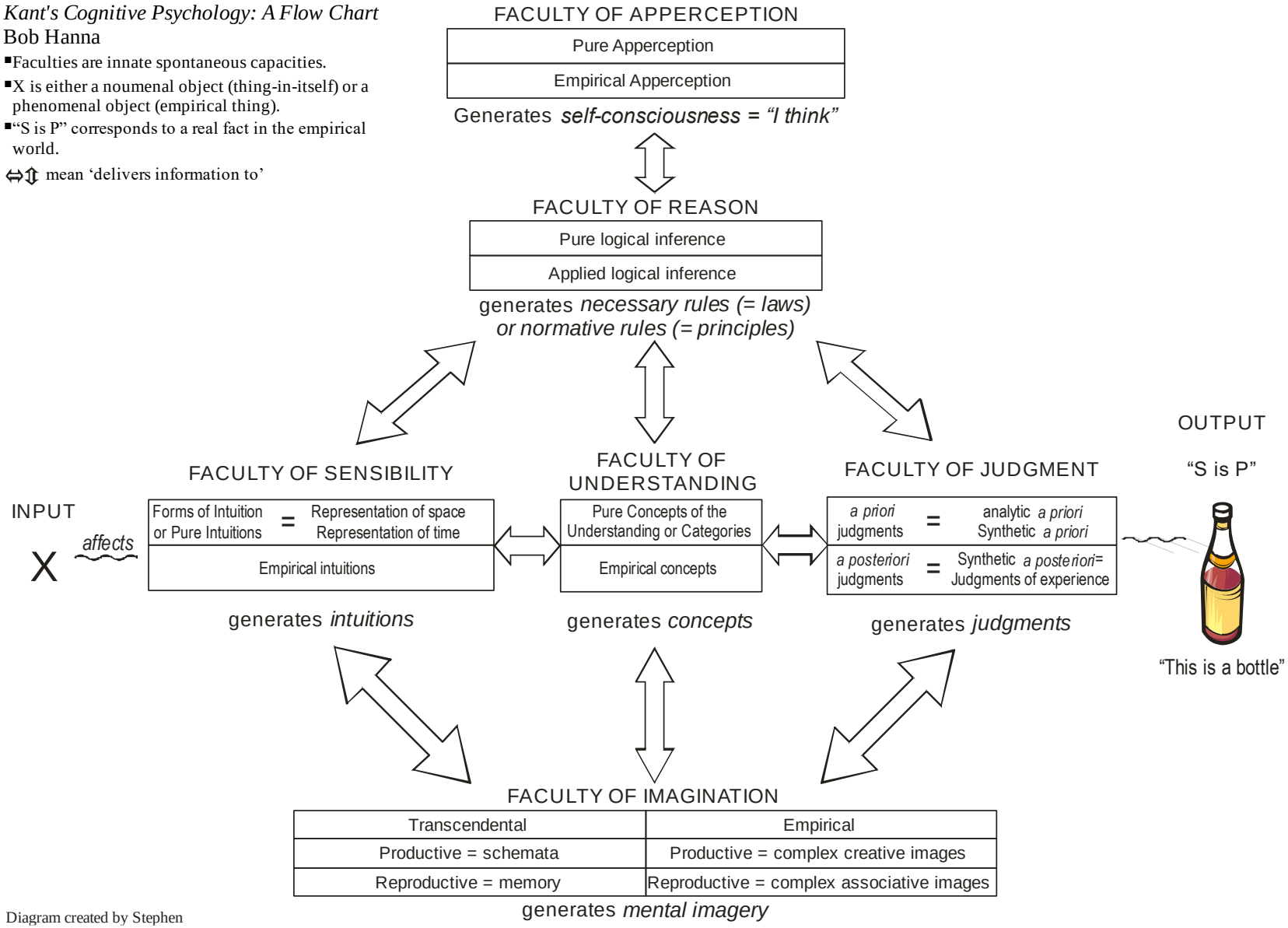



Diagram created by Stephen

## Table 2: The Structure of the *Critique of Pure Reason*

**Preface**
**Introduction**

**I. Transcendental Doctrine of Elements**

**Part I. Transcendental Aesthetic**

**Section I. On Space**
**Section II. On Time**

**Part II. Transcendental Logic**

**Division I. Transcendental Analytic**

**Book I: Analytic of Concepts**

**Clue to the Discovery of all Pure Concepts of the Understanding**
**Deduction of the Pure Concepts of the Understanding**

**Book II: Analytic of Principles**

**Schematism**
**System of all Principles:** **Axioms of Intuition**
**Anticipations of Perception**
**Analogies of Experience**
**Postulates of Empirical Thought**
**Phenomena and Noumena**

**Division II. Transcendental Dialectic**

**Transcendental Illusion**
**Pure Reason as the Seat of Transcendental Illusion**

**Book I: Concepts of Pure Reason**
**Book II: Dialectical Inferences of Pure Reason**

**Paralogisms (Psychology)**
**Antinomies (Cosmology)**
**The Ideal (Theology)**

**II. Transcendental Doctrine of Method**

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**

## LECTURE 2

Covers: CPR: 127-152 = A1-16/B1-30 (Introduction, both editions), 279-283 =A150-158/B189-193 (Supreme principles of analytic and synthetic judgments), 398-399 = A320-321/B376-377, and 684-690 = A820-831/B848-859 (Opining, Knowing, & Believing).

## The Introduction & Beyond: Basic Terms, Notions, & Distinctions

You will have noticed already that the first *Critique* is not an "easy read." One reason for this is Kant's constant use of technical terminology to express with precision his fundamental notions and distinctions. Another reason is that he often uses technical terms many pages before he actually defines them. For example, he does not define and unpack the fundamental notion of a "representation" (*Vorstellung*) until A320/B376—roughly 400 pages into the book! This exemplifies the truth that CPR is in many ways written as a book to be *re*-read, not read. In any case the purpose of these notes is to spell out the basic terms, notions, & distinctions that K. deploys both in the Introduction and throughout the rest of CPR.

### (1) Mental representations vs. objects; mental representational form vs. mental representational matter/content; sensations; cognitions vs. thoughts

According to Kant, the central fact about the mind is its capacity to represent (*vorstellen*), which is to say that the mind has something "to put before" (*stellen … vor*) it, and this something is a mental "representation" (*Vorstellung*). Our mental representational capacity cannot be further explained: it's simply a primitive fact about us. Mental representations, in turn, can be either conscious or nonconscious. The primary cognitive role of consciousness (*Bewußtsein*) is to contribute subjective integrity, or a well-focused and uniquely egocentric organization, to a mental representation.

A conscious mental representation is thus an "idea" in the broadest possible sense. Subjective conscious mental representations are internal or immanent to consciousness and lack fully determinate form or structure. Objective conscious mental representations, by contrast, are determinate ways of referring the mind to any sort of objects (i.e., some topic or target of the mind--what the representation is *about*), including the self considered as an object, as in self-consciousness or apperception. Objects of conscious mental representation also include existent or non-existent objects, and actual or possible objects. In short, conscious objective mental

representation in Kant's sense is essentially what the Scholastics (& later Brentano & Husserl & Meinong) call "intentionality."

For Kant, every objective conscious mental representation has both (i) a "form" (*Form*) and (ii) a "matter" (*Materie*) or "content" (*Inhalt*). The form of an objective conscious mental representation is its intrinsic structure. E.g., sensory perceptions have intrinsic spatial and temporal form or structure, and judgments have intrinsic logical form or structure. *Materie* is qualitative sensory content. *Inhalt* by contrast is intensional content: what Kant calls an objective conscious mental representation's "sense" or *Sinn* and also its "meaning" or *Bedeutung*. The sense, meaning, or intensional content of an objective conscious mental representation is the *information* that the mind has about its objects. Since the same object can be represented in different ways, there is a many-to-one relation between intensional contents (senses, meanings) and their corresponding objects.

Unfortunately, Kant also sometimes uses the term "form" to refer to purely psychological components of our use or grasp of an objective conscious mental representation. Form in this sense is somewhat similar to what Descartes called the "formal reality" of an idea, and the intensional content of an objective conscious mental representation in Kant's sense is somewhat similar to what Descartes called the "objective reality" of an idea. More precisely, for Kant the form of an objective conscious mental representation is what for lack of a better name I will call its *representational character*, by analogy with the "phenomenal character" of phenomenal consciousness. Representational character includes (a) the difference between clarity and unclarity, and between distinctness and indistinctness, (b) different subjective attitudes of all sorts, or what Locke called "postures of the mind," including but not restricted to propositional attitudes, and (c) and our direct conscious awareness of and ability to distinguish between and generalize over types of mental acts or mental operations of all different sorts (e.g., analysis, synthesis, memory, imagination, thought, judgment, etc.), which Kant calls "reflection" (*Überlegung*) and which is somewhat similar to Locke's "ideas of reflection." Nowadays, we'd call this *cognitive phenomenology*.

Conscious mental representations can be either subjective or objective, but in either case are necessarily accompanied by "sensations" (*Empfindungen*). The "matter" or qualitative phenomenal content of sensations--or what we would now call "qualia"--are intrinsic non-relational phenomenal properties of all conscious representations. More precisely, sensation is "the effect of an object on the capacity for representation, insofar as we are affected by it" (CPR A19-20/B34), or in other words, a sensation together with its content is nothing but the subject's direct response to endogenously- or exogenously- caused changes in its own state. Endogenously-caused sensations are "subjective sensations," or feelings, and exogenously-caused sensations are "objective sensations."

Objective conscious mental representations are also known as *cognitions* (*Erkenntnisse*), and this Kantian usage is essentially equivalent with the use of the term 'cognition' in contemporary cognition psychology.

Just to make things confusing however, in the B edition of CPR (see, e.g., at CPR Bxxvi, n.) Kant also sometimes uses the term 'cognition' in a somewhat narrower sense to mean *an objective conscious cognition of an actual or possible object of sense perception, an actual or possible empirical object, or empirical state-of-affairs*: **an empirically meaningful or objectively valid judgment**. This notion of a cognition then directly contrasts with the notion of mere *thought*, which is a conscious conceptual mental representation of any sort of object whatsoever, whether or not it is an object of actual or possible sense perception.

**(2) Sensibility vs. Understanding vs. Power of Judgment vs. Reason vs. Imagination vs. Apperception**

There are six fundamental innate spontaneous mental representational capacities (i.e., cognitive faculties, or cognitive powers), according to Kant:

*Sensibility* (*Sinnlichkeit*)is the cognitive faculty for sensory awareness, i.e., sense-perception, imagination, & also feeling in the broad sense—including pleasure & pain, emotion, and desires.

*Understanding* (*Verstand*) is the cognitive faculty for intellectual awareness or thought (conceptualization, describing).

*Power of judgment* (*Urteilskraft*) is the cognitive faculty for making judgments (framing propositions, framing beliefs).

*Reason* (*Vernunft*) is either (a) the cognitive faculty for logical inference (in particular syllogistic inference) & the systematic organization of thought (theoretical reason), or (b) the innate faculty for decision-making and forming volitional intentions on the basis of desires (practical reason). Otherwise put, reason is the faculty for recognizing and following necessary normative *principles*.

[NB. Since "discursive" means "concerned with or involving concepts," and since both the power of Judgment and Reason presuppose and use concepts, then the Understanding, the power of Judgment, & Reason jointly constitute the three *discursive capacities* of human cognition.]

*Imagination* (*Einbildungskraft*) is a cognitive faculty that has both generic and specific aspects. When taken generically, the imagination is the source or engine of all sorts of synthesis, or mental processing. But when taken specifically as a "dedicated" or task-sensitive cognitive faculty, the imagination also generates (1) the spatial and temporal forms of intuition, (2) mental images in conscious sensory states, (3) reproductive imagery or memories, and (4) "schemata," which are supplementary rules for

interpreting general conceptual rules in terms of more specific figural (spatiotemporal) forms and sensory images.

*Apperception* is the cognitive faculty for self-consciousness or judgment-based self-representation, i.e., second-order thought, or belief about one's own first-order consciousness.

**(3) Intuitions vs. concepts**

Intuitions (*Anschauungen*) for Kant are cognitions that are (i) immediate (directly referential, non-descriptive), (ii) sense-related, (iii) singular, (iv) object-dependent, and (v) prior to thought.

An *empirical intuition* is a either (i) a direct sensory grasp of some individual material object which affects (=causally interacts with) the mind--a perception of that object (e.g., that thing over there now), or else (ii) a direct sensory awareness of the subject's own mental state or condition—a first-order consciousness ("what it's like to be"). The capacity for the first kind of empirical intuition is *outer sense*, and the capacity for the second kind of empirical intuition is *inner sense.*

Concepts (*Begriffe*) for Kant are cognitions that are mediate (indirectly referential, descriptive), general, and essentially bound up with thoughts (*Gedanken*) & thinking (*Denken*).

An *empirical concept* is (i) a general intension, sense, or meaning that has been abstracted from empirical intuitions, (ii) what is expressed by such general words as 'red' and 'bachelor', and (iii) a general rule or categorizing procedure for organizing empirical intuitions. For example, the concept RED is what the word 'red' expresses (its linguistic intension, sense, or meaning), and it also enables the cognizer to recognize red objects.

Empirical concepts are all more or less complex, in the sense that each one contains an analyzable or rationally decomposable intensional content, which also functions as a general description of the set of actual or possible objects falling under that concept: the "comprehension" (*Umfang*) of the concept.

For example, the decompositional microstructure of the intensional content of the concept BACHELOR is:

<ADULT + UNMARRIED + MALE>.

Each of the sub-concepts "contained in" an empirical concept is what Kant calls a "mark" or "characteristic" (*Merkmal*). Hence the meaning-content of an empirical concept can be broken down into an ordered complex of marks. This is the same as its *analytic definition*.

A *pure or formal intuition* is a self-conscious non-empirical intuitional cognition of space or time. For Kant, the representation of space is the necessary a priori subjective form of all outer empirical intuitions (= the form of outer sense), and the representation of time is the necessary a priori subjective form of all inner empirical intuitions (= the form of inner sense). Moreover for Kant, as he argues in the Transcendental Aesthetic, space is "nothing but" the representation of space, and time is "nothing but" the representation of time. So space and time necessarily conform to the forms of our human sensibility; the pure or formal intuition of space is also the self-conscious awareness of the representation of space; and the pure or formal intuition of time is also the self-conscious awareness of the representation of time.

A *pure concept* is a second-order non-empirical concept, or a non-empirical concept that classifies or categorizes empirical concepts. For example, the empirical concept of a chair falls under the pure concept of enduring things (the category of substance).

An *Idea of reason* is a third-order non-empirical concept, or a non-empirical concept that applies to pure concepts in such a way that that pure concept is taken to apply to objects subsisting beyond all possible sensory experience of them (= noumenal objects).

**(4) Judgments, assertion, and knowing**

A *judgment* (*Urteil*) is a logically-organized unity of concepts and concepts, or concepts and intuitions. To judge is to predicate, that is, either (i) to apply or ascribe a concept to a thing or things referred to by intuition, or else (ii) to apply or ascribe one concept to another concept.

Every judgment contains an assertible content, truth-bearer, or "proposition (*Satz*); and to assert a proposition is to "take-it-for-true" (*Fürwahrhalten*) (CPR A820/B848).

Assertion has two distinct modes: believing (*Glauben*) and conviction (*Überzeugung*) (CPR A820-822/B848-850). Believing is assertion based on evidence that is subjectively or individually but not universally or objectively cognitively compelling. Conviction is assertion based on evidence that is universally cognitively intrinsically compelling and also objectively sufficient for the truth of the judgment. Hence conviction is both subjectively certain and objectively grounded, i.e., it constitutes genuine scientific knowing (*Wissen*).

**(5) Transcendental vs. empirical; innateness and spontaneity**

A mental representation is *transcendental* when it is either part of, or derived from, our innate spontaneous (and thus a priori) cognitive capacities.

A representation is *empirical* when it is directly related to, or derived from, what is "given" (i.e., received from beyond the mind) in sensibility, i.e., what is contingently supplied to our sensory capacities by nature, including human nature. So *empirical = sensible + contingent + natural*. For example, both outer and inner sense perceptions

are empirical. A representation is *pure* when it contains no empirical content whatsoever.

The *innateness* of a mental capacity means that the capacity is intrinsic to the mind, and not the acquired result of experiences, habituation, or learning. Correspondingly, the *spontaneity* of a mental capacity implies that the acts or operations of the capacity are

> (i) causally and temporally unprecedented, in that (ia) those specific sorts of act or operation have never actually happened before, and (ib) antecedent events do not provide fully sufficient conditions for the existence or effects of those acts or operations,
> **(**ii) underdetermined by external sensory informational inputs and also by prior desires, even though it may have been triggered by those very inputs or motivated by those very desires,
> (iii) creative in the sense of being recursively constructive, or able to generate infinitely complex outputs from finite resources, and also
> (iv) self-guiding.

**(6) A posteriori vs. a priori**

The phrases 'a posteriori' and 'a priori function either (i) as adverbs that apply to acts of judgment/belief, intuition, conceptualization, etc., or (ii) as adjectives that apply to propositions, intuition-contents, conceptual contents, etc.

By "empirical facts" I mean: *sensory experiences &/or contingent natural objects or facts*.

Roughly speaking, aposteriority is *dependence on experience* and apriority is *independence of experience*. But these rough characterizations are crucially vague & need to be made precise, as follows.

An act of judgment/belief, intuition, conceptualization, etc., is *a posteriori* if & only if empirical facts necessarily determine that act.

(Contemporary translation: An act of judgment/belief, intuition, conceptualization, etc., is *a posteriori* if & only if its existence and specific character are either *grounded on* or *strongly supervene on* empirical facts.)

An act of judgment/belief, intuition, conceptualization, etc., is *a priori* when empirical facts do not necessarily determine—i.e., they modally underdetermine—that act.

(Contemporary translation: An act of judgment/belief, intuition, conceptualization, etc., is *a priori* if & only if its existence and specific character are neither *grounded on* or *strongly supervene on* empirical facts.)

A concept or intuition is *a posteriori* if and only if empirical facts necessarily determine the semantic content or extension of that concept, or the non-conceptual content or reference of that intuition.

(Contemporary translation: A concept or intuition is *a posteriori* if and only if the semantic content or extension of that concept, or the non-conceptual content or reference of that intuition, are either *grounded on* or *strongly supervene on* empirical facts.)

A concept or intuition is *a priori* if and only if empirical facts do not necessarily determine—i.e., they modally underdetermine—the semantic content or extension of that concept, or the non-conceptual content or reference of that intuition.

(Contemporary translation: A concept or intuition is *a priori* if and only if the semantic content or extension of that concept, or the non-conceptual content or reference of that intuition, are neither *grounded on* nor *strongly supervene* on empirical facts.)

A judgment/belief or proposition is *a posteriori* if & only if either its meaning or its truth or its justification is necessarily determined by empirical facts.

(Contemporary translation: A judgment/belief or proposition is *a posteriori* if & only if its meaning or its truth or its justification are either *grounded on* or *strongly supervene* on empirical facts.)

Hence all a posteriori propositions are only contingently true or false. Examples: "Socrates is a philosopher." "Roses are red."

A judgment or proposition is *a priori* if & only if neither its meaning nor its truth nor its justification is necessarily determined by empirical facts = its meaning, truth, & justfication are all modally underdetermined by empirical facts.

(Contemporary translation: A judgment or proposition is *a priori* if & only if neither its meaning nor its truth nor its justification are either *grounded on* or *strongly supervene on* empirical facts.)

Moreover, according to Kant every *a priori* proposition is necessarily true, and also every necessarily true proposition is a priori. Examples: "Bodies are extended." "2+2=4." "Every event has a cause."

A proposition is "absolutely *a priori*" when, assuming it it is logically derived from another proposition, it is derived only from a necessary proposition: hence it is also *absolutely necessary*.

A proposition is "relatively *a priori*" when it is logically derived from a contingent proposition; hence a relatively a priori proposition is only hypothetically or conditionally necessary.

A proposition is *pure* when it contains no empirical constituents whatsoever in its content. Examples include "2+2=4" and "For all propositions *P*, ~(*P* &~*P*)." Otherwise it's "impure."

NB. Some a priori propositions are also *impure* because they specifically contain empirical concepts or content--e.g., "Every event has a cause," "If Socrates is a bachelor, then Socrates is unmarried," "F=ma," & Newton's laws of motion.

**(7) Synthesis vs. analysis; analytic vs. synthetic**

According to Kant, the mind has the ability to carry out two fundamental operations with respect to concepts, intuitions, judgments, and other representational contents: synthesis and analysis.

To *synthesize* is to combine several otherwise disparate representational contents into a structured unity, or pattern, of some sort. Otherwise put, synthesis is information processing, or the mental generation of representations. See also (2) above, under *Imagination*.

To *analyze* is to decompose a concept, intuition, judgment or other mental representation into some or all of its basic constituents, as given in some structure. For example, the analysis of the concept BACHELOR (i.e., the analysandum) produces the decomposed conceptual microstructure <ADULT + UNMARRIED + MALE> (i.e., the analysans).

No analysis of a representation is possible unless the mind has already synthesized that representation's content. Hence all analysis presupposes synthesis.

'Analytic' and 'synthetic', by contrast, are adjectives applying primarily to judgments or propositions.

A proposition is *analytic* if and only if its denial entails a conceptual or logical contradiction, hence it is necessarily logically true. This is the necessary and sufficient condition of analyticity, or "the universal and completely sufficient **principle of all analytic judgments**" (CPR: A151/B191). This criterion applies not only to analytic judgments having subject-predicate ("categorical") form but also to analytic judgments that do *not* have subject-predicate form, e.g., negative judgments, conditional ("hypothetical") judgments, disjunctive ("either…or") judgments, and conjunctive ("both…and …") judgments. More generally, *all* necessary truths of propositional or sentential logic are analytic, and *some* necessary truths of predicate logic are also analytic.

Unfortunately, Kant very misleadingly initially explicates the notion of analyticity *by giving two sufficient conditions for the analyticity of subject-predicate or categorical judgments*:

(i) a subject-predicate judgment is analytic if its predicate-concept is "contained in" its subject-concept, i.e., the predicate-concept is a more general concept that wholly includes the subject-concept as a more specific kind—e.g., BACHELOR is a more specific kind that is wholly included under UNMARRIED and MALE so BACHELOR *contains* UNMARRIED and MALE,

and

(ii) a subject-predicate judgment is analytic if it is true by virtue of the predicate-concept's being identical with [part of] the definition of subject-concept, i.e., the definition of BACHELOR includes UNMARRIED and MALE (and probably also ADULT), so the predicates UNMARRIED and MALE (and probably also ADULT) are identical with [part of] the definition of BACHELOR.

There are two reasons why these initial Kantian explications of analyticity are very misleading:

**first**, they provide *only sufficient conditions* of analyticity, NOT *necessary & sufficient* conditions (i.e, they provide "if" conditions only, NOT "if and only if" conditions),

and **second**, they provide sufficient conditions for analyticity with respect to only *one* kind of analytic judgment (i.e., subject-predicate or categorical judgments), NOT *all kinds* of analytic judgments.

So, again, it cannot be emphasized too strongly or repeated too often that "the universal and completely sufficient **principle of all analytic judgments**" is this: *A proposition is analytic if and only if its denial entails a conceptual or logical contradiction.*

NB. Analytic judgments *can* have empirical content, e.g., "Gold is a yellow metal" and "If Socrates is a bachelor, then Socrates is unmarried."

A proposition is *synthetic* if and only if (a) its denial is conceptually and logically self-consistent, and (b) both its meaning and truth/falsity necessarily depend at least in part on empirical or pure intuition—e.g., "Bodies have weight," "Space has three dimensions only," "2+2=4," "Every event has a cause," "F=ma," and Newton's laws of motion.

**(8) Analytic/synthetic + a posteriori/apriori; objective validity**

The two sets of distinctions given in (6) and (7) can also be interwoven.
An analytic proposition is necessarily and strictly universally true because its truth consists either in an intrinsic or essential connection between concepts, or in its being logically true. So it's true in every logically possible world.

And an analytic proposition is automatically also a priori because even if the concepts contained in the proposition are empirical, its truth-maker depends on conceptual

content or logic *alone*, and therefire it is necessarily underdetermined by all empirical facts.

Synthetic a posteriori propositions are contingent propositions whose meaning and truth/falsity necessarily include empirical intuitions (which makes them synthetic), and also their meaning, truth, and justification is necessarily determined by empirical facts Examples: "Socrates is a philosopher." "Roses are red."

Synthetic a priori propositions are consistently deniable, hence *not logically true*, yet still necessarily true propositions (in the sense that they are true in all and only the members of a complete class of possible worlds, namely the worlds of possible human sense experience) whose truth necessarily depends in part on pure intuition, together with all the other conditions for the possibility of objective experience that are presupposed by pure intuition. Examples: Truths of arithmetic. Truths of geometry. Certain metaphysical propositions, e.g., "Every event has a cause." Natural laws, e.g., Newton's Inverse Square Law.

In contrast with analytic propositions, which tell us only about our concepts and about pure logical laws, synthetic propositions also tell us about the actual world of human experience (reality, nature). More precisely, synthetic a posteriori propositions tell us about contingent facts, and synthetic a priori truths tell us about the actual world's essential or necessary features, or its underlying metaphysics. Hence synthetic a priori propositions are *non-logically metaphysically necessary truths*.

Not every proposition that has the logical form of a synthetic a priori proposition (i.e., that it purports to be non-logically metaphysically necessary, and it is consistently deniable) is either fully meaningful or true/false. Indeed, some propositions having the logical form of a synthetic a priori proposition are intelligible (thinkable, logically consistent) but *not fully meaningful*, hence lack a truth-value. E.g., "God exists."

In order to be fully meaningful, a synthetic a priori proposition (and in fact every other type of conscious objective mental representation as well) has to meet a fundamental condition of meaningfulness: that it applies directly or indirectly to actual or possible objects of human sensory intuition (i.e., objects of possible human experience). If it does apply, then it is *objectively valid* and thereby fully meaningful. (If, in addition, it also applies to some *actual* objects of experience, then it is not merely objectively valid but *objectively real*.)

If the proposition or mental representation doesn't apply to actual or posisble objects of human experience, however—and, in particular, if it could be meaningful or true only if it depended on an *intellectual intuition*, i.e., on a kind of intuition we humans do not possess, but only a divine being could possess—then it is "empty" (*leer*), i.e., not fully meaningful. NB. Not every empty representation is *nonsense* (*Unsinn*). Some of them, i.e., the intelligible ones, enable us to "think" non-sensory objects even if we cannot "cognize" (in the narrow sense) such objects by means of them. For the distinction between thought and cognition in the narrow sense, see **(1)** above.

**(9) Transcendental idealism; appearances vs. things-in-themselves; noumena vs. phenomena; transcendental vs. transcendent**

As we know already from **(5)**, a mental representation is "transcendental" when it is part of, or derived from, our innate spontaneous cognitive faculties. Necessarily, whatever is transcendental is also a priori.

Kant's thesis of *idealism* is that all the proper objects of our specifically human sort of cognition (NB. '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 or noumena.

*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.

Appearances or phenomena come in two flavors: partially or wholly undetermined; and fully determined. Partially or wholly undetermined appearances are unconceptualized objects of empirical intuition. Fully determined appearances are fully conceptualized objects of empirical intuition, also known as *objects of experience*. Noumena, by contrast, are *non-appearances* or *non-phenomena*. But, like appearances or phenomena, noumena also come in two flavors.

*Things-in-themselves* or *noumena in the positive sense* are beings (whether objects or subjects) that are strictly mind-independent (= they can exist even if no minds actually exist or even if no minds can possibly exist), non-sensory, and have a real essence which is a set of intrinsic non-relational properties. Things-in-themselves or noumena in the positive sense are uncognizable (in the narrow sense), 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 positive noumena are Platonic Forms or Ideas, Leibnizian monads, angelic spirits, God, unobservable physical microstructures (e.g., Locke's "real internal constitutions" of physical things) etc.

*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 an intrinsic relational non-sensory property, such as *being four* (as in: "John, Paul, George, & Ringo are four"), *being left-handed, being the cause of,* or *being beautiful, being morally right,* etc.

NB. The difference between intrinsic (i.e., necessary or essential) non-relational properties & intrinsic relational properties is fundamental for Kant. Empirical objects can have only intrinsic *relational* properties, not intrinsic *non-relational* properties. Only things-in-themselves (if they existed) could have intrinsic non-relational properties. Empirical objects can, of course, also have extrinsic (i.e., contingent or accidental) non-relational properties (e.g., something's changing color over time) or relational properties (e.g., something's changing weight over time).

The thesis of *transcendental idealism* is that the essential forms or structures of all appearances or phenomena *necessarily conform to* the innately specified non-empirical forms or structures of our spontaneous cognitive capacities.

In this special formal-structural sense, every appearance or phenomenon is "constructed" (as in: co-structured) by the human mind.

So the essential form or structure of every appearance is isomorphic (and possibly also type-identical, if *strong* transcendental idealism is true) to some or another a priori essential form or structure of the mind.

The particular objects we cognize, in turn, are either literally token-identical (if strong transcendental idealism is true) or else merely isomorphic (if weak transcendental idealism is true) to the objective intensional contents of our objectively valid representations (of those very objects).

It is crucially important to note that *both the existence and non-existence of things-in-themselves or positive noumena is logically consistent with the thesis of idealism*. That is: given the truth of idealism, it is logically possible that things-in-themselves exist and also logically possible that they don't exist. This distinguishes Kant's relatively weak sort of idealism from other stronger types--e.g., Berkeley's.

Moroever, and even more importantly, because things-in-themselves or positive noumena are both uncognizable (in the narrow sense) and scientifically unknowable, then we can't know either whether they exist or what their nature is, or whether they don't exist. *Kant is thus completely and consistently, and indeed radically, agnostic about the existence or non-existence and nature of things-in-themselves or positive noumena*.

Kant's radical ontological agnosticism should be compared & contrasted with Hume's merely *epistemic* agnosticism, which says: We have no decisive empirical reason to believe that either external things beyond the mind, or the "secret connexions" of apparent natural causes & effects, exist. By sharp contrast, Kant says: *we know a priori that we cannot know whether things-in-themselves exist or do not exist.*

But perhaps most importantly of all, also radically unlike Hume, Kant is also committed to the existence of some *negative* noumena: intrinsic relational non-sensory properties of empirical objects (e.g., the *fourness* of The Fab Four).

That which is *transcendent*—as opposed both to that which is *immanent*, or contained in experience, and also to that which is *transcendental*, or necessarily connected with our a priori cognitive faculties—is that which either is, or represents, a positively noumenal entity or thing-in-itself. E.g., God is a transcendent entity; the Idea of God is a transcendent representation.

NB. Unfortunately, and headache-producingly, Kant sometimes fails to observe his own distinction, and uses 'transcendental' when he really means *transcendent*.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**

## LECTURE 3

Covers: CPR: 155-192 = A19-49/B33-73 (Transcendental Aesthetic, both editions) and 630-643 = A712-738/B740-766 (The Discipline of Pure Reason in Dogmatic Use).

## Space, Time, & Mathematics: The Transcendental Aesthetic

*(3.1) The many aims of the TA.*

The Transcendental Aesthetic (TA) is one of the most densely packed sections of CPR. This is because K is trying to do at least five things at once:

(1) Give a transcendental theory of our capacity for sensibility;
(2) Give a theory of the nature of space and time;
(3) Give a philosophy of mathematics;
(4) Establish & explain some synthetic a priori truths;
(5) Prove transcendental idealism.

Two questions immediately arise: What is his strategy for doing all of this in one argument? And, supposing we can formulate the argument, is it sound (= true premises + valid inferential connections)? The two aims of these lecture notes are to spell out the basic moves of Kant's argument, and to initiate a critical evaluation of it.

*(3.2) The nature and goal of Kant's argument.*

Kant says in the first *Critique* that "we have already traced the concepts of space and time to their sources by means of a transcendental deduction, and explained and determined their a priori objective validity" (*CPR* A87/B119-120), and then later in the *Prolegomena* he says that there is a "transcendental deduction of the concepts of space and time" which "explains also at the same time the possibility of pure mathematics" (*Prol* 4: 285).

According to my "cognitive-semantic" approach to Kant's transcendental idealism, I take a transcendental deduction to be a demonstration of the objective validity--the empirical meaningfulness or cognitive significance--of an a priori representation *R* (whether *R* is an a priori concept, an a priori intuition, an a priori necessary proposition, or a systematic corpus of a priori necessary propositions), by means of demonstrating that *R* is the presupposition of some other representation *R**, which is assumed for the purposes of the argument to be objectively valid (*CPR* A84-94/B116-127, A156/B195).

It follows from these points, that Kant believes that a single line of transcendental argumentation establishes, in one fell swoop, both the objective validity of the a priori representations of space and time and also the objective validity of mathematics. What Kant wants to prove is this:

> The representation of space (r-space) and the representation of time (r-time), which are (i) the a priori necessary subjective forms of all empirical intuitions of appearances and (ii) identical to space and time respectively, are the conditions of the possibility of the fact that mathematical truths are synthetic a priori.

He puts this same argument-goal somewhat less compactly in the following way:

> Time and space are accordingly two sources of cognition, from which different synthetic cognitions can be drawn *a priori*, of which especially pure mathematics in regard to the cognitions of space and its relations provides a splendid example. Both taken together are, namely, the pure forms of all sensible intuition, and thereby make possible synthetic *a priori* propositions. But these *a priori* sources determine their own boundaries by that very fact (that they are merely conditions of sensibility), namely that they apply to objects only so far as they are considered as appearances, but do not present things in themselves. (*CPR* A38-39/B55-56)

(3.3) *A thumbnail sketch of the argument.*

Kant's overall argument can be broken down into four distinct steps:

STEP I: R-space and r-time are the a priori necessary subjective forms of all empirical intuitions of appearances.

(See the "Metaphysical Exposition of the Concept of Space" and the "Metaphysical Exposition of the Concept of Time").

STEP II: Space and time are strongly "transcendentally ideal," i.e., space and time are "nothing but" r-space and r-time, the a priori necessary subjective forms of all empirical intuitions of appearances, i.e., space and time necessarily conform to our representations of space and time.

(See the "Conclusions from the above Concepts [of Space]," "Conclusions from these Concepts [of Time]," and "Elucidation.")

***Commentary: I'm leaving open the possibility that the transcendental ideality of space & time is either a <u>strong</u> thesis, which identifies space and time with r-space and r-time, or a <u>weak</u> thesis, which says just that space and time necessarily conform to r-space and r-time, even though they exist, to a certain extent, independently of our minds, in the sense that even if there are or were no human minds—say, 10 million years ago—space and time still exist. See the criticism of STEP II in section (E).***

STEP III: Geometric and arithmetic truths alike are synthetic a priori.

(See section V of the Introduction, and section 13 and "Remark I" of the first part of the *Prolegomena*.)

STEP IV: True mathematical propositions are possible if and only if r-space and r-time are the a priori necessary subjective forms of all empirical intuitions of appearances.

(See the "Transcendental Exposition of the Concept of Space," section 1, chapter I of the "Transcendental Doctrine of Method," and sections 10-12 of the first part of the *Prolegomena*).

From these the conclusion mentioned in (B) follows directly:

CONCLUSION: R-space and r-time, which are (i) the a priori necessary subjective forms of all empirical intuitions of appearances and also (ii) "nothing but" space and time respectively, are the conditions of the possibility of the fact that mathematical truths are synthetic a priori.

*(3.4) A step-by-step reconstruction of the argument*

Prove: R-space and r-time, which are (i) the a priori necessary subjective forms of all empirical intuitions of appearances and also (ii) "nothing but" space and time respectively, are the conditions of the possibility of the fact that mathematical truths are synthetic a priori.

STEP I: Prove that r-space and r-time, as the forms of intuition, are the a priori necessary subjective forms of all empirical intuitions of appearances.

(1) Empirical intuitions are singular representations of undetermined apparent or sensible objects, and those representations in turn possess both matter and form.

"The undetermined object of an empirical intuition is called **appearance**." (*CPR* A20/B34)

"I call that in the appearance which corresponds to sensation its **matter**, but that which allows the manifold of appearance to be intuited as ordered in certain relations I call the **form** of appearance (*Form der Erscheinung*)." (*CPR* A29/B34)

(2) Appearances or objects of the senses are represented in empirical intuition by means of either outer (or spatial) sense or inner (or temporal) sense. R-space and r-time are the mutually distinct and jointly exhaustive (although not mutually exclusive) forms of intuition, and also the subjective forms of outer and inner sense respectively.

"By means of outer sense (a property of our mind) we represent to ourselves objects as outside us, and all as in space. In space their shape, magnitude, and relation to one another is determined, or determinable. Inner sense, by means of which the mind intuits itself, or its inner state, gives, to be sure, no intuition of the soul itself, as an object; yet it is still a determinate form, under which the intuition of its inner state is alone possible, so that everything that belongs to the inner determinations is is represented in relations of time." (*CPR* A22-23/B37)

"Time can no more be intuited externally than space can be intuited as something in us." (*CPR* A23/37)

"[R-]space is nothing other than merely the form of all appearances of outer sense, i.e., the subjective condition of sensibility, under which alone outer intuition is possible for us." (*CPR* A26/B42)

"[R-]time is nothing other than the form of inner sense, i.e., of the intuition of our self and our inner state." (*CPR* A33/B49)

(3) R-space and r-time are necessary conditions for the empirical intuition of appearances in outer and inner sense.

"[R-]space is a necessary representation, a priori, which is the ground of all outer intuitions. One can never represent that there is no space, although one can very well think that there are no objects to be encountered in it." (*CPR* A24/B38)

"[R-]time is a necessary representation that grounds all intuitions. In regard to appearances in general one cannot remove time, though one can very well take the appearances away from time." (*CPR* A31/B46)

(4) R-space and r-time, the forms of intuition, by means of an act of self-consciousness, can also be treated as "pure intuitions" or "formal intuitions," that is, singular nonconceptual representations of themselves as unique abstract relational totalities or formal-structural frameworks, thereby in turn representing space and time as singular infinite given wholes.

"[R-]space is not a discursive or ... general concept of relations of things in general, but a pure intuition." (*CPR* A24-25/B39)

"Space is represented as a given infinite magnitude." (*CPR* A25/B39)

"[R-]time is no discursive or ... general concept, but a pure form of sensible intuition." (*CPR* A31/B47)

"The infinitude of time signifies nothing more than that every determinate magnitude of time is only possible through limitations of a single time grounding it. The original representation, [r-]time, must therefore be given as unlimited." (*CPR* A32/B48)

"[R]-space and [r]-time and all their parts are **intuitions**, thus individual representations along with the manifold that they contain in themselves (see the Transcendental Aesthetic), thus they are not mere concepts by means of which the same consciousness is contained in many representations, but rather are many representations that are contained in one and in the consciousness of it; they are thus found to be composite, and consequently the unity of consciousness, as **synthetic** and yet as original, is to be found in them. This singularity of theirs is important in its application." (*CPR* B136 n.)
"[R]-space, represented as **object** (as is really required in geometry), contains more than the mere form of intuition, namely the **putting-together** (*Zusammenfassung*) of the manifold given in accordance with the form of sensibility in an **intuitive** representation, so that the **form of intuition** (*Form der Anschauung*) merely gives the manifold, but the **formal intuition** (*formale Anschauung*) gives unity of the representation." (*CPR* B160 n.)

(5) R-space and r-time are a priori. (From (3), (4), and the definition of 'a priori' as absolute experience-independence, or modal underdetermination by all empirical facts.)

"[W]e will understand by a priori cognition not those that occur independently of this or that experience, but rather those that occur *absolutely* independently of all experience." (*CPR* B3)

(6) Since r-space and r-time are (a) mutually distinct and jointly exhaustive (although complementary) necessary forms of the empirical intuition of appearances, (b) subjective forms of outer and inner sense, and (c) able to to be treated, via self-consciousness, as pure a priori nonconceptual intuitions of themselves as unique relational totalities or formal-structural frameworks, they are therefore the a priori necessary subjective forms of all empirical intuition of appearances. (From (1)-(2) and (5).) **QED**

STEP II: Prove that space and time are strongly transcendentally ideal, i.e., that space and time are "nothing but" the a priori necessary subjective forms of all empirical intuition of appearances, i.e., that space and time necessarily conform to r-space and r-time respectively.

***Commentary: It's an argument by trilemma: Either P or Q or R. Not P and not Q. Therefore R.***

(7) Space and time are either (a) things-in-themselves, (b) ontologically dependent on things-in-themselves (either as monadic intrinsic properties of things-in-themselves or as extrinsic relations between things-in-themselves), or else (c) strongly transcendentally ideal, i.e., "nothing but" the a priori necessary subjective forms of all empirical intuitions of appearances. And there are no other alternatives.

"Now what are space and time? Are they real essences (*wirkliche Wesen*)? Are they only determinations or relations of things, yet ones that would pertain to them even if they were not intuited, or are they relations that attach only to the form of intuition alone, and thus to the subjective constitution of our mind, without which these predicates could not be ascribed to anything at all?" (*CPR* A23/B37-38)

(8) But space and time are neither things-in-themselves nor ontologically dependent on things-in-themselves (either as intrinsic monadic properties of things-in-themselves or as extrinsic relations between things-in-themselves).

"Those .. who assert the absolute reality of space and time, whether they assume it to be subsisting or only inhering, must themselves come into conflict with the principles of experience. For if they decide in favor of the first ... then they must assume two eternal and infinite self-subsisting non-entities (space and time), which exist (yet without there being anything real) only in order to comprehend everything real within themselves. If they adopt the second position ... and hold space and time to be relations of appearances ... that are abstracted from experience ... then they must dispute the validity or at least the apodictic certainty of a priori mathematical doctrines in regard to real things (e.g., in space), since this certainty does not occur a posteriori." (*CPR* A39-40/B56-57)

(9) Therefore space and time are strongly transcendentally ideal, i.e., space and time are "nothing but" the a priori necessary subjective forms of all empirical intuition of appearances, i.e., space and time necessarily conform to r-space and r-time respectively. (From (7) and (8).) **QED**

***Commentary: The strong transcendental ideality of space & time does <u>not</u> itself logically yield the strong transcendental ideality of all things in space & time (i.e., that they're nothing but appearances or phenomena), because the spatial & temporal properties of those things might still be merely <u>extrinsic</u> relational properties of them.***

***So in order to derive the strong transcendental idealityof all things in space & time, Kant must also assume the truth of another principle, which I'll call <u>the intrinsicness of space & time</u>, which says that the spatial & temporal properties of all things in space & time are <u>intrinsic</u> relational properties of those things. E.g., the left- or right-handedness of a hand is an intrinsic relational property of that hand.***

STEP III: Prove that mathematical truths are synthetic a priori.

(10) Mathematical truths are a priori and necessary, not a posteriori and contingent.

"[M]athematical propositions are always a priori judgments and are never empirical, because they carry necessity with them, which cannot be derived from experience. But if one does not want to concede this ... I will restrict my proposition to **pure mathematics**, the concept of which already implies that it does not contain empirical but merely pure a priori cognition." (*CPR* B14-15)

"Here [in pure mathematics] is a great and established branch of knowledge ... carrying with it thoroughly apodeictical certainty, i.e., ... necessity, which therefore rests on no empirical grounds." (*Prol* 4: 280)
(11) Mathematical truths are synthetic, not analytic.

***Commentary: Mathematical propositions are synthetic, because (a) they fail the criteria of analyticity (in particular, they are consistently deniable); and (b) they are intuition-dependent.***

***Here's a sample of how one might prove syntheticity in the case of geometry:***

***(1) The syntheticity (& apriority) of any proposition P is established by showing (a) that P is consistently deniable, and (b) that the truth of P presupposes pure intuition. [Assumption]***

***(2) Proposition P: Necessarily, cone A and cone B (which have identical dimensions yet are left-right mirror images of one another, i.e., they're "enantiomorphs") are incongruent. [Assumption]***

***(3) But there are conceivable or thinkable worlds in which Proposition P is not true. [Thinkability]***

***(4) Only if the pure intuition of 3-D orientable Euclidean space is presupposed, can the truth of proposition P be made consistent with the truth of (3). So it follows that P is synthetic. [From (1), (2), and (3)]***

***Explication of (4): An orientable space is a space with intrinsic directions (e.g., up-down, left-right, backwards-forwards, inside-outside, above-below, etc.). Non-orientable Euclidean spaces are also logically possible, e.g., spaces in which the Möbius Strip or the Klein Bottle is embedded.***

***Thus (4) is saying that necessarily (in all and only 3-D orientable Euclidean spaces) cones A and B are incongruent counterparts. Which seems a priori true. But it is also logically possible for A and B to be congruent in a 4-D orientable Euclidean space – Wittgenstein, e.,g., makes this point in the Tractatus at prop. 6.36111). If so, then A and B are not logically necessarily incongruent. And in non-orientable spaces Proposition P is also not true, simply because it has no truth-value in that world. So Proposition P is not logically necessarily true, even though it is non-logically (=synthetically) necessarily true.***

***(5) The same general point goes, with only trivial variations in the reasoning, for <u>any</u> truth of geometry. [Generalization of (4)]***

***(6) So geometry is synthetic (a priori). [From (5)]***

"The concept of twelve is by no means already thought merely by my thinking of that unification of seven and five, and no matter how long I analyze my concept of such a possible sum I will not find twelve in it.... That 7 **should** be added to 5 I have, to be sure, thought in the concept of a sum = 7+5, but not that this sum is equal to the number 12. The arithmetic proposition is therefore always synthetic." (*CPR* B15-16)

"Just as little is any principle of pure geometry analytic. That the straight line between two points is the shortest is a synthetic proposition. For my concept of **the straight** contains nothing of quantity, but only a quality. The concept of the shortest is therefore entirely additional to it, and cannot be extracted out of the concept of the straight line by any decomposition (*Zergliederung*)." (*CPR* B16)

(12) Therefore mathematical truths are synthetic a priori. (From (10) and (11).) **QED**

STEP IV: Prove that r-space and r-time are the conditions of the possibility of mathematical truths.

(13) R-space and r-time are necessary conditions of the objective validity of the truths of geometry and arithmetic.

"Geometry is a science that determines the properties of space synthetically and yet a priori." (*CPR* A25/B41).

"Now the intuitions which pure mathematics lays at the foundation of all its cognitions and judgments which appear at once apodeictic and necessary are space and time.... Geometry is based on the pure intuition of space. Arithmetic attains its concepts by the successive addition of units in time.... Both representations, however, are merely intuitions." (*P* 4: 283)

(14) R-space and r-time (when taken together with pure general analytic logic) are the jointly sufficient conditions of the objective validity of the truths of geometry and arithmetic.

"To determine an intuition a priori in space (shape), to divide time (duration), or merely to cognize the universal in the synthesis of one and the same thing in time and space and the magnitude of an intuition in general (number) which arises from that: that is a **rational concern** through construction of the concepts, and is called **mathematical**." (*CPR* A724/B752)

"[T]he intuitions which pure mathematics lays at the foundation of all its cognitions and judgments which appear at once apodeictic and necessary are [r-]space and [r-]time. For mathematics must first exhibit all its concepts in intuition, and pure mathematics in pure intuition, i.e., it must construct them." (*Prol* 4: 283)

"The ground of mathematics actually is pure intuitions, which make its synthetic and apodeictically valid propositions possible." (*P* 4: 285)

(15) Therefore r-space and r-time are the conditions of the possibility of mathematical truths. (From (13) and (14).) **QED**

CONCLUSION

(16) R-space and r-time, which are (i) the a priori necessary subjective forms of all empirical intuitions of appearances and also (ii) "nothing but" space and time respectively, are the conditions of the possibility of the fact that mathematical truths are synthetic a priori. (From (6), (9), (12), and (15).) **QED**

"Time and space are accordingly two sources of cognition, from which different synthetic cognitions can be drawn a priori, of which especially pure mathematics in regard to the cognitions of space and its relations provides a splendid example. Both taken together are, namely, the pure forms of all sensible intuition, and thereby make possible synthetic a priori propositions. But these a priori sources of cognition determine their own boundaries by that very fact (that they are merely conditions of sensibility), namely that they apply to objects only so far as they are considered as appearances, but do not present things-in-themselves." (*CPR* A38-39/B55-56)

*(3.5) Criticism of the argument.*

STEP I

See (2): Granting that the forms of outer and inner sense (the representations of space and time) are mutually irreducible, are they mutually exclusive too? Is the temporal form of inner sense really more basic than the spatial form of outer sense? What are the implications of taking the representation of time to be prior to the representation of space, as opposed to taking them to be complementary and co-primordial?

See (4): How is it possible to have *pure* intuitions of space and time, i.e., spatial and temporal intuitions without any sensory content whatsoever?

See (4) again: Are the representations of space and time intuitions, or are they in fact conceptual representations?

STEP II

Is Kant's elimination of the two alternatives narrowly specific to Newton and Leibniz, or is it generalizable?

Has Kant spelled out and eliminated *all* possible alternatives to strong transcendental ideality? Is there in fact a Missing Alternative—e.g., that space and time can consistently be forms of sensibility *and* things-in-themselves? Let's call this the Classical Missing Alternative (CMA). The CMA is what the Fischer-Trendelenberg controversy was all about. But the CMA seems incoherent. If space & time are forms of sensibility, then they imply the existence of our sensibility; but if space & time are things-in-themselves, then

they are necessarily non-sensible. Furthermore, since the only way we have of telling space apart from time is in relation to our inner & outer sense, then if space & time were things-in-themselves, then we couldn't individuate them.

Now suppose that the CMA is incoherent. Are there any other possible missing alternatives? My own view is that there is at least one other missing alternative, which says that space & time are (a) the isomorphic satisfiers of our representations of space & time, (b) that space & time *can't* exist unless it's necessarily possible for us to exist, but (c) that space & time *can* actually exist even if we don't actually exist. So space and time exist if and only if necessarily, *were* human minds to exist, then we *would* be able to represent space and time a priori correctly & directly. So according to this new missing alternative, space & time are what I call *weakly transcendentally ideal*.

If we add to this new missing alternative the principle of the intrinsicness of space & time, then it follows that all things in space & time are only weakly transcendentally ideal, not strongly transcendentally ideal. This then gives us two distinct versions of transcendental idealism: strong TI and weak TI. So then the $64, 000.00 question is: quite apart from the exegetical question, what's more likely to be actually true—strong TI or weak TI?

STEP III

Can Kant safely assume without further argument that mathematical propositions are a priori?

How will the corresponding case be made for the syntheticity of arithmetic? Will it have to rely wholly on the downright dopey-seeming "finger counting" and "large number" arguments, or can it be more charitably reconstructed?

STEP IV

See (13) and (14): What are we to make of the fact that in the "Transcendental Exposition of the Concept of Time" in the Aesthetic, instead of arguing that *arithmetic* presupposes the representation of time, Kant argues that the "general doctrine of motion" (i.e., physics) presupposes the representation of time? More generally, is there is a basic asymmetry in Kant's accounts of arithmetic and geometry--i.e., is the representation of time to arithmetic, as the representation of space is to geometry? And if not, then what explains the asymmetry?

See (15): has Kant adequately established the 'the' or uniqueness? Could something *other* than the representations of space and time independently account for the synthetic apriority of math?

CONCLUSION

Even supposing that this argument is sound, has Kant offered us a theory of mathematical *knowledge*? If not, how will such a theory go? Clue: look at CPR: A713-738/B740-766.

What are the further metaphysical implications of the thesis that space and time are strongly transcendentally ideal, and of the corresponding thesis that all things in space and time are strongly transcendentally ideal? Are these theses fully intelligible? And even if they're fully intelligible, are they ultimately defensible? This of course raises the question of the intelligibility & defensibility of transcendental idealism itself.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 4

Covers: CPR: 185-192 = A41-49/B59-73 (General Remarks on TA), and 338-365 = A235-260/B294-315 (Phenomena and Noumena)

## Transcendental Idealism

*(4.1) What Transcendental Idealism is.*

In Kantian terminology, something is "transcendental" when it is part of, or derived from, our innate spontaneous cognitive capacities. Everything transcendental is also a priori.

Correspondingly, K's thesis of *representational transcendentalism* says that all the forms or structures of the representational contents of human cognitions are imposed a priori by our innate cognitive faculties.

K's thesis of *cognitive idealism* says that all the proper objects of our specifically human sort of cognition (NB. 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.

Notice that K's cognitive idealism is logically distinct from both Berkeley's *metaphysical or dogmatic idealism*, which says

> (a) that matter is impossible &
> (b) that necessarily (x) (x is either an idea in a conscious mind or x is a conscious mind),

and also from Cartesian *skeptical or problematic idealism*, which says

> that possibly nothing exists outside my own conscious states.

In contrast to Berkeleyan metaphysical or dogmatic idealism, (i) K's cognitive idealism does *not* apply to *all* objects whatsoever, (ii) K's cognitive idealism does *not* say that matter is impossible, and (c) K's cognitive 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)*.

In contrast to Cartesian skeptical or problematic idealism, K's cognitive idealism does *not* say that it is possible that nothing exists outside my conscious states (i.e., inner sense); on the contrary, K's cognitive idealism implies *that necessarily something actually exists outside my conscious states (i.e., inner sense) in space* = the conclusion of the Refutation of Idealism.

K's thesis of *Transcendental Idealism* says that *the spatiotemporal structures of the natural or physical world of appearances or phenomena necessarily conform to the mentalistic structures of our innate spontaneous cognitive capacities* = The Conformity Thesis = Kant's "Copernican Revolution."

In other words, the form or structure of every appearance is at least isomorphic to (weak TI), and perhaps also type-identical to (strong TI), some a priori form or structure of the mind. And the particular objects we cognize, in turn, are at least isomorphic to (weak TI), and perhaps also token-identical to (strong TI), the objective intensional contents of our objectively valid representations of those very objects.

*(4.2) K's five basic arguments for TI*

(i) The argument for TI from the possibility of cognitive-semantic luck.

(1) We have a priori cognitions.
(2) A priori cognitions correspond to their objects or truth-makers.
(3) If the connection between the objects or truth-makers and a priori cognitions is merely a massive coincidence, then a priori knowledge is impossible. For if it's merely a massive coincidence, then the connection is merely accidental in the actual world and in other introspectively indistinguishable possible worlds the connection fails altogether, hence it can always be doubted whether the connection really & truly holds. Let's call the possibility of massive coincidence *the problem of cognitive-semantic luck.*
(4) Assuming that our minds conform to objects, the best available alternative theories for avoiding the problem of cognitive semantic luck are Platonism, Cartesian clear & distinct intuition + a non-deceiving God, Leibnizian pre-established harmony, and Empiricism.
(5) But if either Platonism, Cartesianism, or Leibnizianism is true, then the connection between a priori cognition and its objects or truth-makers is a metaphysical mystery.
(6) And if Empiricism is true, then the connection between cognition and its objects or truth-makers yields *aposteriority*, not apriority.
(7) Therefore we must postulate that objects necessarily conform to our innate spontaneous cognitive capacities, i.e., we must conclude that TI is true because it is the only adequate way of ruling out the possibility of cognitive-semantic luck.

(ii) The argument for TI from the strong transcendental ideality of space and time.

> (1) Space and time are "nothing but" our a priori necessary subjective forms of sensory intuition, r-space and r-time (= the strong transcendental ideality of space and time).
> (2) Spatial properties and temporal properties are *intrinsic structural properties* of all actual and possible things in space and time (= the intrinsicness of space and time).
> (3) Therefore TI is true.

NB If K's Three Alternatives Argument fails & there is a genuine missing alternative, namely *the weak transcendental ideality of space & time* =

> Necessarily, space and time are the proper satisfiers of our human *representations* of space and time, and thus space and time exist only if it is necessarily really possible that human representers of space and time exist, and necessarily, if human cognizers *were* to exist, then they *would* be able to know space and time directly via pure intuition, but space and time *can* exist even if human minds *do not* exist,

then premise (1) is false. What would be the result of replacing premise (1) with the weak transcendental ideality of space & time?

(iii) The argument for TI from the gap between objective validity and truth.

> (1) It is possible for any judgment to be objectively valid (meaningful) but not true.
> (2) So all meaningful judgments might be false.
> (3) The only way to close the skeptical gap between objective validity and truth is to hold that TI is true.
> (4) Therefore TI is true.

(iv) The argument for TI from the synthetic a priori.

> (1) There are some objectively valid & true synthetic a priori judgments in mathematics, fundamental physics, and the transcendental metaphysics of human experience.
> (2) The best overall explanation of the objectively validity and the truth of synthetic a priori judgments is TI.
> (3) Therefore TI is true.

(v) The argument for TI from the third Antinomy

> (1) The only way to provide a coherent solution to the third Antinomy of Pure Reason (= the antinomy of freedom and natural determinism) is to hold that TI is true, and that freedom is noumenal while natural determinism is phenomenal. So they're consistent with each other.
> (2) Therefore TI is true.

*(4.3) Kant's theory of phenomena and noumena*

*Appearances* or *phenomena* are mind-dependent objects of actual or possible human sense perception.

Appearances or phenomena come in two 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*.

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

*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 K.'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 that they are *non-spatiotemporal*.

Things-in-themselves or noumena in the positive sense are beings (whether objects or subjects) that are uncognizable (in the narrow sense), 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.

*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 a non-sensory aspect. But in principle a negative noumenon can *also* be an empirical object or phenomenon. E.g.: The fact that John, Paul, George, & Ringo are 4 includes a non-sensory or negatively noumenal property, their fourness. The fact that Paul & Ringo exist includes a non-sensory or negatively noumenal property, their existence. The fact that a sunset is beautiful includes a non-sensory or negatively noumenal property, its beauty. The fact that you are a person includes a non-sensory or negatively noumen al

property, your personhood. Etc. Roughly speaking, to say that something empirical has a negatively noumenal property is to say that it has an *abstract* property.

*Things in general* comprise the class of all objects whatsoever, whether phenomena or noumena. *The transcendental object = X* is the generic notion of an object, or a representational variable ranging over all things in general.

*(4.4) Three different theories about noumena and phenomena*

The Two World (or Two Object) Theory:

Things-in-themselves (positive noumena) exist but are unknowable by us, & the classes of positive noumena and of phenomena comprise two necessarily mutually exclusive classes of really existing objects.

**Problems: substance dualism + Cartesian skepticism about existence claims + the double affection problem (= causal overdetermination).**

The Two Aspect (or Two Standpoint) Theory:

Version 1: There exists one and only one class of real objects, each member of which can be taken by us as either in-itself (noumenal standpoint) or for-us (phenomenal standpoint), and these two standpoints do not constitute intrinsic properties of those objects--which would then be to ascribe contradictory properties to the same objects--but rather are nothing but converse intentional relational properties of those objects (e.g., the relational property of x's being loved by y is a converse intentional property of x).

**Problem: what *are* the real objects then?**

Version 2: There exists one and only one class of real objects, *namely the appearances or phenomena*, each member of which can be taken by us as either in-itself (noumenal standpoint) or for-us (phenomenal standpoint), and these standpoints do not constitute intrinsic properties of those objects--which would then be to ascribe contradictory properties to the same objects-- but rather are nothing but converse intentional relational properties of those objects.

**Problem 1: Isn't it conceptually incoherent to ascribe positively noumenal properties to *phenomenal* objects?**

**Problem 2: if the real objects are phenomena that are only *regarded* noumenally, then the freedom of a phenomenally real human animal is only a function of belief, & we're not really free.**

The Two Concept (or Two Property) Theory:

Thesis 1: There is one and only one class of real objects, the authentic appearances or phenomena.

Thesis 2: Because things-in-themselves (positive noumena) are both uncognizable and scientifically unknowable, then we can't know whether they exist or don't exist, or what their nature is. So we must be completely and consistently agnostic about both the existence or non-existence and nature of things-in-themselves. From the standpoint of metaphysics & epistemology we can completely ignore them. Let's call this *methodological eliminativism about things-in-themselves*.

Thesis 3: There are logically consistent & semantically coherent but non-objectively valid concepts of things-in-themselves or positive noumena, hence things-in-themselves or positive noumena are merely logically possible.

Thesis 4: There are objectively valid, actually instantiated concepts of phenomenal things, hence there are real phenomenal properties.

Thesis 5: There are also objectively valid, actually instantiated negatively noumenal concepts of phenomenal things, hence there are real negatively noumenal properties.

Thesis 6: So there are are two irreducibly different sets of properties of phenomenal things: phenomenal properties & negatively noumenal properties.

**The Two Concept or Two Property Theory is a version of *property dualism without substance dualism*.**

**It follows that if freedom is a negatively noumenal property of phenomenally real human animals, then we *are* really free.**

*(4.5) Kant's empirical realism*

Kant both mitigates the sting and also enriches the substance of his transcendental idealism by explicitly defending, at the same time, *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)

In other words, K. is saying that when we eliminate thing-in-themselves as possible objects of human sensible cognition (although we remain capable of *thinking* about them

abstractly), focus exclusively on appearances instead, and then identify *them* with the real material objects in space, it follows that we perceive real material objects in space through our senses without any further intermediary (let us call this Kant's *direct perceptual realism*) and also that all the essential properties of real material objects in space are macrophysical directly perceivable or observable properties (let us call this Kant's *manifest realism*).

So for Kant the classical "veil of mere appearances" becomes *the field of authentic appearances*, in which all things are precisely what they seem to be.

In this sense, K.'s idealism is also paradoxically the most robust realism imaginable. Indeed, in the B edition he offers an explicit *refutation of idealism*.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 5

Covers: CPR: 425-431 = A366-380 (Fourth Paralogism) and 121n.-122n. + 326-329 = Bxxix-xli + B274-279 (Refutation of Idealism).

## The Refutation of Idealism

*(5.1) Kant's transcendental idealism and his anti-skeptical anti-idealism*

In the lecture notes for LECTURE 4, we saw that Kant's TI differs sharply from both Berkeley's *metaphysical or dogmatic idealism*, which says (a) that matter is impossible & (b) that necessarily (x) (x is either an idea in a conscious mind or x is a conscious mind),

and also from Cartesian *skeptical or problematic idealism*, which says that possibly nothing exists outside my own conscious states.

Now in contrast to Berkeleyan metaphysical idealism,

> (a) K's cognitive idealism does *not* apply to *all* objects whatsoever,
> (b) K's cognitive idealism does *not* say that matter is impossible, &
> (c) K's cognitive 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)*.

And in contrast to Cartesian skeptical or problematic idealism, K's cognitive idealism does *not* say that it is possible that nothing exists outside my conscious states (i.e., inner sense); on the contrary, K's cognitive idealism implies *that necessarily something actually exists outside my conscious states (i.e., inner sense) in space*. And this is in fact the conclusion of the Refutation of Idealism or RI.

Indeed, Kant regards both of these views, i.e., the Cartesian and Berkeleyan views, as inherently skeptical. And in the B Preface he famously says of his philosophical predecessors that

> it 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.)

Why do these purported refutations of idealism all fail? Kant's diagnostic insight, which he works out at length in the Paralogisms of Pure Reason (*CPR* A341-405/B399-432), is that the purported refutations all to some extent presuppose *the Cartesian model of our mind*, which effectively generates the very worries the refuters are trying so hard to refute.

According to the Cartesian model of our mind, the inner world of conscious experiences and the outer world of material objects are at once (i) essentially different and ontologically distinct from one another, in that their basic natures are incompatible (because the inner or mental is intrinsically immaterial and non-spatial, whereas the outer or physical is intrinsically material and extended in space), so it is metaphysically possible for one to exist without the other (= **ontological dualism**), and also (ii) epistemically mutually independent of one another, in that from the veridical cognition or knowledge of the one, no veridical cognition or knowledge of the other can ever be directly accessed or immediately inferred (= **indirect realism**). Hence the anti-skeptic is driven by a sort of philosophical despair to rely upon either a rationally undemonstrated "faith" in the existence of a non-deceiving God (as, e.g., in Descartes's epistemology) or in the dictates of common sense (as, e.g., in Thomas Reid's common sense realism).

In other words, one basic aim of the RI is to provide *a fundamental critique of the Cartesian model of our mind, & reject both its ontological dualism & its indirect realism.*

But Kant believes that if we get rid of the Cartesian model of our mind, we can be both transcendental idealists *and* empirical realists. So the RI is intended not only to refute skeptical idealism & to criticize the Cartesian model of the mind, but also to offer a radically new post-Cartesian **anti-dualist** & **direct realist** model of the mind, in order to make way for his own empirical realism.

*(5.2) The overall strategy of the RI*

The nerve of the RI can be found at B275-276 in the "Postulates of Empirical Thought" section in the first *Critique*; but I will also take into account the three "Notes" that immediately follow it in the text at B276-279, as well as a crucial footnote that Kant added at the last minute to the second Preface (*CPR* Bxxxix-xli).

The view that Kant is aiming to refute is what he officially calls "skeptical idealism" or "problematic idealism":

[T]he **skeptical idealist** [is] one who **doubts** [the existence of matter], because he holds [matter and its existence] to be unprovable. (*CPR* A377)

Problematic idealism ... professes only our incapacity for proving an existence outside us from our own [existence] by means of immediate experience. (*CPR* B275)

Skeptical or problematic idealism (the Cartesian skeptic's view) says that *possibly* the external world does not exist. This is to be sharply contrasted with what he officially calls

"dogmatic idealism" (Berkeley's view), which takes the modally stronger position that the external or material world "is false and impossible," i.e., that the external or material world both actually and necessarily does not exist (*CPR* B274).

Since Kant takes on the modally weaker form of skepticism, he has of course given himself a heavier burden of proof than would be required to refute Berkeley, since it is always harder to show that something is impossible than to show merely that its denial is actual or possible. But on the other hand, if Kant can show that problematic idealism is false, then that will also suffice to show that dogmatic idealism is false, and more generally that "material idealism"--which is the inclusive disjunction of problematic and dogmatic idealism (*CPR* B274)--is false. So the RI, if sound, will kill three skeptical birds with one argumentative stone.

*(5.3) A step-by-step reconstruction of the RI*

Let's now look at the RI in detail. For each step I will offer a short commentary. Then in the next section I will develop some criticisms of the overall argument.

> (1) "I am conscious of my existence as determined in time" (*CPR* B276).

*Commentary on step 1*. Obviously, this step can be directly compared & contrasted with Descartes's *cogito*: Necessarily, "I am, I exist" is true whenever I say or think it. Kant begins with what he elsewhere in the first *Critique* calls "empirical apperception" (*CPR* A107). Empirical apperception is empirical self-consciousness, or empirical reflective consciousness. So what Kant is saying here is that I have an empirical reflective consciousness *of* myself, as I consciously exist in "inner sense." Inner sense for Kant is the subject's intuitional awareness of a temporal succession of representational contents (*CPR* A22/B37, A107, B152-155, A357-359, A361-363, B420, B422-423 n.). Intuitional awareness, in turn, is (i) immediate or directly referential, (ii) sense-related, (iii) singular, (iv) object-dependent, and (v) logically prior to thought or nonconceptual (*CPR* A19/B33, A51/B75, B132, B146-147, A320/B377) (*PAFM* 4: 281-282). Occasionally in the first *Critique* Kant confuses inner sense and empirical apperception by calling them both "consciousness." But when he is being careful, we can see that he invokes a distinction between (i) a first-order unreflective reflexive consciousness of the phenomenal contents (whether objectively representational or merely sensory) of one's own mental state, and (ii) a second-order reflective consciousness of first-order consciousness. In one of the *Reflexionen* and in the *Prolegomena* he says this of inner sense:

> (The inner sense) Consciousness is the intuition of its self. (*R* 5049; 18: 72)
>
> [The ego] is nothing more than the feeling of an existence without the slightest concept and is only the representation of that to which all thinking stands in relation. (*Prol* 4: 334 n.)

By contrast, he says of empirical apperception that it is "**one** consciousness of myself" through which "I can say of all perceptions that I am conscious of them" (*CPR* A122). And in the *Anthropology* he distinguishes usefully between what he calls "taking notice of oneself" (*das Bemerken*), that is, an unreflective reflexive consciousness of oneself in inner sense at a given time, as opposed to "observing oneself" (*Beobachten*) (*Anthropology* 7: 132), that is, the introspective function of empirical apperception, which is repeatable over time and directly accessible via memory.

This difference between two levels of consciousness is crucial to Kant's argument against problematic idealism. To use some non-Kantian terminology borrowed from William James and Thomas Nagel, inner sense is both a "stream of consciousness" and also captures "what it is like to be for an organism": inner sense is a constantly-changing succession of unreflectively reflexive egocentric phenomenal states in a human or nonhuman animal cognizer. In other words, inner sense is the *phenomenal consciousness* of an animal cognizer. Empirical apperception, by contrast, is a second-order judgmental consciousness of myself as a singular or individuated first-order stream of unreflectively reflexive representations. The propositional element in empirical apperception makes it imperative that we further distinguish it from what Kant variously calls "pure apperception," "transcendental apperception," or "the original synthetic unity of apperception." This is an a priori or empirically underdetermined, spontaneous (i.e., unconditioned or unprecedented, creative), innate capacity for anonymous content-unification and for propositional and conceptual self-representation in general: more precisely, it is a universal capacity for attaching the cognitive prefix "I think" to any concept-involving representational content of the mind whatsoever (*CPR* B131-139, A341-348/B399-406). Empirical apperception, which presupposes transcendental apperception, is perhaps best regarded as the realization of that innate spontaneous capacity under concrete empirical conditions. Through empirical apperception, by carrying out an introspective judgment, I become conscious of my own first-order consciousness *as* constituting a determinate conscious human individual: "**I**, as a thinking being, am an object of inner sense, and am called 'soul'" (*CPR* A342/B400).

Kant's idea in this first step, then, is that even the most refractory skeptic would have to allow for the bare fact of such empirical introspection. To deny it would entail either (i) that we are always unconscious, or (ii) that even if we are sometimes conscious, then we are never conscious of our own consciousness ("meta-conscious"), or (iii) that even if we are sometimes meta-conscious, then we are never able to make first person psychological reports. There may well be living creatures that are always unconscious (e.g., humans in persistent vegetative states), or animals that have consciousness without meta-consciousness (e.g., newborn human infants and cats), or animals who have meta-consciousness without the capacity for introspective judgment (e.g., human toddlers and adult apes): but these are not creatures sharing our rational human cognitive constitution.

> (2) "All determination in time presupposes something **persistent** in perception" (*CPR* B276).

*Commentary on step 2*. For Kant, to "determine" something *X*, is either (necessarily or contingently) to ascribe or apply some definite attribute (a quality or property) to *X*, or to show how *X* enters (necessarily or contingently) as a relatum into some definite relation, and thereby takes on the attribute of belonging to that relation, or to show how *X* (necessarily or contingently) supports some definite relation. That all time-determination presupposes "that which persists," is a direct consequence of the arguments given by Kant in support of the first Analogy of Experience, the "principle of the permanence of substance" (*CPR* A182-189/B224-232). In the first Analogy Kant asserts that

> that which persists, in relation to which alone all temporal relations of appearances can be determined, is substance in the appearance, i.e., the real in the appearance, which as the substratum of all change always remains the same. (*CPR* B225)

The rationale behind this is the plausible thought that every change of attributes or relations in time requires something which remains the same throughout those changes. Now when we apply Kant's reasoning to strictly psychological phenomena, it grounds the conclusion that every determinate sequence of successive changes of conscious mental contents in time requires some or another unchanging substratum (something which persists) to which those changes are directly ascribed or applied. We need not, for our purposes, accept Kant's questionable further thesis--which seems to reflect a quantifier shift fallacy--to the effect that therefore there exists one and only one unchanging substratum to which every change of attributes or relations whatsoever is ascribed or applied, in order to buy into his original point. Nor need we, for our purposes, accept his questionable assumption that the unchanging substratum is either absolutely or even sempiternally persistent, rather than only relatively or temporarily persistent.

In any case, the crucial point Kant is driving at in step 2 has to do with psychological "determination in time." This phrase could be read as referring merely to the application of temporal predicates to my experiences. But I think that by using this phrase Kant is instead invoking something slightly stronger than this, namely, the *individuating determination* of my stream of experiences. This seems to be clearly implied by his use of the unusual phrase "my existence (*meines Daseins*) as determined in time"--as opposed to, say, "my experiences as determined in time"--and by his telling remark in the B Preface footnote to the effect that

> this consciousness of my existence in time is thus *bound up identically* (*identisch verbunden*) with the consciousness of a relation to something outside of me (*CPR* Bxl, emphasis added).

Kant's idea is that if I am to exist in inner sense as a constantly changing yet individuated stream of consciousness, and as an object of empirical apperception, then that stream must be essentially discriminable or uniquely identifiable, in the sense that it is distinguishable from any other such flow. To individuate my stream of consciousness is to confer upon all the separate elements of that stream--sensations, conceptions, images, judgments, etc.--a contingent yet *particular* ordering. This ordering activity occurs

primarily through what Kant calls "the synthesis of reproduction" (*CPR* A100-102), which I think is best construed as our cognitive capacity for (short-term, long-term, semantic, episodic, and procedural) memory. In any case, what reproductive synthesis does is to convert that otherwise undifferentiated stream of mental contents into a single personal history or autobiography, whereby my inner life takes on a definite psychological shape or profile. Now according to Kant the individuating determination (through reproductive synthesis) of any such flow of changing mental contents, requires a relatively fixed underpinning or matrix, that is, a psychological persistent factor which "as the substratum of all [psychological] change always remains the same."

It is hard to know precisely what Kant means by this, but I think that an analogy taken from physical geography is quite illuminating. A given river can be individuated only in relation to a spatially fixed material underpinning or matrix that includes its banks and riverbed, its beginnings and its terminus, and more generally the total path or locus it follows in getting from one end of the river to the other. Let us call this total path or locus its "geophysical route." The Mississippi, for example, flows south along a certain route from northern Minnesota to the Gulf of Mexico, and could not be the self-same river unless it did so. Since the actual quantity of water in the Mississippi at any given time is always changing and running off into the Gulf of Mexico, the Mississippi would then seem to uniquely defined by three physical factors: first, the fact that it is always water that is flowing in it, and not (say) beer or gasoline; second, the actual history of all the water that has already flowed through it over the years; and third, its route. This geophysical route can of course vary slightly within certain parameters, due to flooding or erosion, but those defining parameters continue to exist in a fixed way all the same. Like water in the Mississippi, which is always changing and running off into the Gulf, the contents of my stream of consciousness are always changing and running off into the past. So, analogously, my own individual psychological life would seem to be uniquely defined and distinguished from all other such "streams of consciousness," or conscious psychological processes, by three analogous psychological factors: first, the fact that only conscious human sensations, memories, concepts, etc., are flowing in it, and not (say) either non-sensory or "intellectual" intuitions or completely alien sorts of sense perceptions (*CPR* B71-72); second, the actual history of the various conscious mental contents that have already occurred in my psychological life; and third, its "psychological route": a fixed underpinning or matrix that remains invariant in relation to the constantly changing flow of my sense-qualities and representations in time. All the psychological changes in my inner life must be changes *of*, or changes ascribed or applied *to*, this particular fixed something, which in turn functions as a source of unity for my otherwise ever-changing stream of consciousness. But just as a river cannot be individuated without its geophysical route (its underlying geophysical substratum or defining parameters--that which geophysically persists in relation to it), so too the individuation of my stream of consciousness requires a psychological route (its underlying psychological substratum or defining parameters--that which psychologically persists in relation to it). And also by analogy we can predict, as in the case of the river's geophysical route, that small variations within my individuating psychological determining substratum will also be permissible, so long as they always remain within certain fixed parameters.

> (3) "But this persisting element cannot be an intuition in me. For all the determining grounds of my existence that can be encountered in me are representations, and as such they themselves need something persisting distinct from them, in relation to which their change, and thus my existence in the time in which they change, can be determined" (*CPR* Bxxxix n.).

*Commentary on step 3*. This is the first of the two most crucial steps of the proof. From step 2 we know that every changing conscious individuating determination of myself in time presupposes something that persists, in relation to which I can uniquely determine the conscious stream of contents in my inner sense. But this persistent thing must be outside my own conscious mental states, and not merely inside me. For if it were merely inside me, it would then *belong* to the ever-changing stream of consciousness, and so could not provide a uniquely determining substratum for the mental modifications I experience directly. Hence it must fall outside the proper domain of my inner sense, that is, outside the series of first-order phenomenally conscious representational states that I am directly aware of via my second-order introspective consciousness or empirical apperception.

Now at this point it might well occur to us that something else in inner sense might plausibly function as the "persisting element," namely the *form* of inner sense, as opposed to its contents. And indeed according to Kant the form of inner sense always remains the same, since it is invariantly presupposed by any actual or possible inner experience (*CPR* A22-23/B37, A31/B46). But the form of inner sense is nothing other than the representation of *time*. And it is incoherent to suggest that either the representation of time or time itself could be a persisting or enduring thing *in* time. Either the representation of time or time itself is a necessary formal precondition for the series of changes in my stream of consciousness. Now to hold that the representation of time occurs in time, would be to confuse properties of the psychological *vehicle* of a representation (which does indeed occur in inner sense, hence in time) with semantic properties of its representational *content*. And also it would plainly be conceptually incoherent to hold that time itself occurs in time. So neither the representation of time nor time itself could also function as an enduring substance or substratum *to which* my changing conscious representational states are ascribed or applied. Hence nothing in either the content or the form of inner sense can function as the persistent element or substratum that is required for the individuation or unique determination of my stream of consciousness.

> (4) "Thus the perception of this persistent thing is possible only through a **thing** outside me and not through the mere **representation** of a thing outside me. Consequently the determination of my existence in time is possible only by means of the existence of actual things that I perceive outside myself" (*CPR* B275-276).

*Commentary on step 4*. This is the second crucial step. In order uniquely to determine or individuate my own successive existence in time I must presuppose the existence of outer things perceptually represented by me, and not merely the existence of my internally flowing conscious representations of those outer things. The radical nature of what Kant

is saying here cannot be overemphasized. He is saying that any individuating temporally determinate introspective awareness of myself is necessarily also a direct nonconceptual veridical representation of some real material thing existing *outside* my stream of conscious experiences and *at a distance from me in space*. The latter factor is especially to be noted. In the Transcendental Aesthetic Kant argues that "in order for certain sensations be referred to something outside of me" they must be referred to "something in another place in space from that in which I find myself" (*CPR* A23/B38). My unique individuality at the level of first-order phenomenal consciousness is therefore *inherited* from the world of distal physical objects. In this way, despite the fact that via empirical apperception in a loose and everyday sense we introspect "an object of inner sense [which is] called 'soul'," there is strictly speaking for Kant no independent "inner object" of inner sense:

> inner sense, by means of which the mind intuits itself, or its inner state, gives, to be sure, no intuition of the soul itself, as an object. (*CPR* A22/B37)

That is, what is truly inner for Kant cannot be reified: it is neither a noumenal inner thing nor a phenomenal inner thing. And as he puts it in the first Note concerning the Refutation, "inner experience itself is … only mediate and possible only through outer experience" (*CPR* B277). So insofar as I am aware of myself in empirical apperception as a uniquely determined psychological being, then I must directly and nonconceptually ascribe or apply the changing contents of my mental states to the objective furniture of the distal material world.

This crucial point needs further emphasis. Far from having the problem of escaping from a "Cartesian box" into the outer world, *Kant's* problem in the first *Critique* is instead that of distinguishing himself from various surrounding material objects in the outer world! This problem comes out clearly if we put it in non-Kantian terminology, this time borrowed from G.E. Moore and Jean-Paul Sartre. Kant's view of inner sense in the Refutation comes very close to an amazing doctrine defended by Moore in his 1903 essay, "The Refutation of Idealism," a doctrine which he calls the "transparency of consciousness":

> [W]hen we refer to introspection and try to discover what the sensation of blue is, it is very easy to suppose that we have before us only a single term. The term 'blue' is easy enough to distinguish, but that other element which I have called 'consciousness'... is extremely difficult to fix. That many people fail to distinguish it at all is sufficiently shown by the fact that there are materialists. And, in general, that which makes the sensation of blue a mental fact seems to escape us: it seems, if I may use a metaphor; to be transparent --we look through it and see nothing but the blue. We may be convinced that there *is something* but *what* it is no philosopher, I think, has yet clearly recognized.[1]

Here consciousness is *not an inner thing*, in spades: instead it is nothing but a noetic searchlight on outer things. Later, in the 1930s (but presumably without having read

[1] Moore, "The Refutation of Idealism," p. 37.

Moore), Sartre pushes this idea of transparency one step further and describes something he calls "the transcendence of the ego."[2] Sartre's idea is that the ego receives its first-order unreflective reflexive subjective unity solely and directly from the outer things it is transparently conscious of. So this is not merely *content*-externalism: it is also *phenomenal consciousness*-externalism. The conscious mind is much "out there in the world" as it is "in here." *Ego*centricity is representational *ec*centricity. Phenomenal consciousness is nothing but *consciousness-of* or intentionality. Essentially the same view is held by contemporary defenders of the "first-order representational theory of consciousness,"[3] and, even more radically, by defenders of the "extended conscious mind thesis."[4]

All of these later affinities shows how radical and philosophically prescient Kant's doctrine really is. Add the Sartrean transcendence of the ego, the first-order representational theory of consciousness, and the extended conscious mind thesis to the Moorean transparency of consciousness, and you have, in effect, Kant's doctrine in step 4.

> (5) Now consciousness [of my existence] in time is necessarily bound up with consciousness of the [condition of the] possibility of this time-determination. Therefore it is also necessarily combined with the existence of the things outside me, as the condition of time-determination. (*CPR* B276)

*Commentary on step 5*. This step is fairly straightforward. Insofar as I am empirically self-aware, and individuate myself in time, I must also be directly consciously aware of this act of time-determination. Hence I must also be directly consciously aware of the existence of a distal persistent thing outside me that individuates me, since this is the necessary condition of time-determination.

> (6) "I.e., the consciousness of my existence is at the same time (*zugleich*) an immediate consciousness of the existence of other things outside me" (*CPR* B276).

*Commentary on step 6*. This adds a crucial factor to step 5. The "immediate consciousness of" something is a direct veridical consciousness of that thing. So Kant is saying that for any particular empirical apperception of myself as uniquely determined in inner sense, I am also *simultaneously* directly veridically perceptually aware, via outer intuition, of some existing or actual distal material object in space as the individuating substratum to which I ascribe or apply the changing conscious representational contents of my mind.

So to sum up the whole Refutation: Necessarily, if I am determinately aware of myself in empirical apperception, then I am also thereby at that very same moment directly veridically perceptually aware of some actual distal material object in space

---

[2] See Sartre, *The Transcendence of the Ego: An Existentialist Theory of Consciousness*.
[3] See Carruthers, "Natural Theories of Consciousness."
[4] See, e.g., Rockwell, *Neither Brain Nor Ghost*.

*(5.4) Criticism of the RI*

I now move on to critical objections. It seems to me that both steps 1 and 3 are acceptable, assuming the correctness of both Kant's philosophical psychology (of inner sense, outer sense, and apperception) and of the "weak" reading of the First Analogy that I proposed.

Yet critics of the Refutation often hold that the fundamental gap in the proof is to be found in step 3. Why, such critics ask, is it necessarily the case that the intuition of that which is persistent, be an intuition of something outside me? Why couldn't it instead be an intuition of some persistent thing *inside* me--that is, of some "thinking thing"? This option immediately fails, however, when we remember just what sort of intuition an inner intuition is:

> the determination of my existence can occur only in conformity with the form of inner sense, according to the particular way in which the manifold that I combine is given in inner intuition, and therefore I have **no cognition** of myself **as I am** but merely as I **appear** to myself. (*CPR* B157-158)

In other words, in empirical self-consciousness I am not directly aware of myself as a Cartesian ego-in-itself. That would require an "intellectual intuition" (*CPR* B72) of myself. But as a finite human cognizer who is not merely rational but also an animal, my intuition is strictly sensory and not intellectual: in inner sense, I am directly aware only of the phenomenal flotsam and jetsam of consciousness. That is, in inner sense, I am directly aware only of *my phenomenally conscious states and their phenomenal contents (whether objectively representational or not)*, not of some deeper noumenal substratum of those phenomenally conscious states.

But even if steps 1 to 3 hold up tolerably well, nevertheless in my opinion steps 4 to 6 as they stand are highly questionable. Here is a worry about step 4. Even granting that my empirical self-consciousness of my stream of consciousness in inner sense requires an outer intuition of something persistent, nevertheless it does not seem to follow that inner intuition in general requires any outer intuition of actually existing distal material objects in space. For so long as *space alone, as an object,* can be represented by means of a "pure intuition" or "formal intuition," as Kant explicitly argues in the Transcendental Aesthetic and again later in the B edition's Transcendental Deduction of the Pure Concepts of the Understanding (*CPR* B160 n.), then *that* seems sufficient to meet the requirement that there be a single persistent thing over against me, to which I must intuitionally refer and ascribe my ever-changing conscious inner states. And the pure or formal intuition of space does not logically require the existence of any distal material objects *in* space. Kant says explicitly that "one can never represent that there is no space, although one can very well think that there are no objects to be encountered in it" (*CPR* A24/B38-39).
What is the pure or formal intuition of space? Five features are at least individually necessary for it. First, the pure or formal intuition of space is *a non-empirical presupposition of all empirical intuitions of objects in space*: "[this representation of space] is a necessary representation, a priori, which is the ground of all outer intuitions"

(*CPR* A24/B39)[5]. Second, the pure or formal intuition of space is *nonconceptual*: "[this representation of] space is not a discursive or, as is said, general concept of relations of things in general, but a pure intuition" (*CPR* A25/B39). Third, the pure or formal intuition of space represents space as a *unique object*: "one can represent only a single space" (*CPR* A25/B39). Fourth, the pure or formal intuition of space represents space as a *unified structured manifold*:

> Space and time are represented a priori not merely as **forms** of sensible intuition, but as **intuitions** themselves (which contain a manifold), and thus with the determination of the **unity** of this manifold... Space, represented as **object** (as is really required in geometry), contains more than the mere form of intuition, namely the **putting-together** (*Zusammenfassung*) of the manifold. (*CPR* B160, text and note combined)

Fifth and finally, the pure or formal intuition of space represents space as an *infinite totality*: "space is represented as a given infinite magnitude" (*CPR* A25/B40). For our purposes, we need not unpack Kant's extremely interesting doctrine of pure spatial representation any further. My point right now is only (a) that the pure or formal intuition of space is an a necessary priori nonconceptual representation having a referent--i.e., space itself--which is represented as a unique unified structured manifold and an infinite totality, and (b) that this unique unified structured manifold and infinite totality has not been ruled out as the persisting element Kant needs in order to meet the requirement of step 2. It is incoherent to suppose that time itself might function as a persisting entity in time, but *not* incoherent to suppose that space itself might function as a persisting entity in time. And if space itself can meet that requirement, then since Kant explicitly says that space can be represented as empty of all material objects (*CPR* A24/B38-39, A291/B347), it follows that Kant has not ruled out the possibility that I ascribe or apply my changing mental states directly to empty space itself.

Just in case my objection to step 4 is not convincing however, here is another objection, this time to step 6. Even if we grant what I think we should *not* grant--namely, that my self-conscious awareness of my stream of consciousness in inner sense entails that I have some direct outer experiences of actual distal material objects in space--it does not seem to follow from that, that on every occasion of self-awareness I must be *simultaneously* directly correctly perceptually aware of a distal material external object. What about dreams and hallucinations? In Note 3 of the Refutation, Kant himself admits that

> from the fact that the existence of outer objects is required for the possibility of a determinate consciousness of our self it does not follow that every intuitive mental representation of outer things includes *at the same time* (*zugleich*) their

[5] There is, however, an interpretive subtlety here: the pure or formal intuition of space is *a* presupposition of all empirical intuitions of objects in space, but it is not *the* presupposition: only the bare intuitional representation of space, the form of our outer intuition, is uniquely presupposed. The pure or formal intuition of space entails the form of outer intuition, but the form of our outer intuition does not entail the pure or formal intuition of space. For the important distinction between *formal intuitions* and *forms of intuition*, see *CPR* 160.

> existence, for that may very well be the mere effect of the imagination (in dreams as well as in delusions). (*CPR* B278, emphasis added)

So Kant is certainly aware of the dream problem, and he must then implicitly grant that step 6 as it stands, with the simultaneity condition, is false.

Where does this leave us? By virtue of his admission of the dream problem, Kant has implicitly admitted that not every self-conscious awareness of my own uniquely determined conscious existence in time entails a simultaneous direct correct perception of a distal external object. So since Kant is certainly no fool, it seems to me that his concluding step 6 is most charitably and plausibly interpreted as saying the same as these two alternative formulations of the conclusion of the Refutation:

> The proof that is demanded must therefore establish that we have **experience** and not merely **imagination** of outer things, which cannot be accomplished unless one can prove that even our **inner experience**, undoubted by Descartes, is possible only under the presupposition of outer experience. (*CPR* B275)

> The mere, but empirically determined, consciousness of my own existence proves the existence of objects in space outside me. (*CPR* B275)

Taken together, these formulations say that my having a self-conscious awareness of my individuated stream of inner consciousness entails my also having some direct correct perceptions of distal material objects in space. Even so, there is no necessity that I have a direct correct outer perception of a distal material object at the very *same* time that I am in one of these self-conscious states, so long as I also have some direct correct outer perceptions of distal spatial objects at *other* times. The simultaneity condition can be dropped.

This charitable interpretation is backed up by a footnote appended to the Refutation, which says that even when we are dreaming or hallucinating, and merely *imagining* space, it is presupposed that we already have an outer sense through which we do sometimes get direct correct perceptual access to outer material things:

> In order for us even to imagine something as external, i.e., to exhibit it to sense in intuition, we must already have an outer sense, and by this means immediately distinguish the mere receptivity of an outer intuition from the spontaneity that characterizes every imagining. For even merely to imagine an outer sense would annihilate the faculty of intuition, which is to be determined through the power of imagination. (*CPR* B276-277 n.)

In other words, space cannot be even imagined without our already having a capacity, sometimes actualized or realized, for directly and correctly perceiving or empirically intuiting distal material objects in space. And this reading is in turn backed up by two other texts. First, in the *Anthropology* Kant notes that imagination "cannot bring forth a representation that was *never* given to the power of sense; we can always trace the

material of its representations" (*A* 7: 168). And second, in one of the *Reflexionen* he is even more explicit:

> Dreams can represent to us things as outer, which are not there; however, we would never be able to represent something as outer in dreams, if these forms were not given to us through outer things. (*R* 5399; 18:172)

So it seems to be Kant's view that even our capacity for "imagination of something as external" is parasitic upon some direct correct outer sense perceptions of distal material objects, at some time or another. If he is right, and if we interpret step 6 in such a way as not to commit Kant to the implausible thesis that every individuating act of empirical self-consciousness requires a simultaneous direct correct perceptual awareness of a distal material object, then he in fact avoids the dream problem.

But even so, *is* he right? Well it seems likely that it is empirically true, as a fact in cognitive psychology, that normal image-construction and manipulation is originally funded by direct correct sense-perception of distal material objects. But is it *necessarily true* for creatures like us? Surely we can conceive of a possible human being whose empirical imagination-content is entirely funded by some source other than direct correct sense-perception of distal material objects. Or, to put it another way, if a creature had been born with or developed a capacity for imagining external things that was entirely empirically funded in some non-standard way which was systematically insulated from direct correct perceptual contact with the distal outer world--suppose, e.g., that someone was fitted from birth with a microscopically thin computer-driven "virtual reality suit" covering her entire body, or that (as in *The Matrix*) she was born hooked up to the Matrix, so that again all her perceptions were in fact false digital images--would she thereby fail to be *one of us*? I think not. Such a human cognizer, cocooned inside her all-encompassing perceptual prosthetic, or unconsciously supplied with a massively complex and detailed but still phoney digital image of her actual surrounding world, would certainly be odd, and perhaps somehow slightly cognitively handicapped (or perhaps not even slightly handicapped, in light of the actual empirical fact of "neural plasticity," as manifest in the effective neural and behavioral adaptation of actual human cognizers to inverting lense goggles, Tactile-Visual Substitution Systems, etc.): but she would certainly nevertheless, I think, still fully share our human cognitive constitution. So Kant's thesis of the dependency of imagination on correct perception is false, and the Refutation is therefore unsound.

*(5.5) Or is the RI sound after all?*

I would like now to shift philosophical gears, and move from the negative evaluation of Kant's Refutation given at the end of the last section, towards a more positive evaluation. Indeed I think that the Refutation implicitly contains something of real and even fundamental philosophical significance. Suppose, now, that steps 4 and 6 are indeed fallacious as they stand. Nevertheless it seems to me the case that Kant *has* indeed proved *this* weaker thesis:

> [I]nner experience in general is possible only through outer experience in general. (*CPR* B278-279)

My reasoning is this. Crucial to this thesis are two phrases: 'inner experience in general' and 'outer experience in general'. I want to read 'inner experience in general' as meaning "to have a self-conscious awareness of myself in inner sense," and I want to read 'outer experience in as meaning "to have an actual outer sense." That is, I want to read the thesis as saying:

> To have a self-conscious awareness of myself in inner sense is possible only through my also having an actual outer sense,

*not*:

> Each and every inner self-conscious experience of a given mental state of my own is possible only through some direct correct outer perception of an actual distal material object in space.

That my proposed reading is at least plausibly Kantian is also well-supported by a passage in the *Reflexionen*:

> The question, whether something is outside of me, is just the same as to ask, whether I represent to myself an actual space. For this is outside of me. (*R* 5400; 18: 172)

Otherwise put, I want to distinguish quite sharply between three distinct meanings of the phrase '*X* is outside my own conscious mental state':

> (1) *X* is a mind-independent substance;
> ` (2) *X* is a material object in another part of space from that in which I am located = *X* is a distal material object; and
> (3) *X* is necessarily spatial in character.

What I would want to argue on Kant's behalf is that in order to refute (*), or skeptical idealism, it is necessary only to prove that *I myself* satisfy (3), not to prove that *something else* satisfies (1) or (2). The issue on the table right now is whether a dreamer or hallucinator could have a capacity for imagining external things without having an actual outer sense. Again, I think not. That is, I would want to argue on Kant's behalf that a capacity for imagining external things, even in dreams or hallucinations, is not possible without an actual outer sensibility. Given the general definition of outer sense as what stands in an immediate or a mediate relationship to my body, this in turn is equivalent to the thesis that a subject's capacity for imagining external things is not possible without *her possessing a body in space*.

Reconstructed in this way, the RI conforms very smoothly to Kant's original idea that all conscious changes in inner sense are necessarily immediately ascribed to an actual

spatially existing persisting thing or substratum. For we can now see that the most natural way of reading this is as saying that necessarily the contents of my own consciousness literally belong to my own body. The big problem with steps 4 and 6 in the original argument was the assumption that the external substratum in question was *distal*, not *proximal*, in relation to the self-conscious subject.

But suppose that the external substratum Kant is talking about is strictly proximal: suppose that the external substratum is none other than my own body. Then what Kant is actually saying in the RI is that in order to individuate myself psychologically and as a unique member of my own species, then I must ascribe each of my mental states directly to my own body in space. In other words, the ascription of my mental states to my own body individuates my mental states by *locating them uniquely*. And this seems to me correct. So the RI seems to me to be a sound demonstration, against the Cartesian skeptic and also the Berkelyan idealist, of the existence of my own material body in space.

That is, Kant is saying in the RI that necessarily, if I am self-consciously aware of myself as an individuated stream of consciousness in inner sense, then my body also exists as a uniquely located material being in space.

And this seems to me to be **objectively true**.

(Notice, however, that it doesn't follow from the soundness of of this version of Kant's RI *that I know specifically what my body is like, e.g., whether I really have hands, etc*. So it doesn't support Moore's attempted refutation of problematic or skeptical idealism from common sense in "A Defense of Common Sense." According to Kant's RI, I know only *that* my body exists as a uniquely located material being in space.)

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 6

Covers: CPR: 193-218 = A50-83/B74-116 (The Idea of a Transcendental Logic, On the Clue to the Discovery of all Pure Concepts of the Understanding).

## Concepts, Logic, & Judgment: The Metaphysical Deduction of the Categories

*(6.1) From the Transcendental Aesthetic to the Transcendental Logic.*

One of the central aims of the Transcendental Aesthetic was to show that sensibility has two necessary a priori forms--r-space and r-time--which are necessarily applicable to all sensory appearances, i.e., to the objects accessible through empirical intuition.

In the Introduction to the Transcendental Logic Kant wants to prove a parallel result for concepts and judgments: that is, he wants to show that there exist pure a priori concepts of the understanding (henceforth PCUs, a.k.a., "categories") that are necessarily applicable to objects of intuition in general, i.e., to anything whatsoever that can be represented via intuitions, concepts, and judgments. This argument, which K. later labels "the metaphysical deduction" of the PCUs (CPR: B159), in fact has two phases: (I) a "logical" phase, and (II) an "objectual" phase.

That is, the first part of the Metaphysical Deduction (MD) ties the PCUs to *logic*, and the second phase ties the PCUs to the cognition of *objects*. For clarity's sake, I'll spell out each phase separately.

*(6.2) Four Background Doctrines.*

Neither phase of the Metaphysical Deduction makes sense, however, unless it is placed against the background of the following four doctrines:

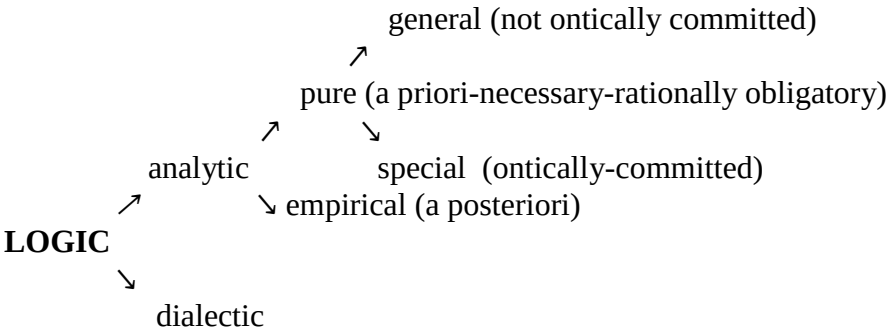
(A) the mutual irreducibility and complementarity of intuitions and concepts;
(B) Kant's logical theory;
(C) Kant's theory of truth; and
(D) Kant's theory of judgment.

All four of these doctrines are covered in an article of mine, "Kant's Theory of Judgment," in *The Stanford Encyclopedia of Philosophy*, which can be found at the following URL: = <http://plato.stanford.edu/entries/kant-judgment/>.

Here is the correlation of the four doctrines mentioned above to corresponding sections in the *Stanford Encyclopedia* article:

(A) → section 1.3.1
(B) → section 2.1.1 – 2.1.2
(C) → section 1.3
(D) → section 1.2.

*(6.3) Kant's logical theory: an outline*




(1) **LOGIC** = the science of the necessary rules (i.e., laws) of (rational) thinking.

(2) Analytic logic = the logic of (necessary) truth, valid consequence, and soundness.

NB. 'Analytic' in this context does *not* mean the same as 'analytic' in the context of the analytic-synthetic distinction (see (4) below).

(3) Dialectical logic = the logic of (necessary) falsity—including paradox, fallacy, & illusion.

(4) Analytic pure general logic = monadic logic = sentential logic + the logic of one-place predicates & quantification into them = (roughly) syllogistic logic.

NB. Monadic logic is consistent, sound, complete, & decidable, & seems to capture all our basic beliefs about analyticity. No other logic is so *rationally & systematically formally perfect.* So pure general logic is, arguably, the *a priori normative core or essence* of all analytic logic.

(5) Analytic pure special logic = any pure logic with special ontic commitment, e.g., to all & only the objects of possible human experience, i.e., *transcendental logic*.

NB.1 Contemporary "classical" elementary logic would *also* count as an analytic pure special logic, although it doesn't have ontic commitment to all & only the objects of possible human experience. So there can be many different kinds of transcendental logic.

NB.2 Every analytic pure special logic, i.e., every transcendental logic, is *synthetic a priori*.
(6) Analytic empirical logic = the empirical psychology of logic.

NB. So Kant's theory of pure (general or special) logic is *not* psychologistic.

*(6.4) The Logical Phase of the Metaphysical Deduction*

Prove: The PCUs are the same as the logical forms of pure general logic.

(1) A PCU is a concept that is intrinsic to the understanding itself: hence the PCUs are pure, non-intuitional, basic, and the roster of such concepts is complete (CPR: A64-65/B89-90).

(2) The transcendental clue to the discovery of the PCUs is the general logical use of the understanding (CPR: A67/B92).

(3) Implicit Premise: All cognition in the narrow sense (which I'll indicate from now on by adding a subscript to 'cognition', i.e., '$\text{cognition}_{ns}$') occurs through the combination of intuitions and concepts, and only through such a combination. And all $\text{cognition}_{ns}$ by means of concepts is governed by pure general logic, the science of the necessary rules of the understanding in general (= the a priori science of the laws of thought) (CPR: A50-52/B74-76).

(4) Concepts are "functions of unity" or rules (of synthesis) for bringing the various data supplied by intuition into a unity (CPR: A68/B93).

(5) Now the only use, or application, of a concept is to judge by means of it (CPR: A68/B93).

(6) The basic sort of judgment, namely a subject-predicate judgment, is a representation of a representation of an object. That is, it is a representation that consists in the application of a conceptual representation (the predicate) to another representation (the subject), which is either a concept or an intuition, and which in turn relates indirectly or directly to objects of intuition (CPR: A68-69/B92-93).

(7) Because judgments are representations of the sort described in step (6), they are higher-order functions of unity among our conceptual and intuitional representations (CPR: A69/B93-94).

(8) All acts of the understanding are to be traced back to judgments (CPR: A69/B94).

(9) Hence the basic functions of the understanding can be exhaustively determined by exhaustively determining the logical functions of unity in judgments = the table of judgments (CPR: A70-76/B95-101).

(10) Since the table of judgments yields an exhaustive list of the logical forms governing all acts of the understanding, and since the logic of the understanding is both pure and general, its logical forms must also be the PCUs. QED

*(6.5) The Objectual Phase of the Metaphysical Deduction*

Prove: The PCUs, which (by the Logical Phase) are the logical forms of pure general logic, are also necessary a priori conditions of the possibility of the cognition$_{ns}$ of objects of intuition in general, and thus metaphysical "categories" in (roughly) Aristotle's sense of that term.

(1) All cognitions$_{ns}$ rest on synthesis: the unification of many diverse representations into unified representations, by means of the power of imagination (CPR: A76-78/B102-104).

(2) The PCUs are functions of pure synthesis, i.e., grounds of the unity of pure synthesis, i.e., basic rules of pure synthesis (CPR: A78/B104).

(3) The cognition$_{ns}$ of an object arises in the following way: a manifold is given in intuition; this manifold is synthesized in the imagination; and the synthesis of the manifold is brought under concepts (CPR: A78-79/B104).

(4) But all cognitions$_{ns}$ are judgments. (From the Logical Phase.)

(5) And all judgments presuppose the PCUs as the logical forms of pure general logic. (From the Logical Phase.)

(6) Since the function which gives unity to the synthesis in a judgment is the same as the function which gives unity to the synthesis which represents objects in intuition--i.e., a PCU--it follows that the logical forms of pure general logic, i.e., the PCUs, are also necessary a priori conditions of the possibility of the cognition$_{ns}$ of objects of intuition in general, and thus metaphysical "categories" in (roughly) Aristotle's sense of that term (CPR: A79/B104-105). QED

*(6.6) Some critical worries.*

(1) To what extent, if any, is K.s argument in the MD of the PCUs vitiated by the fact that his formal logic borrows heavily from Aristotelian/Scholastic logic, and by the further—notorious—fact that he believed logic in this sense to be a finished and complete body of doctrine? More generally, what are the philosophical implications of the differences between K.'s formal logic and modern formal (= symbolic, mathematical) logic?

(For an attempt to answer these two questions, see my "Kant's Theory of Judgment," section 2.1.2.)

(2) To what extent, if any, is K.'s argument in the MD vitiated by the fact that his conception of pure general logic is explicitly cognitivist (= mental representationalism + faculty-innatism)? In particular, does K.'s logical cognitivism imply logical psychologism (= the thesis that logic is reducible to empirical psychology)?
(3) How does K. actually get from the table of judgments to the table of categories? Is he just making this up as he goes along, or is there some deeper rationale at work?

[Anticipatory answer to (3): We can ask *how does the necessary truth, apriority, & categorical normativity of pure general logic place constraints on the way the world has to be?*]

(4) If K. has just proved that the PCUs are the necessary a priori conditions of the possibility of the $\text{cognition}_{ns}$ of objects in general, then why does he think he need a Transcendental Deduction of the PCUs? Or is the latter just philosophical obsessing and argument-overkill? In other words, what precisely is K.'s rationale for the transition from the MD of the PCUs to the TD of the PCUs?

[Anticipatory answer to (4): Only the TD of the PCUs explicitly and specifically restricts intuition to *our* sensibility, as structured by *our* formal representations of space and time.

For all we know from the Metaphysical Deduction alone, the categories might in fact apply *only* to objects delivered by intuitions governed by some wholly *alien* form of sensible intuition, e.g., the representation of *spime*. If so, then it is possible that there are *spimey* objects that don't fall under the Categories.

This "spimey object" worry, in turn, is analogous to another, more serious worry about the scope of the Categories that I'll cover in the next lecture, which I call *the bottom-up problem*, or *the problem of rogue objects*.]

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
**LECTURE 7**

Covers: CPR: 219-244 = A84-130 (Deduction of the Pure Concepts of the Understanding, A edition).

## The Transcendental Deduction of the Categories in the A Edition

*(7.1) From the Metaphysical Deduction to the Transcendental Deduction.*

You'll remember (see the notes for LECTURE 3) that Kant's argument in the Transcendental Aesthetic (TA) yields the thesis that sensibility has two necessary a priori subjective forms—(the representations of) space and time—that are necessarily applicable to all appearances, or empirical objects, given in empirical intuition. The TA also yields, by means of the argument for the transcendental ideality of space and time (together with the thesis of the intrinsicness of space and time), the thesis of transcendental idealism (TI): All objects of cognition are nothing but phenomena (= the intersubjectively shareable contents of our sensible representations), not things-in-themselves, and strictly conform to the transcendental faculties (or innate spontaneous capacities) of the human mind.

You'll also remember (see the notes for LECTURE 6) that in the Metaphysical Deduction (MD) of the pure concepts of the understanding (PCUs) Kant argues (i) that the PCUs are the same as the logical forms of pure general logic, and (ii) that the PCUs, which are the logical forms of pure general logic, are also necessarily applicable to objects of intuition in general, and thus metaphysical "categories" in (roughly) Aristotle's sense of that term.

Kant's A edition Transcendental Deduction of the PCUs or categories (for short, the A Deduction of the categories) presupposes the results of the TA and the MD of the PCUs.

Here is a little bit of relevant scholarly background information. Kant's term '*Deduktion*'is first and foremost an 18th century juridical or jurisprudential term, and *not* a bit of logical jargon: what we now call a "deduction" in the formal logical sense is best translated by the 18th century German logical term '*Schluss*', which roughly means the same as 'inference' or 'proof'. By contrast, a deduction in the 18th century juridical sense is the legal justification offered by a lawyer for someone's de facto possession of something (say, some piece of property).

So by the concept of a "transcendental deduction" of the categories Kant does not mean a formal logical deduction, but instead an argument justifying the philosophical legitimacy of our de facto possession of the PCUs. This is the same as to demonstrate the necessary applicability of the categories to objects of human experience (CPR: A84-92/B116-124). And thus the function of a transcendental deduction of the categories is to give an argument for the *objective validity* or empirical meaningfulness of the categories. This is to be contrasted with the "empirical deduction" of a concept, which shows its applicability or objective validity by factual means or empirical tests alone.

By Kant's own reckoning, the A Deduction falls into two parts: a "subjective deduction" and an "objective deduction" (see CPR: Axvi-xvii). Sadly however, Kant himself is not absolutely explicit as to just where these two sub-deductions occur.

Nevertheless, I will treat section 2 of the A Deduction (CPR: A95-114) as the subjective deduction, and section 3 (CPR: A115-130) as the objective deduction. These two parts of the A Deduction are preceded by section 1 (CPR: A84-A94/B116-129) which, with the exception of one paragraph, appears in both the A and B editions.

*(7.2) What Kant is Arguing in the A Deduction.*

In light of the TA and the MD of the PCUs, Kant wants to argue in the A Deduction (and again in the B Deduction too) that the categories are a priori necessary conditions of the possibility of all objects of experience.

(Notice however that the categories are not *the only* a priori necessary conditions of the possibility of all objects of experience, because (the representations of) space and time are *also* a priori necessary conditions of the possibility of all objects of experience.)

The *subjective deduction* yields this conclusion as the result of an analysis of the nature of our transcendental cognitive capacities alone (i.e., via "transcendental psychology"); and the *objective deduction* yields the same conclusion as a result of an analysis of the relation between those same capacities and the objects of experience.

Now one might ask: Why does Kant even *need* to carry out a transcendental deduction of the categories if he has already established the applicability of the categories to objects in general in the MD of the PCUs? And the answer is: a transcendental deduction of the categories must be undertaken for two basic reasons.

First, the MD of the PCUs does not explicitly restrict the categories to *appearances* (the undetermined objects of *our* empirical intuition) or *objects of experience* (the fully determined objects of *our* $\text{cognitions}_{ns}$ or empirical judgments

And second, unlike the objective validity of the a priori necessary subjective forms of of intuition, the objective validity of the categories (especially that of causality) cannot be transcendentally deduced from the fact of empirical intuition alone. For empirical intuition might, at least in principle, latch onto appearances that do not or cannot satisfy

all the conditions required for the applicability of concepts: "appearances can certainly be given in intuition without functions of the understanding" (CPR: A90/B122).***** So the objective validity of the categories must be deduced instead from the fact of empirical judgments about objects of experience.

****NB. Eventually, this point is going to create a fundamental worry about the soundness of the B Deduction, that I call the problem of The Gap = the worry that the scope of the Categories & concepts & the human understanding is narrower than the scope of pure intuitions, empirical intuitions, & human sensibility, hence there are actual or possible appearances beyond the Categories, hence the B Deduction fails. See Lecture 8, & also my essay, "Kant's Non-Conceptualism, Rogue Objects, and the Gap in the B Deduction," International Journal of Philosophical Studies 19 (2011): 397-413, available online at URL =*** <https://www.academia.edu/1018554/Kants_Non-Conceptualism_Rogue_Objects_and_the_Gap_in_the_B_Deduction>.

*(7.3) The Subjective Deduction.*

Prove: That the PCUs or categories are necessary a priori conditions of the possibility of all objects of experience.

(1) All experience of objects involves the contributions of several cognitive capacities or powers, and each of these capacities or powers has a "transcendental constitution." (CPR: A97)

(2) Synthesis is the basic act of mental processing, or combining several lower-level representations into a unitary, higher-level representational content; and all syntheses whasoever are produced by the power of imagination (see CPR: A77-78/B102-103). Insofar as a manifold of intuitive content is collected together and internally organized, this must occur by means of a *threefold synthesis*: (i) the synthesis of apprehension in intuition (that is, in empirical intuition and under the pure forms of intuition); (ii) the synthesis of reproduction in the imagination (that is, in the empirical imagination and under a "pure schema" of the transcendental imagination); and (iii) the synthesis of recognition in a concept (that is, in an empirical concept and under a pure concept or category). (CPR: A98-103)

(3) Now this threefold synthesis necessarily introduces a "unity of rule" into the manifold of sensory representational content in each empirical intuition, and also into the manifold of empirical intuitions, and makes possible the experience of a fully determined object. This unity of rule is also expressible as the concept of "the transcendental object = *X*" (CPR: A105-106).

(4) The concept of the transcendental object = *X*, in turn, is possible only if the synthesis occurs under the governance of a single but purely formal consciousness that expresses our innate capacity for self-awareness or self-consciousness. Such an innate capacity for self-consciousness is called "transcendental apperception." (CPR: 106-107)

(5) Experience as a whole must stand in a synthetic unity. And just as the forms of intuition are necessary a priori conditions of the synthetic unity of the manifold in and of the empirical intuitions given in our sense perceptions of appearances, so too the categories are a priori necessary conditions of the synthetic unity introduced into experience by means of empirical concepts. (CPR: A110-111)

(6) Now the ground of the synthetic unity of the manifold in and of the empirical intuitions given in our sense perceptions of appearances, is transcendental apperception; correspondingly, the ground of the synthetic unity of concepts is that very same transcendental apperception. (CPR: A111-112)

(7) Implicit premise: An object of experience is nothing but the rule-governed, synthesized manifold in and of the intuitions given in our sense perceptions of appearances. This follows from TI and the theory of synthesis.

(8) Therefore, since (a) the very same transcendental apperception grounds both the synthesis of concepts and also the synthesized manifold in and of the intuitions given in our sense perceptions of appearances, and since (b) the synthesis of concepts is necessarily governed a priori by the categories, then it follows (c) that the synthesis of the manifold in and of the empirical intuitions given in our sense perceptions of appearances is *also* necessarily governed a priori by the categories. (CPR: A112)

(9) That is: The PCUs or categories are necessary a priori conditions of the possibility of all objects of experience. QED

*(7.4) The Objective Deduction.*

Prove: That the PCUs or categories are necessary a priori conditions of the possibility of all objects of experience.

(1) The unity of apperception governs the unity of the manifold in and of the intuitions given in our sense perceptions of appearances. (CPR: A116-117)

(2) This unity involves a synthesis which introduces a special form or structure into all intuitions a priori. This synthesis is called the "pure (productive) synthesis of the imagination" or "transcendental synthesis of the imagination" (TSI).[6] (CPR: A 118-119)

(3) The capacity or power of understanding is in fact analytically decomposable into two more basic transcendental powers: (i) the transcendental unity of apperception, and (ii) the TSI. (CPR: A119)

(4) The understanding contains pure concepts (PCUs) applicable to objects in

[6] In the B edition, the TSI is also called the "figurative synthesis" or the *synthesis speciosa*, and sharply distinguished from the *intellectual synthesis* of the understanding.

general--i.e., the categories--which at bottom are nothing but the purely intellectual modes of the TSI in its a priori application of the unity of apperception to the manifold in and of the intuitions given in our sense perceptions of appearances. (CPR: A119)

(5) Implicit premise: An object of experience is nothing but the rule-governed synthesized manifold in and of the intuitions given in our sense perceptions of appearances. This follows from TI and the theory of synthesis.

(6) Given (4) and (5), TSI both brings conceptual form to intuitions, and also introduces necessary rule-governedness (= law-governedness) into all appearances in order to determine them fully. This produces *nature*, which for Kant is the same as the law-governed totality of appearances. (CPR: A120-128)

(7) Therefore, since an object of experience is nothing but the rule-governed synthesized manifold in and of the intuitions given in our sense perceptions of appearances, it follows that the categories are necessarily applied to objects of experience and make them possible. (CPR: A128)

(8) So the PCUs or categories are necessary a priori conditions of the possibility of all objects of experience. QED

*(7.5) Some critical worries.*

There are many critical worries that could be and have been raised about the A Deduction. And some of these will come out in our discussion of the B Deduction. But for the time being, I am particularly interested in four of these worries.

First, why should we accept Kant's doctrine of the threefold synthesis? And more generally, what is the nature and philosophical status of transcendental psychology? This is pertinent even if the subjective deduction is not crucial: for some of the cognitive powers described in the subjective deduction *also* play essential roles in the objective deduction.

Second, what precisely is the function of "the transcendental object = *X*" and how does it relate (a) to transcendental apperception, (b) to judgments, (c) to appearances as undetermined objects of empirical intuitions, (d) to objects of experience, and (e) to negative or positive noumena?

Third, what are we to make of that philosophical loose cannon, the capacity or power of imagination? How can it at once be the source of all syntheses whatsoever (in the objective deduction) and yet also be restricted to a pure condition for empirical reproduction (in the subjective deduction)? And what is logically and cognitively more basic: the transcendental imagination, or transcendental apperception?

Fourth and most importantly, Kant interweaves with the A Deduction a doctrine

of what he calls the "empirical affinity" of the manifold (CPR: A114), which is the thoroughgoing coherence of all empirical nature. Such an empirical affinity or total coherence of appearances is said to be identical to the causal law-governedness of empirical nature. And this law-governedness in turn is said to follow follow directly ("as the mere consequence") from the fact that the categories are necessary a priori conditions of the possibility of all objects of experience, which is called the "transcendental affinity" of the manifold (CPR: A114, A122, A126-128).

But in relation to the doctrine of empirical affinity, I see three nasty problems lurking on the horizon:

> (1) The "top-down problem," or "problem of transcendental rule application," according to which the empirical affinity of the manifold does *not* follow automatically from the transcendental affinity of the manifold.
>
> (2) Hume's skepticism about the existence and/or knowability of necessary connections in nature, and his related skepticism about justifying induction.
>
> (3) The "bottom up problem," or "problem of essentially rogue objects," according to which there might be not merely undetermined but in-principle *undeterminable* objects given in "blind" or non-conceptual empirical intuition.

Can Kant solve these nasty problems? If so, how?

The chapter on the schematism of the PCUs is an attempt to solve problem (1).

The system of all principles of pure understanding, and especially the second Analogy of Experience (together with the two other Analogies) is an attempt to solve problem (2).

As for problem (3), unfortunately there is no serious attempt on K's part to solve it in the first *Critique*. There *are* some brief but highly suggestive remarks in the appendix to the Transcendental Dialectic, "On the Regulative Use of the Ideas of Pure Reason," at A650-651/B678-679: but they only hint at a possible solution.

These sketchy remarks are however the basis of K's later serious & in fact unfinished attempt to solve the problem in the First Introduction to the *Critique of the Power of Judgment*, the Introduction to the *Metaphysical Foundations of Natural Science*, and the *Opus postumum*.

The basic idea of this later "post-Critical" project is that what's given in "blind" or non-conceptual empirical intuition is necessarily some proper part of the unified totality of matter, which K ultimately identifies with a non-atomistic physical *aether* which is constituted by primitive attractive & repulsive moving forces, & which is the fundamental ontological basis of all inert & living material beings (including nonconscious animals & conscious animals) alike.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 8

Covers: CPR: 245-266 = B129-169 (Deduction of the Pure Concepts of the Understanding, B edition).

## The Transcendental Deduction of the Categories in the B Edition

*(8.1) From the A Deduction to the B Deduction.*

Probably the most important change in the B edition of the first *Critique* is the completely rewritten version of the transcendental deduction of the pure concepts of the understanding (PCUs) or categories. The overall purpose of the B Deduction, just as in the case of the A Deduction, is to prove that the categories are necessary a priori conditions of the possibility of objects of experience. So too, just as in the case of the A Deduction, the B Deduction occurs against a dual backdrop: it assumes both (1) the main conclusions of the Transcendental Aesthetic (including the doctrine of transcendental idealism), and (2) the main conclusions of the Metaphysical Deduction.

What are the main differences between the A Deduction and the B Deduction? Three particularly stand out.

**First**, the so-called "subjective" and "objective" Deductions in the A Deduction are smoothly combined in the B Deduction within a single argument-structure.

**Second**, the to-die-for doctrine of the three syntheses in the A Deduction is (mostly—but see §26 for a crucial reference to the synthesis of apprehension) replaced in the B Deduction by a detailed discussion of the nature of a judgment.

**Third**, the strong emphasis upon the pure productive imagination--or transcendental synthesis of the imagination--in the A (objective) Deduction is downplayed in the B in favour of an increased reliance upon transcendental apperception and upon the *power* (innate faculty) of judgment.

**Fourth**, given the first three differences, it is now supposed to follow that *necessarily, every appearance corresponds to an object of experience*, hence the necessary applicability of the Categories to all objects of experience also entails "the application of the catgories to objevts of the senses in general" (CPR: B150).

So the essential difference between the A and B Deductions seems to be a "turn towards judgment," and more specifically, a turn towards *judgments of experience*. Indeed, this

turn towards judgments of experience is particularly evident in sections 18-23 of the *Prolegomena*—which was published in 1783, between the A (1781) and B (1787) editions of the first *Critique*. Be very wary, however, of the doctrine of "judgments of perception" in the *Prolegomena*. The basic problem is that since judgments are defined by K as objectively valid, necessary unities of representation (see, e.g., section 19 in the B Deduction), while the so-called "judgments of perception" are contingent unities of representation and at best subjectively valid, the latter could not possibly *be* judgments. *Only judgments of experience, not so-called "judgments of perception," can genuinely count as judgments.* This is a mistake that Kant quietly rectified by the time he did the B edition, and never mentions again. But the more general problem of how we are to make sense of cognitions of appearances that fall short of satisfying all the Categorial conditions of objectivity (i.e., rogue objects) lingers on in the The Gap problem for the B Deduction.

K however helpfully tells us just where the nub of the B Deduction occurs: in section 20 (B143-144) and in section 26 (B159-B161). Put together, these two short texts provide a single two-phased argument which I will reconstruct by quoting, and then commenting on, the several premises and conclusions in each phase.

Nevertheless, it doesn't follow that the rest of the B Deduction is irrelevant; on the contrary, the highly-compressed argumentation given in sections 20 and 26 makes sense only in light of all the other sections. So they must be assumed & understood as context.

*(8.2) A Map of the B Deduction*

Here is the structure of the B Deduction in schematic form:

**(A) The Run-up to Section 20**

§15: the poss. of combination in general
§16 the Original Synthetic Unity of Apperception (OSUA)
§17 the principle of the synthetic unity of app is the supreme principle of all use of the understanding
§18: what objective unity of self-consc is
§19: the logical form of all judgments consists in the objective unity of the apperception of the concepts contained therein

**(B) Phase I: The Argument in Section 20**

**(C) The Transition to Phase II & the Run-up to Section 26**

§21: remark (on the givenness of the manifold in intuition)
§22: the category has no other use for the cognition of things than its application to objects of experience
§23: (objective validity/objective reality of the categories)
§24: application of the categories to objects of the senses in general

* * *: (inner sense, apperception, and judgment)
§25: (more on apperception)

**(D) Phase II: The Argument in Section 26**

**(E) The Follow-up to Section 26**

***: (rules of perception and causes)
***: (natural laws)
§27: result of this deduction of the concepts of the understanding

*(8.3) Phase I: The Argument in Section 20*

(1) "The manifold that is given in a sensible intuition necessarily belongs under the original synthetic unity of apperception, since through this alone is the **unity** of the intuition possible (#17)." (CPR: B143)

Commentary: The manifold of sensory content in intuition is subject to a fundamental unity provided by the pure formal intuitions of our forms of intuition = the representation of space (r-space) and the representation of time (r-time). The source of the unity of these pure formal intuitions, in turn, is the original synthetic unity of apperception (=the innate capacity for unifying representations which is necessarily involved in any consciousness of an object, plus the cognitive function for self-ascribing the contents of representations: "I think x," where x = the object of representation.

(2) "That action of the understanding, however, through which the manifold of given representations (whether they be intuitions or concepts) is brought under an apperception in general, is the logical function of judgments (#19)." (CPR: B143)

Commentary: Any unification of the manifold, under the unity of apperception, is a synthesis involving concepts and therefore also the understanding. The understanding insofar as it is applied to the manifold is a judgment. So the unification of the manifold is essentially an act of the power of judgment.

(3) "Therefore all manifold, insofar as it is given in **one** empirical intuition, is **determined** in regard to one of the logical functions for judgment, by means of which, namely, it is is brought to a consciousness in general." (CPR: B143)

Commentary: Judgments are classified under fundamental types, each of which specifies a certain logical function (universal or particular quantification, negation, predication, conditionalization, modalization, etc.). Therefore since the unification of the manifold of intuitions is essentially an act of the power of judgment under the unity of apperception, any particular empirical intuition in the manifold must fall under at least one of these logical functions.

(4) "But now the **categories** are nothing other than these very functions for judging, insofar as the manifold of a given intuition is determined with regard to them (#13)." (CPR: B143)

Commentary: We know already from the Metaphysical Deduction of the PCUs that the PCUs are the logical forms of judgment, and that they are also the categories--i.e., they are necessarily applicable to all objects of intuition in general cognizable$_{ns}$ through judgments. And for sensory cognizers just like us, such objects are originally given only in empirical intuition.

(5) "Thus the manifold in a given intuition also necessarily stands under the categories." (CPR: B143)

Commentary: Now a given empirical intuition represents an object by means of a manifold or field of sensory content: sensory qualities plus spatiotemporal organization. But the categories are necessarily invoked as classificatory forms in any act of judgment, and any act of the power of judgment is a unification of the manifold of intuitions. Hence the sensory manifold or field of any given intuition is necessarily governed by the categories.

*(8.4) The Transition to Phase II*

This marks the end of the first phase of the argument. The categories have been shown to be necessarily applicable to the sensory manifolds or sensory fields in all particular empirical intuitions. Now what Kant needs to do is to get the categories "onto" full-blown objects of experience. For a mere empirical intuition (= a synthesis of apprehension directed to an appearance which is a "blind" or "obscure" sensory manifold = an undetermined sensory individual) is not the same as the representation of an object of experience (= a well-ordered empirical state-of-affairs, represented by a rule-governed sequence of perceptions). Moreover, he also needs to show that categories apply to *all* objects of our sensory intuition.

Therefore it is only "by the explanation of [the category"s] *a priori* validity in regard *to all objects of our senses* [that] the aim of the deduction will first be fully attained" (CPR: B145, emphasis added).

*(8.5) Phase II: The Argument in Section 26*

(6) "First of all I remark that by **synthesis of apprehension** I understand the composition of the manifold in an empirical intuition, through which perception, i.e., empirical consciousness of it (as appearance), is possible." (CPR: B160)

Commentary: Perception is not just a passive sensory consciousness of a manifold of given qualities. In order to perceive, a certain spontaneous act of the mind is necessary, and this is the synthesis of apprehension. And we know from the A Deduction that the synthesis of apprehension is the act of bringing together sensory contents under a unitary

spatiotemporal form—i.e., the form of a unified sensory field at a given moment and as spread out in a given space. The internal content of the field is represented as a holistic, unified spatial and temporal pattern or array (a *Gestalt*) but is otherwise unstructured.

(7) "We have **forms** of outer as well as inner sensible intuition *a priori* in the representation of space and time, and the synthesis of apprehension of the manifold of appearance must always be in agreement with the latter, since it can only occur in accordance with this form." (CPR: B160)

Commentary: The representations of space and time, as a priori subjective forms of intuition that arew invoked in the synthesis of apprehension, are also necessary conditions of the perception of objects. This is simply a reminder of something already proved in the Transcendental Aesthetic, namely that (the representations of) space and time are the necessary a priori forms of all empirical intuitions (sense perceptions) of appearances.

(8) "But space and time are represented *a priori* not merely as **forms** of sensible intuition, but also as **intuitions** themselves (which contain a manifold), and thus with the determination of the **unity** of this manifold in them (see the Transcendental Aesthetic)." (CPR: B160)

Commentary: In the Transcendental Aesthetic Kant argued that space and time could be thought as empty of all sensory objects, and that pure intuitions of space and time--as infinite given wholes--were also possible. And at CPR: A712-A738/B740-B766 he also argues that geometric truths are demonstrated by "constructions" employing the representation of pure space. But it is crucial to recognize that space and time are not empirical objects of any sort, & if we try to imagine them as merely abstracted from empirical objects they are strictly speaking nothing at all: space or time in this sense would be merely an "**empty intuition without an object, *ens imaginarium***" (CPR: A292/B348). *Nevertheless according to Kant it is still possible to become reflectively self-conscious of r-space and r-time.* To do so is thereby to generate not merely subjective a priori necessary forms of empirical intuition, but instead pure or "formal intuitions" which non-conceptually represent space and time as total unified frameworks of relations independently of any particular empirical objects in space and time (CPR: B160 n.).

(9) "Thus even **unity of the synthesis** of the manifold, outside or within us, hence also a **combination** with which everything that is to be represented as determined in space or time must agree, is given *a priori* along with (not in) these intuitions, as condition of the synthesis of all **apprehension**." (CPR: B161)

Commentary: Because the synthesis of apprehension presupposes the pure or formal intuitions of space and time, it follows that the formal unity of these pure intuitions--as distinct from the sensory content of the intuitive manifold--is given as an experience-independent presupposition of the synthesis of apprehension.

(10) "But this synthetic unity can be none other than that of the combination of the manifold of a given **intuition in general** in an original consciousness, in agreement with the categories, only applied to our **sensible intuition**." (CPR: B161)

Commentary: The formal unity of pure intuition, as a subjective unity, must be grounded in the original synthetic unity of apperception. But apperception is also the ground of the unity of the understanding, and every synthetic act of the understanding presupposes the categories. So since the apperception which underlies the formal unity of pure intuition is the same as the apperception which underlies the unity of the understanding, it follows that the synthesis of apprehension falls necessarily under the categories.

(11) "Consequently all synthesis, through which even perception itself becomes possible, stands under the categories, ...." (CPR: B161)

Commentary: Perception is an empirical intuitional consciousness of the sensory object or appearance. Now perception is possible only via a synthesis of apprehension, and hence by step (10) presupposes the categories.

(12) ".... and since experience is cognition through connected perceptions, ...." (CPR: B161)

Commentary: Just as perception is distinct from mere sensory consciousness of affection, so too experience is distinct from mere perception. A perception is a consciousness of a sensory manifold or field, as apprehended (=an appearance at a moment or in a space). An experience by contrast is a rule-governed complex of perceptions, and its object is not a bare or unconceptualized apparent object but rather an empirical state-of-affairs--a complex, determinate object represented by means of concepts and judgment. Hence an experience is both distinct from a perception, and yet *built up from* perceptions by conceptualization and judgmental synthesis.

(13) "...., the categories are conditions of the possibility of experience, and are thus also valid a priori of all objects of experience." (CPR: B161)

Commentary: Notice the crucial move here from ***"****conditions of the possibility of* ***experience of objects"*** to ***"****conditions of the possibility of* ***objects of experience."*** Let's call "experiences of objects" EOs and let's call "objects of experience" OEs. So in making this crucial move, Kant is presupposing the thesis that necessarily, EOs = OEs. Now it's clear that *only* the thesis of *strong* transcendental idealism will make this presupposition true and thus validate the step from the conditions of the possibility of EOs to the conditions of the possibility of OEs. So, interpreting Kant charitably, he must be assuming the truth of strong transcendental idealism for the purposes of the B Deduction. Because the categories are necessarily applicable to all perceptions (step (11)), and because experience is synthetically generated from perceptions by means of conceptualization and judgment (step (12)), it follows that the categories are necessarily applicable to, and hence necessary conditions of the possibility of, experience. But all experiences are representations *of objects*; and by the thesis of strong transcendental

idealism, the contents of those representations are *token identical to* the objects of experience. Therefore the categories are necessary a priori conditions of the possibility of objects of experience. QED

*(8.6) Some critical worries.*

(i) What is the OSUA?

The central philosophical notion of the B Deduction is Kant's theory of the original synthetic unity of apperception (OSUA). The OSUA is neither inner sense, nor empirical self-consciousness, nor a Humean bundle of perceptions, nor a Cartesian ego, nor a noumenal subject of any other sort. So what is it? How does it compare and contrast with Descartes's and Hume's conceptions of the self? How does it relate to inner sense (not to mention outer sense) and to empirical self-consciousness? How does it differ from the "analytical unity of apperception" (B133)? And how does it relate to the power of judgment? See also the Paralogisms of Pure Reason (both editions).

(ii) The top-down problem again.

K. notes that "the pure faculty of understanding does not suffice ... to prescribe to appearances through mere categories a priori laws beyond those on which rests a **nature in general"** (CPR: B165). Is this simply an admission that the B Deduction fails to solve the "top-down problem" (see lecture notes for WEEK 8)?

(iii) The problem of empirical laws.

By conceding the top-down problem, K. also opens a window for another difficulty, the problem of "empirical laws" or "particular laws." The problem is that every law of nature is necessary for K., hence a priori. But then how can a law of nature be necessary and a priori if "experience must be added in order to come to know particular laws **at all**" (B165)?

(iv) Humean skepticism again.

Does the B Deduction answer Hume's skepticism about causation and induction?

(v) The bottom-up (or rogue object) problem again.

A serious problem for Kant arises from the fact although empirical intuitions without concepts are "blind" and "obscure," nevertheless "blind" or non-conceptualized intuitions are *possible* (CPR: A90/B122). This leads to what I called the "problem of essentially rogue objects" (see the lecture notes for WEEK 9): given that non-conceptualized empirical intuitions are possible, couldn't (at least in principle) the "undetermined object" (CPR: A20/B34) of such an intuition *essentially* evade the nets of the synthesis of recognition, transcendental apperception, and the categories (= conditions of the possibility of a scientifically law-governed nature)?

That is, might such an object be *intuitable* but totally *unconceptualizable*? Kant raises this worry himself at CPR: A90-91/B123. Now here are some examples of essentially rogue objects:

(i) incongruent counterparts like my right and left hands (see the *Prolegomena* and Lecture 3),
(ii) the arbitrarily-ordered stream of consciousness in inner sense (see the Second Analogy of Experience and Lecture 10),
(iii) organisms, to the extent that they inherently exhibit teleological self-organization and are not merely machines (see the second half of the *Critique of the Power of Judgment*), and
(iv) spontaneous, conscious, and freely-controlled, but also not reflectively planned, intentional body-movements like ecstatically flinging my arm into the air while dancing (see the Third Antinomy of Pure Reason and Lecture 12).

Notice that essentially rogue objects are all aspects of the lives of embodied, conscious, living, free rational human animals that do *not* seem to be reducible to the inherently mechanistic facts explicable by the natural sciences. So the problem of essentially rogue objects is *also* the problem of determining the legitimate scope and limits of the natural sciences and natural-scientific explanation.

If essentially rogue objects are really possible, however, then apparently the conclusion of phase I will not go through. But if phase I does not go through, phase II does not go through either. So, like philosophical wild elephants, do the essentially rogue objects ultimately stomp the TD of the PCUs to bits?

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 9

Covers: CPR: 267-295 = A130-176/B169-218 (Schematism, Axioms, Anticipations).

## The System of Principles I: Schematism, Axioms of Intuition, & Anticipations of Perception

*(9.1) Life after the B Deduction.*

If the B deduction is sound, then Kant has demonstrated that the pure concepts of the understanding (PCUs) or categories are necessary a priori conditions of the possibility of the experience of objects (EOs) and of objects of experience (OEs) alike. Therefore, the categories are applicable to all objects of experience (i.e., they are objectively valid or empirically meaningful). But he has not yet demonstrated just *how* they are applicable.

This arises as an issue for four reasons.

**First**, K. has not yet given an explanation of how concepts (as cognitive universals) are applicable to objects of experience (as cognitive particulars).

**Second**, since the categories are pure but objects of experience are empirical and concrete, there is the general problem of how non-empirical items apply to concrete items.

**Third**, since the categories are PCUs and not pure intuitions, there is the added complication that the categories will not apply *directly* to empirical objects but only via judgments.

And **fourth**, left over after the B Deduction is the "top-down" problem, or the problem of how universal transcendental rules can be specified in terms of "particular laws" or empirical laws of nature. K. addresses all these issues in the chapter on the schematism of the PCUs (CPR: A137-147/B176-187).

*(9.2) A Problem about Rule-Application in Judgments.*

In the introduction to the Analytic of Principles (CPR: A131-136/B169-175), K. indicates another way of formulating the four issues mentioned in (1), namely, as a problem about how rules are to be applied in judgments (a.k.a. the problem of "subsumption" of specific rules, or of particular objects, under general rules). According to K., concepts are rules for organizing the perceptual manifold; and judgments are to be understood as rules for

applying those concept-rules to particulars. Now if one raises a worry about how a judgment is to be carried out, then there seem to be only two options: the judger can already apply the rule immediately and "blindly" or without further guidance (& this anticipates a basic theme in the work of the later Wittgenstein); or the judger can appeal to another higher rule. But for K. the first answer begs the question, and the second leads to a vicious regress of higher rules. K. solves the dilemma by denying that the two options are exhaustive: there is to be a "third thing"—the "schema"—that gets between objects and concepts, and mediates the application of the latter to the former. This leads to his doctrine of the "schematism." Schematism is the primary function of the "productive" or "figurative" imagination and its *synthesis speciosa*.

*(9.3) What the Schematism Is.*

The general ideas behind the schematism are (1) that where both empirical and pure concepts are concerned, the productive or figurative imagination can be employed to generate schemata, and (2) that schemata make it possible to apply concepts to objects. So what is a schema?

This much we do know: schemata are representational contents that are in one way singular (like intuitions) but in another way general (like concepts). Therefore, as regards their representational character, they fall somewhere between spatiotemporal representations and discursive representations: they are essentially *intermediate* representations between intuitional information-content and conceptual information-content.

Now schemata are not mere "mental images" (*Bilder*) of sensory objects, because these are too particular and concrete. Here K. agrees with Berkeley's famous objections to Locke's theory of general ideas. But on the other hand the empirical schema for an empirical concept (say, DOG) is the "representation of a universal procedure of the imagination in providing a [mental] image for a concept" (CPR: A140/B179-180); hence there is an important connection between empirical schemata and mental imagery even if they are not quite the same.

By contrast, the transcendental schema for a pure concept is

> only the pure synthesis, in accord with a rule of unity according to concepts in general, which the category expresses, and is a transcendental product of the imagination, which concerns the determination of inner sense in general, in accordance with conditions of its form (time) in regard to all representations, insofar as these are to be connected together *a priori* in one concept in accord with the unity of apperception. (CPR: 183; A142/B181).

*(9.4) Schemata as Mental Models or Partial Interpretations.*

Put in simpler language, what K. seems to be proposing is that a schema is a *mental model* or *partial interpretation* for a concept. Think of it this way: a concept is a bit of

general descriptive information. This cannot be applied to particular cases without some way of interpreting that information in more specific and concrete terms. A schema gives a more specific and concrete yet still significantly formal and only partial interpretation, of that information. As more specific, it is applicable to classes of instances; as more concrete, it can be used to map directly into the target domain of intuitional content; but as formal and partial it cannot simply be identified with concrete instances. Otherwise put, it gives a *diagrammatic procedure* for applying the concept. Aspects of the model are systematically correlated with aspects of the general information on the one hand, and with aspects of particular objects on the other. In this way, schematic information enables us to indicate or *show* the way the world is, even if we can't describe or *say* it. K. himself characterizes schemata as "monogram[s] of … imagination" (CPR: A570/598).

Understood in this way, an empirical schema for an empirical concept is what might be called a "stereotype": a generic mental image, or a few salient sensible features (in any sense modality—visual, auditory, tactile, gustatory, olfactory, or proprioceptive) loosely organized into an easily remembered format. The stereotype is shared by all, or most, of the things that fall under the concept. Such stereotypes are in a constant process of construction, deconstruction, and reconstruction over time; their only fixed constraint is that they remain correlated with the *definition* of the concept on the one hand, and with actual items on the other hand. The acquisition of a stereotype is an output of the empirical productive imagination.

By contrast, the transcendental schema is generated by exploiting the properties of (the representation of) time, the pure form of intuitions in inner sense. As the Aesthetic indicated, time has some general necessary features such as successiveness, duration, a uniquely asymmetric flow (sometimes called "time's arrow"), the relation of before and after, and the past/present/future structure. These properties are now exploited in various ways by K. to provide temporal models, or partial interpretations, for the categories.

*(9.5) How to Schematize the Categories.*

According to K., as the Metaphysical Deduction shows, every category derives from a logical form of pure general logic by relating that form to objects in general. This gives the category a "thin" or purely logico-metaphysical meaning without regard to the *specific sorts* of objects that might fall under it. We know from the A and B Deductions that the proper objects of cognitions are objects of experience or fully-determined appearances, not things-in-themselves. So in order to mediate between categories and objects of experience, what is needed is a representation that is intermediate between pure conceptual content and the intuition-based representations of macroscopic material empirical objects or state of affairs. The pure forms of intuition, space and time, are the obvious candidates for the job.

For K., to schematize a category is simply to *correlate* its pure conceptual or metaphysical meaning with some formal feature or property of (the representation of) time. Because we know from the Transcendental Aesthetic that (the representation of)

time is transcendentally ideal and empirically real, and also a form that applies to every object of empirical intuition whatsoever, it will follow that schematizing the category will (a) restrict the category to appearances, and (b) give it an application to every object of experience insofar as it realizes that temporal property.

Here's an example. Take the metaphysical categories of quantity (unity, plurality, and totality), which derive from the judgment-forms of universality (All S is P), particularity (Some S is P), and singularity (This S is P). How are quantitative notions to be applied to objects of experience? Answer: Via the *successive* property of time, namely that it is always given as a constantly-growing recursive series of singular temporal moments running one-by-one from some initial moment up to any later moment. Now what sort of quantitative object encodes the formal temporal properties just mentioned?

Answer: the natural numbers including the integers (cardinal numbers, counting numbers). Hence the *transcendental schema* of quantity is "number," and any object of experience that can be counted in any way falls under that schema. In principle at least, each of the other categories can be similarly correlated with some special property of time, and thereby schematized.

*(9.6) The Systematic Function of the Transcendental Schemata.*

Suppose now that every category has been supplied with a time-schema (see CPR: A142-145/B182-185). K. is then in a position to tell us just how each categorial notion can have an application to objects of experience. Henry Allison usefully dubs the several propositions expressing the several schematizations, "schema judgments." In turn, there are synthetic a priori truths that are based directly on the schema judgments, and explicitly express the applications of the several categories to objects of experience. These are truths are what K. calls Principles of Pure Understanding. And this of course relates the schematism directly to K.'s overall project of showing how synthetic a priori judgments are possible.

*(9.7) Two Worries about the Schematism.*

(1) <u>Top-down, but not bottom-up?</u>

Obviously the Schematism makes some progress towards answering the "top-down" problem. But what about the "bottom-up" problem of nonconceptual empirical intuitions that might latch onto unconceptualizable objects (i.e., essentially rogue objects) falling beyond the scope of all functions of the understanding? The specific form of the worry in this connection is that there might be sensory intuitions whose essentially rogue objects inherently don't conform to the four a priori time-determinations mentioned at CPR: A145/B184-185. One important example of this would be the arbitrarily-ordered stream of consciousness in inner sense (see the Second Analogy of Experience and Lecture 10).

(2) Why just r-time?

Why is the representation of time *alone* used for the generation of transcendental schemata? Why not r-space, or r-time-plus-r-space? And what would spatial—or spatiotemporal—schemata of the same categories be like? If r-time is the immediate form of *inner* sense, then isn't it possible that schematizing the categories solely in terms of time might restrict the meaning or objective validity of the categories too narrowly to purely mental or subjective contexts? For some remarks which strongly suggest that Kant himself had similar worries, see CPR: A157/B196 and B291. And of course there is also the Refutation of Idealism, which says that self-representation in inner sense is possible only through outer experience (CPR: B275-279). The latter argument certainly suggests that the pure representations of time and of space are semantically and metaphysically complementary; but if so, then it would seem to follow that the transcendental schematism would have to invoke both spatial and temporal, or even spatiotemporal, schemata.

*(9.8) What "Principles of Pure Understanding" Are, & Why We Care about Them.*

Principles in general for Kant are normative rules. These can be either practical (e.g., prudential rules or moral rules) or theoretical (e.g., scientific rules or logical rules). The principles of pure understanding (PPUs), according to Kant, are the highest scientific rules: primitive synthetic a priori judgments that govern not only all empirical judgments but also all other synthetic a priori judgments--which he also sometimes rather confusingly calls 'principles'--including mathematical truths and a priori truths of natural science (i.e., propositions stating causal natural laws).

But even in the realm of principles, PPUs are special. That's because they're the *basic metaphysical truths* of K's transcendental idealism.

Unlike mathematical truths, which are derived more or less directly from pure or formal intuition, principles of pure understanding (PPUs) are derived from the pure concepts of the understanding (PCUs) or categories by means of the transcendental schematism. The purpose of the Analytic of Principles is thus to explain how the PPUs are meaningful and true, in light of both the Transcendental Deduction of the PCUs and the Schematism, and thereby to explain how the basic propositions of metaphysics are meaningful and true. So in the rest of this lecture I want to cover the general theory of PPUs, and the first two Principles (CPR: A148-176/B187-218).

*(9.9) The Criterion of Truth for PPUs.*

For K., as we know, to give a transcendental deduction of some representational content or judgment is to give an argument demonstrating the necessary and sufficient conditions of that content's empirical meaningfulness from a priori grounds; and this is the same as to supply non-empirical necessary and sufficient conditions for its being objectively valid, hence true or false.

In the A and B Deductions and the Schematism K. worked out two thirds of a three-step argument for semantically grounding PPUs: (1) the Transcendental Deduction of the PCUs shows *that* the PCUs or categories must apply to all actual and possible objects of experience, and (2) the transcendental part of the Schematism shows *in general how* the PCUs or categories apply to all actual and possible objects of experience. Then (3) the Analytic of Principles shows *specifically how* the PCUs or categories apply to all actual and possible objects of experience by explaining the primitive synthetic a priori truths which express the application of the several PCUs to all actual and possible objects of experience.

Truth, as we also know, is the correspondence ("agreement") of a judgment ("cognition") with its object. The negative criterion of the truth of any judgment is that it be logically or conceptually self-consistent (i.e., not contradictory); and the positive criterion of truth for analytic judgments is that their denial entails a contradiction. But what is the positive criterion for the truth of a synthetic a priori proposition? K.'s answer is that "the **possibility of experience** is ... that which gives all of our cognitions *a priori* objective reality" (CPR: A156/B195) and that

> [t]he supreme principle of all synthetic [*a priori*] judgments is, therefore: Every object stands under the necessary conditions of the synthetic unity of the manifold of intuition in a possible experience. (CPR: A158/B197).

That is, he is saying that a synthetic a priori proposition is true if and only if it applies not only to all actual objects of experience but also to all possible objects of human experience. Put in my lingo, this is to say that a synthetic a priori proposition is true if and only if it is true "in every experienceable world." And that's precisely what the argument in the Analytic of Principles is intended to demonstrate—assuming, of course, the soundness of the Transcendental Aesthetic, the Metaphysical Deduction, the A and B Deductions, and the Schematism.

*(9.10) Mathematical vs. Dynamical Principles.*

Kant draws a distinction between two types of PPUs: (1) those which express conditions necessary for the application of mathematical truths to nature, and (2) those which express conditions necessary for the application of truths of physics to nature. Roughly speaking, the distinction is this: (1*) for the mathematical principles, the conditions expressed in them are such that mathematical truths apply strictly to all objects of possible experience whatsoever; but by contrast, (2*) for the dynamical principles, the conditions expressed in them include the requirement that a certain existential assumption be satisfied--hence necessary a priori truths of natural science hold only for all possible empirical objects *under that existential assumption*.

The relevant assumption for dynamical principles is that *just this kind of matter* exists: that is, just the kind of matter we find in the actual world = inert matter = matter that operates by exclusively by mechanical principles & extrinsic forces. Therefore not every possible object of experience to which mathematics applies is an object to which

Newtonian mechanistic physics applies: at least in principle, we could still have sense experience in a world in which the laws of nature were radically different, or in which matter were radically different (e.g., a world in which there was *living matter*, or organismic matter that operates according to teleological principles & intrinsic "living" forces: the metaphysical thesis of *hylozoism*).

And here is another case in which K. seems to have been highly prescient. The modern science of *complex systems dynamics*, e.g., is a mathematical theory that applies to non-mechanistic natural phenomena and irreversible thermodynamic processes, especially biological processes. Of course, complex systems dynamics doesn't itself entail hylozoism. But it's consistent with hylozoism, and if hylozoism were true, then universal natural mechanism (as, e.g., in Newtonian science) would be false.

*(9.11) Axioms of Intuition.*

The first categorial principle applies the categories of quantity (schematized, we will remember from *(5)* above, as number) to objects of intuition. The principle is this: "all intuitions are extensive magnitudes" (B202). I take this to mean that necessarily for any possible object of experience, if it is an object of intuition then it is given as an extended aggregate of parts in space or time in such a way that its extensive parts are countable or measurable. So no matter *how* an object of intuition is divided into parts, its parts are such that they can be added, subtracted, etc., and more generally, such that they can be treated mathematically.

*(9.12) Anticipations of Perception.*

Not only are objects of intuition necessarily such that they have (or can have) parts that are spread out (at the very least) in time; they also are such that they *fill* or *occupy* time. For K., the filling or occupation of time is the schema of the category of reality: something is real only insofar as it *occurs* in time, moment by moment. This leads directly to his principle that "in all appearances, the real, which is an object of the sensation, has intensive magnitude, that is, a degree" (CPR: B207). Intensive magnitude is simply the *force* of something: how much it psychologically strikes you; or how something physically impresses itself upon something else; or how two or more things physically attract one another; or how two or more things physically repel one another.

So I take this principle this to mean that necessarily for any object of experience, if it is an object of sensation, then each of its sensory qualities occupies a fixed position within a certain continuous range of distinct but internally-related force-qualities in space or time. In other words, each of its sensory qualities is a spatiotemporalized force-qualitative *determinate* under a force-qualitative *determinable* (see Prior and Searle on "determinables" and "determinates"). For example, for something to be red is for it to have a certain *shade* or hue of red; for it to fall within the red band of the visible spectrum as opposed to the orange or yellow band, etc. Moreover, there is an infinitely divisible range of such degrees of quality.

One interesting consequence of such a view is that although we cannot know in advance of experience just *which* sensory force-qualities will be apparent, we do know that necessarily they will occupy a fixed position somewhere within certain spatiotemporal system of degrees of those force-qualities: hence to that extent we can "anticipate" the qualitative structure of any object of sense perception. This may provide the beginnings of an answer to Hume's notorious "missing shade of blue" problem: we are given individual colors as inherently belonging to a larger *structure* of colors. It may also play a significant role in solving the more recent but equally famous problem of the interpretation of "color-incompatibility propositions" (e.g., Necessarily nothing can be simultaneously red all over and green all over, because red inherently excludes green--but this could not have been analytically derived from the concept RED nor could it have been analytically derived from the concept GREEN).

A second interesting consequence of such a view is that when the Axioms are combined with the Anticipations, we derive a spatiotemporal sensory manifold with two distinct but systematically coordinated *dimensions* of infinite divisibility, extensive and intensive. It is arguable that this will allow us to represent *real* numbers in Kantian terms, not merely denumerable quantities (natural numbers or rational numbers).

And a third interesting consequence of the Anticipations is that according to K. the perceived world has no genuinely "empty" areas in it; what appears as "empty space" or "empty time" must in fact be filled with dynamic material forces corresponding to sensory content in order to be real, no matter how small the degree of intensity of its quality.

*(9.13) Some worries about the Axioms and Anticipations*.

(1) Of time and the deep blue sky: a problem about the Axioms.

Is Kant saying that the intuited world is necessarily always *immediately* given as an extensive magnitude, or merely only that the intuited world is necessarily always such that it *can* be given as an extensive magnitude? The former claim seems false. Sometimes, I think, we perceive or intuit spatial objects as perfectly undivided or "seamless" wholes, for instance our normal awareness of the blue sky on a cloudless day. So too as Henri Bergson argued in his *Introduction to Metaphysics,* our unreflective experience of time is not successive but durational and seamless. We can in principle divide the cloudless blue sky and temporal durations into extensive regions or parts, but they don't seem to be originally given that way.

Now these facts make one wonder whether objects of intuition are *ever* immediately given as extensive magnitudes. Mightn't the Axioms reflect only what Edmund Husserl in the *Crisis of European Sciences* calls a "Galilean idealization" of the immediately given "lifeworld"? If so, then even assuming that the empirical world is immediately given as spatiotemporal, it seems that this spatiotemporal structure might not be immediately or automatically appropriate for the application of the formal or natural sciences to it.

(2) The indeterminacy of partitions: another problem about the Axioms.

Let's assume now that a given intuited object is actually displayed as an extensive aggregate. Is Kant is saying that for every such intuited object, there is a *unique* division into parts, hence a single way of calculating over it? For that seems false too. Consider, for example, the perception of a house: how many parts has it got? One answer is that it has three parts, namely (1) the roof, (2) its walls, and (3) its basement and foundations. Another answer is that it has a great many parts, namely all the separate bricks of which it is made. And so on. More generally, how an object is intuitively divided into parts does not seem to be unique. Rather it seems to depend on background assumptions concerning the *sort* of thing we are partitioning and our *interest* in partitioning it. But an appeal to human interest slides off into pragmatic factors which K. seems definitely to want to exclude.

(3) Substances, trope-bundles, or states of affairs? A worry about the Anticipations.

What, according to Kant, *is* the real object of a sense perception, that is, the empirical *thing* (*Ding*)? It cannot be part of the sensory experience, because then it could not "correspond" to the sensation (CPR A168/B209). Is it then an empirical substance which *has*—is a substrate for—sensory qualities? Or is it merely a bundle of instances *of* qualities (what contemporary metaphysicians call "tropes")? Or is it a relation of some sort? If it is a substantial substrate for qualities, then it seems false to say that *it* has intensive magnitude--rather, only its qualities have intensive magnitude. If on the other hand it is nothing but a bundle of quality-instances or tropes, then while these tropes can certainly have intensive magnitude, a trope can't also *be* what "has" qualities. So a real empirical object or thing seems to be essentially a *relation* between a substantial substrate and its qualities (= a fact or state of affairs), which is *also* intrinsically related to the sensing subject.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 10

Covers: CPR: 295-321 = A176-218/B218-265 (Analogies of Experience).

## The System of Principles II: Analogies of Experience

*(10.1) The Metaphysical Significance of the Analogies.*

In some ways, the Analogies of Experience section is the metaphysical core of the first *Critique*. This is because in it Kant offers, in effect, solutions to three fundamental metaphysical problems: Berkeley's problem of how to account for the objectivity of a world made up entirely of wholly subjective sensory objects (i.e., a world of ideas), Hume's problem of the nature of object-identity over time, and Hume's other problem about the validity of our idea of causation or necessary connection in nature.

B. argues, you will remember, that matter is impossible and that to be an object is to be perceived by a thinking subject. B.'s solution to the objectivity problem is that a divine mind imposes an order upon the totality of subjective sensory objects by systematically affecting us in sensibility.

H. argues, you will also remember, that continuity of object-identity over time cannot be either directly experienced or legitimately inferred from experiences, & is nothing but a projection of the mind from the repeated association of similar experiences.

H. also argues, perhaps most famously, (a) that the ideas of causally necessary connections we naturally ascribe to perceived objects are false and vacuous because of the contingency of all temporal connections immediately presented to us in sensory impressions (skepticism about causal necessity), and (b) that even if causally necessary connections can in some sense exist "secretly" behind mere sensory objects, they are totally unknowable by means of the senses (metaphysical agnosticism).

H.'s "skeptical solution" to his problem about the idea of causation is that we non-rationally form habits of mind in experiencing constantly conjoined sensory events, and unconsciously project our habitual expectations, in the form of a belief that a necessary connection exists between all events of those types, onto the sensory data (radical psychological empiricism).

K.'s transcendental solutions to these problems avoid both B.'s appeal to a transcendent being, and also H.'s skepticisms about object-identity and causal necessity, and his radical psychological empiricism. K.'s idea, in a nutshell, is that creatures minded like us

cannot represent the material world without also representing it as substantially objective, as objectively identically enduring over time, and as causally necessitated in time and space, whether successively or simultaneously—so, assuming strong transcendental idealism, the world of appearances must conform to our minds and therefore be this way too.

*(10.2) The General Principle of the Analogies.*

Kant says that "[The Analogies'] principle is: Experience is possible only through the representation of a necessary connection of perceptions" (B218).

What he means is this. The three Analogies correspond to the categories of relation (substance/attribute, cause/effect, community), which in turn correspond to the relational forms of judgment (subject/predicate, hypothetical, disjunctive). The temporal schemata for the three categories are, respectively, persistence (existence of a thing through time), succession (the passage of events), and coexistence (simultaneity).

In each case, K. wants to say, assuming transcendental idealism, that the sensory objects given in experience will take on necessary temporal structures that are strictly transcendental--imposed by the subject. The application of all of these structures constitutes a *substantially-objective, identically enduring, causally law-governed* empirical world in time and space. That is, transcendental idealism + categories + schematization = a world that is fundamentally metaphysically appropriate for the application of necessarily and empirically true propositions in physics & ordinary human experience.

*(10.3) The First Analogy.*

The first Analogy is: "in all changes of appearances substance persists, and its quantum is neither increased nor diminished in nature" (B224).

What does *that* mean? Think of it this way: it's a transcendental *conservation of matter* principle. Take the logical form of a subject/predicate proposition, and metaphysically interpret it by applying it to objects in general. The result is the notion of a substance (an independently existing thing that supports properties), and its accidents (the contingent properties of the substance). Now take the metaphysical notion and give it a temporal interpretation (schema) in terms of duration. The result is the notion of something which exists "persistently" through time, and is also the substrate for the various changes in properties that occur through time. Well, what is the thing that exists persistently throughout all time and supports various changes in properties? Answer: physical matter. So K. is saying that necessarily every changing contingent property in appearances must be applied to, or predicated of, a material object that endures through time by virtue of its *intrinsic* (relational) properties.

Here K. rather puzzlingly talks about two distinct levels of the material substrate of empirical nature: on the one hand, he talks of a single substratum that exists persistently

throughout *all* time--that is the totality of matter; but on the other hand, he talks about a plurality of "lesser" substances that exist persistently for a while, and then go out of existence--those are the particular material beings. These two perspectives on substance can be reconciled, I think, only by assuming that the plurality of substances are *apparent parts* of the *real whole*, or One Big Substance, which is the totality of matter. In fact, there is one and only one substance (let's call it "primary substance"). This primary substance, the totality of matter, is preserved through the coming-to-be and passing-away of the many "secondary substances" by virtue of the fact that particular substances are simply individual organizations of matter, all of which eventually "break up." But the "quantum," or total supply, of matter is permanently preserved or persists.

This makes it possible for K. to give a very simple doctrine of "alteration" (*Veranderung*) or change. For something to alter is for a material substrate to have a succession of changing properties in a single or unique time; otherwise put, things come to be or pass away by virtue of the succession of properties.

The objects we experience are temporal complexes consisting of the One Big Substance + changing properties. Otherwise put, particular objects of experience are nothing but particular events in the long career of the One Big Substance. The unity of an object is nothing more than a certain *orderliness* imposed on the succession of properties applicable to primary substance.

*(10.4) Two Worries about the first Analogy.*

First, there is the <u>One Substance/many substances problem</u>.

Is it plausible, even within the framework of K.'s theory, that there is really only One Big Substance? What about particular empirical intuitions—surely they are not all directed to the same massive super-individual? Not only that, if K. is right, then every subject-term in every judgment of experience is ultimately applied to the same One Big Substance. So in that sense, we'd always be judging and talking about the same One Big Substance, no matter how differently meaningful our judgments of experience might initially seem. But on the contrary, it seems to be a manifest and even necessary feature of our experience that there be *many real substances*.

Second, there is the <u>One Substance/one time problem</u>.

Kant seems to believe that the unity of time is tied necessarily to the One Big Substance. For he argues that if there were many substances that came to be and passed away, then there would be many distinct times and not one time, which is absurd (CPR: A188-189/B231-232). But while it does seem to be correct that there is something absurd in the very idea of a plurality of times, it does not seem to be the case that the concept of a plurality of substances *entails* a plurality of times: why couldn't there be necessarily only one time, but contingently many real material substances?

*(10.5) The Second Analogy.*

This is the most famous of the Analogies, because it contains Kant's answer to H.'s skeptical analysis of our idea of causal necessity in nature. It goes like this: "all alterations occur in accordance with the law of the connection of cause and effect" (CPR: B232). This principle clearly builds on the first Analogy. In a nutshell, Kant is saying that the temporal succession of changing "states" or (*Zustände*) of a single One Big Substance (= "alterations" or *Veränderungen*) must include within itself a necessary connection between earlier and later states, such that the earlier states are nomologically sufficient for later ones.

That is because the category of cause/effect derives from the logical form of the hypothetical, which Kant understands as: Logically necessarily (which I'll abbreviate as "L-NEC") if P then Q (i.e., the antecedent is strictly sufficient for the consequent under a logical law). The cause/effect relationship, in other words, is the logical consequence relation as applied to objects in general.

Now the temporal schema restricts this relation to asymmetrically successive moments in time ("time's arrow"). Hence the schematized category of cause/effect is the logical sufficiency relation as mapped onto asymmetrically successive moments in time, which by virtue of its dependence on time makes it a *synthetically sufficient connection under a general law*, i.e., a nomologically synthetically sufficient connection.

A "state" or *Zustand* of the One Big Substance is the instantiation of a property at a time somewhere in the material world. So necessarily, whenever an earlier state is nomologically sufficient for a later state, *then* the later one is the effect and the earlier one is the cause. Perceptions of states that are ordered in this way (K.'s example is the successive positions of a boat floating downstream) are *objective or law-governed* orderings. By contrast, perceptions of states that are not so ordered are merely *subjective or arbitrary orderings* (K.'s example is the succession of sensory objects of someone's gaze flitting over a house).

The difference between the objective ordering and the merely subjective or arbitrary ordering is also the difference between the objective material world given in outer sense and determinately represented by judgments of experience (= successive states of the external or material world), and the merely subjective conscious world given in inner sense and represented by empirical apperception (= successive spontaneous conscious mental states, or the arbitrarily-ordered stream of consciousness). But notice that this implies the following striking Kantian doctrine: *the spontaneity of the arbitrarily-ordered stream of consciousness* is a necessary condition of the representation of an objective external world, even though, as we learned in Lecture 8, such streams of consciousness are *essentially rogue objects*.

The objective orderings are also called temporal "events" (*Begebenheiten, Ereignisse*), and should be contrasted both with (i) mere "states," which are just instantiations of

properties at times somewhere in the material world, and also with (ii) subjective orderings of perceptions or the arbitrarily-ordered spontaneous stream of consciousness.

K.'s theory of events has two parts. The first part says that synthetically necessarily (which I'll abbreviate as "S-NEC") for any x and for any two distinct properties P1 and P2, x is a *simple event* or *Begebenheit* (a.k.a. an "occurrence" or "something that happens") if and only if there exists an earlier state of x such that this moment in x instantiates a quality P1 that is synthetically sufficient for a later state of x which in turn instantiates another quality P2.

This is the same as to say that synthetically necessarily the earlier state of a simple event *causes* its later state as its *effect*. If you find quasi-formalizations helpful, what Kant is saying is this:

S-NEC (x) (P1) (P2) {x is a simple event ←→ (∃y) (∃z) [y and z are both states of x & y is prior to z & S-NEC (P1 is instantiated in y → P2 is instantiated in z)]}

In the second part of K.'s theory of events, he extends the account of simple events to events that are made up of simple events, i.e., *complex events* or *Ereignisse*, which contain at least two simple sub-events, as follows:

S-NEC (x) {x is a complex event ←→ (∃y) (∃z) [y and z are both simple events contained in x & y occurs earlier than z & S-NEC (y → z)]}

And similiarly for complex events containing three simple sub-events, and so-on.

But this is the crucial point: for Kant the objects of experience are *complex events built up out of simple events*, and these simple events, in turn, are *successive sequences of necessarily and nomologically connected states of the One Big Substance*.

In this way, for K. *the natural world is the totality of causally-structured simple or complex events*, not mere atomistic "things."

This solves both B.'s and H.'s problems in single swipe by proposing that objectivity, continuing object-identity over time, and causality in nature are all the result of our transcendentally imposing the r-time-schematized Second Analogy of Experience on sensory appearances.

*(10.6) Three worries about the Second Analogy.*

**First**, there is <u>the non sequitur problem</u>. In *The Bounds of Sense*, Peter Strawson says that K. commits a "non sequitur of numbing grossness." This is the fallacious inference from the premise that the *ordering* between two events is necessary (i.e., at time t1 boat B is higher up the stream, and at later time t2 boat B is lower down the stream—but B couldn't have gone downstream without first being higher upstream) to the conclusion

that the events follow each other necessarily (e.g., that boat B.'s being further down the stream at t2 is a necessary consequence of its having been higher up the stream at t1).

K. can solve this problem, however. The inference described by Strawson is of course fallacious, but K. is not arguing *from* necessary temporal ordering *to* necessary event-connection. Instead, what he is saying is that what *constitutes* x's being an objective item in nature is that x is an event which contains not only a necessary temporal ordering but also a necessary event-connection. So it's not an inference, it's a *metaphysical analysis*.

**Second**, there is the problem of simultaneous or synchronic causation. It is a consequence of K.'s metaphysical analysis of causation and objectivity that causal relations hold exclusively between earlier and later phases of events. But what about the many physical phenomena that are apparently both simultaneous or synchronic and yet causal, e.g.: the centripetal force of the earth's gravity now acting on me; and the light now illuminating this room?

Kant can also solve this problem. In fact he does so simply by offering the Third Analogy of Experience: "All substances, insofar as they can be perceived in space as simultaneous, are in thoroughgoing interaction" (CPR: B256). That is, all simultaneous or synchronic substances stand to one another in necessary and mutual interactive dynamical relations of co-determination.

Fair enough. But then what I want to know is: why isn't this a form of *causation*? Does causation *have* to be successive or diachronic? Why can't causation be simultaneous or synchronic?

The simple answer is: it can be! In the end, since according to the Third Analogy the simultaneity of material substances for K. necessarily involves reciprocal dynamic causal interaction, it seems to be no more than a *terminological convention* to say that causation has to be successive or diachronic.

But setting terminological conventions aside, K.'s *actual* doctrine is that causal-dynamic relations between material objects are BOTH successive or diachronic (the Second Analogy) AND simultaneous or synchronic (the Third Analogy). So now we can contrast either of the two basic kinds of causation (i.e., successive/diachronic causation & simultaneous/synchronic causation) with *mere coincidence*.

**Third** and finally, there is the problem of dream skepticism. Kant's criterion of natural-world objectivity is that x is objective if and only if x involves a necessary law-governed succession of states of a substance in time: but what is to prevent this structure from being imposed on dream-images or hallucinatory images (Kant calls these "phantoms of the brain")? I can, it seems, *dream about or hallucinate a boat floating downstream*. So, given K.'s account, why couldn't the "objective world of appearances" be simply a huge causally-structured dream?

This worry is particularly acute in light of the fact that K. *concedes* in the Refutation of Idealism that his argument doesn't in and of itself solve the dream-skeptical problem (CPR: B278-279).

In this connection, what Kant says at B291 is extremely revealing:

> In order to understand the possibility of things in accordance with the categories, and thus to establish the **objective reality** of the latter, we do not merely need intuitions, but always **outer intuitions**.

This strongly suggests to me that K. thinks that the categories don't apply to anything unless they're also applied to appearances in space. But doesn't that imply that the categories are also schematized spatially and not merely temporally?

And this takes us back to the Refutation of Idealism.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 11

Covers: CPR: 384-408 = A293-338/B349-396 (Transcendental Illusion, Transcendental Ideas) and 590-604 = A642-668/B670-668 (Regulative Use of the Ideas of Pure Reason)

## Transcendental Dialectic and Transcendental Ideas

*(11.1) What the Transcendental Dialectic is.*

For Kant, pure general logic is the science of the a priori & necessary laws of thought, and transcendental logic is pure general logic as restricted to objects of some sort or another.

Transcendental logic then divides into (1) transcendental analytic = the transcendental logic of *truth*, i.e., of (necessary) truth, validity, & soundness, and
(2) transcendental dialectic = the transcendental logic of *falsity*, i.e., of (necessary) falsity—including paradox, fallacy, and illusion.

Otherwise put, transcendental analytic is logic insofar as it applies to phenomena or appearances, and transcendental dialectic is logic insofar as it (supposedly) applies to positive noumena or things-in-themselves.

More generally in the Dialectic, K. undertakes the self-critique of pure reason from the standpoint of his transcendental metaphysics of human experience & exposes the main errors & confusions of traditional speculative metaphysics.

All of the errors & confusions ultimately have the same source however, which is the humanly natural but nevertheless tragic attempt to extend concepts & principles that are objectively valid with respect to appearances or phenomena or objects of experience, beyond their proper scope to positive noumena or things-in-themselves.

Our problem, basically, is that we're desperately finite limited embodied mortal fallible creatures whose deepest desire is nevertheless to be nothing less than god: we can't help trying to transcend ourselves towards noumenal grounds in order to justify ourselves & to confer meaning on our lives from the standpoint of what is radically outside us and hidden from us. This leads us directly into metaphysical confusion and skepticism.

The Kantian response to this is *critique*: a steady sober mature careful self-disciplined

undespairing & stoical awareness of our own essential limitations as knowers & practical agents. We *can't* know things-in-themselves. So get used to it, & learn to love the appearances.

On the other hand, we *can* always freely do the right thing (*ought* entails *can*), but because we're crooked timbers in a big bad world, we possess the capacity for screwing up in all sorts of more or less morally bad ways, all the way up to freely choosing immoral things for their own sake (a.k.a. "radical evil" or "perversity of the heart" in *Religion within the Boundaries of Mere Reason*). So demand a lot from yourself, but take it easy on others.

At the same time however, K. also thinks that what we are essentially denied in speculative metaphysics & epistemology, we are nevertheless able to secure for ourselves by means of our practical freedom or autonomy. Indeed, traditional metaphysics is nothing but sublimated ethics (Prol: 102-103; 4: 362-363).

*(11.2) The Ideas of Pure Reason.*

We know from the Metaphysical Deduction that all human thought is constrained & structured by the pure concepts of the understanding or PCUs, which in turn are a priori necessary *2nd-order* concepts (hence concepts about concepts) generated by our understanding, that guide the application all of 1st-order empirical concepts in judgments of experience.

The Ideas of Pure Reason or *notions* however are a priori necessary *3rd-order* concepts (hence concepts about the PCUs) generated by our reason, which consist in absolutizing & hypostatizing extensions of the PCUs beyond all possible experience into noumenal domains.

The basic Ideas of Reason are *God (absolute ground or ideal)*, *freedom (absolute causal power or spontaneity)*, and *immortality (absolute subject or soul)*. But in fact there are many such Ideas. Roughly speaking, for every domain of cognition or action, there will be a set of corresponding Ideas.

How are Ideas generated by the human mind? The logical function of reason is to draw inferences from premises. This can be absolutized. The transcendent metaphysical role of reason is therefore either to seek out a completed totality of logically antecedent premises or grounds for any given claim (regressive series) or else to seek out a completed totality of logical consequences for any given claim (progressive series). In either case speculative or transcendent reason always looks for the completed totality of grounds or consequences.

*(11.3) The Regulative Use of the Ideas of Pure Reason (IPRs)*

At the end of the System of Principles, there are three important leftover problems for Kant's transcendental project:

(i) top-down (the problem of empirical laws)
(ii) bottom-up (the problem of essentially rogue objects)
(iii) dream skepticism (the problem of objective reality)

Problem (iii) could be handled by requiring that contents have objective reality with respect to space & by schematizing the categories in space as well as time.

K. addresses problems (i) and (ii) in the Appendix to the Transcendental Dialectic on the Regulative Use of the IPRs, & again much more fully in the *Critique of the Power of Judgment* and the *Opus postumum.* Here is an outline of K.'s account in the Appendix.

*(11.4) Constitutive vs. regulative uses of concepts or judgments*

Constitutive = (i) the use is not conditional or dependent on any assumptions, and (ii) the use implies the objective reality of the relevant concept or the truth of the relevant judgment.

Regulative = (i) the use is dependent on some subjunctive agnostic assumption about a certain judgment: we cannot legitimately assert or legitimately deny that the judgment is correct, and (ii) under this assumption, the regulative use implies only the subjective necessity of *cognizing or acting as if* we believed the relevant judgment.

Compare & contrast this version of the constitutive vs. regulative distinction with the version given at CPR: A179-180/B221-222, where it is mapped onto the earlier distinction between mathematical & dynamical principles at CPR: A160/B199.

*(11.5) There are no legitimate **constitutive** uses of the IPRs*

This is shown by the dialectic of pure reason: paralogisms, antinomies, and the ideal of pure reason.

*(11.6) There are some legitimate **regulative** uses of the IPRs*

Theoretical uses:

(i) Scientific: inferences to the best explanation in natural science that use IPUs as heuristic guides for inquiry = nature must be *cognized as if we believed that* it is systematically & coherently & universally lawlike, for the purposes of effective progress in natural science.

(ii) Transcendental: subjective a priori necessity = nature must be *cognized as if we believed that* it were designed to conform to our cognitive faculties.

Practical uses:

(i) God & Immortality: we must always *act as if we believed that* God had arranged things so that all & only the morally virtuous people are happy, & also that all & only the wicked people (i.e., all of us in our current condition of self-incurred rational immaturity, i.e., our current unenlightened, radically evil condition) must face up to their wickedness & try to reverse their wicked ways and turn them into morally good lives in an endless life (lest we become skeptics about the very idea of a complete good for persons). This is what *believing-in* God really means for K.

(ii) Freedom: we must always *act as if we believed that* we possess both transcendental freedom & also practical freedom or autonomy (lest we become hard determinists & skeptics about moral responsibility & the categorical imperative). This is what *believing-in* free will really means for K.

Upshot: Via the regulative use of the IPRs, beyond the exact sciences, human reason systematically bootstraps itself into cognitive & practical success. Thus we must deny **scientific knowing** (*Wissen)* in order to make room for **faith** *(Glauben)* (CPR: Bxxx).

NB. 'Glauben' also means *belief*, and Kant is of course a moral rationalist. So other more accurate although less colloquial translations of 'Glauben' here would be **moral belief** or **moral faith** or **rational faith**.

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 12

Covers: CPR: 459-469 = A405-425 (Antinomy of Pure Reason), 484-489 = A444-451/B472-479 (Third Antinomy), 532-550 = A532-567/B560-595 (Resolution of the Third Antinomy).

## The Third Antinomy, Freedom, & Determinism

*(12.1) The Context of the Third Antinomy.*

The Antinomy of Pure Reason follows the Paralogisms of Pure Reason, and continues the job of transcendental dialectic: the logical diagnosis of our natural tendency to pure rational metaphysical (necessary) falsity—including paradox, fallacy, and illusion.

According to Kant, we cannot in fact ever entirely remove this kind of illusion precisely because it is *natural* for rational humans= animals; nevertheless we can come to terms with it by self-reflectively exposing the metaphysical error that underlies it, and learning from our mistake, even though we can't help making that mistake. But there are two crucial differences between the Paralogisms and Antinomies.

**First**, whereas the Paralogisms investigates pure rational metaphysical illusion concerning the subject of cognition, the thinking *subject*, the Antinomy investigates pure rational metaphysical illusion concerning the *object* of cognition: that is, the totality of appearances. We know from our study of the Analogies of Experience that this totality of appearances is equivalent to the empirical world, or nature, and that nature is a dynamical system of events in space and time governed by strict deterministic causal natural laws.

**Second**, whereas the Paralogisms exposes a basic fallacy in metaphysical reasoning about the thinking subject (roughly, that the fact of self-consciousness or apperception entails the existence of a Cartesian thinking substance), the Antinomies expose a paradox or hyper-contradiction in our metaphysical reasoning about the object of cognition.

*(12.2) What an Antinomy is, How It Can Be (Dis)solved, and What Can Be Learned from it.*

An antinomy is no ordinary inconsistency or contradiction (= a judgment that is, or entails, a judgment of the form '*P* and ~ *P*'). It is, rather, a paradox or hypercontradiction: on the assumption of the thesis, *P*, a contradiction can be derived; and on the assumption of the antithesis, ~ *P*, another contradiction can be derived. Hence the defender of the thesis can "prove" his claim by a *reductio ad absurdum* argument on the antithesis; and

the defender of the antithesis can "prove" her claim by the same *reductio* strategy as applied to the thesis.

Kant believes that there are four basic forms of the Antinomy, corresponding to the four sets of categories: (1) quantity, (2) quality, (3) relation, and (4) modality.

The first Antinomy concerns the question as to whether the world is finite in time and space (thesis) or infinite (antithesis).

The second Antinomy concerns the question as to whether the world is made of ultimate atoms (thesis), or is infinitely composite (antithesis).

The third Antinomy deals with the question as to whether nature includes some spontaneous causes or freedom (thesis), or includes no such causes because it is completely determined by the laws of nature (antithesis).

And the fourth Antinomy covers the question as to whether the world includes or has as its cause a necessary being (thesis), or whether it neither includes nor has as its cause a necessary being (antithesis).

For K., the logical clue to the (dis)solution of the antinomy lies in the logical distinction between "contradictories" and "contraries." Both are forms of inconsistency. But whereas two contradictories cannot both be false and cannot both be true (i.e., one of them must be true), two contraries cannot both be true but both *can* be false.

For example, "All *A*s are *B*s" and "Some *A*s are not *B*s" are contradictories, but "All *A*s are *B*s" and "No *A*s are *B*s" are contraries (= they can both be false if some *A*s are *B*s and some *A*s are not *B*s).

In each form of the Antinomy, what we discover is that the thesis and antithesis are really *contraries*, not contradictories. For in each case we discover that both thesis and antithesis share a false presupposition: both sides falsely presuppose that there is no distinction between phenomena and noumena, hence both sides falsely presuppose that they must apply their principles to the *same* domain of substances or properties. But at least in principle they could still each apply to *different* domains of entities: phenomena *or* noumena. Hence the Antinomy is not a genuine paradox after all.

Kant's interest in the Antinomy is only methodologically skeptical, because he wants to disclose, by a negative route, some a priori truths about the world or nature. This is what he calls the "transcendental solution" to the Antinomy. The key to the transcendental solution of the Antinomy, not too surprisingly, is transcendental idealism. Each antinomy can then be positively analyzed in terms of transcendental idealism. But the first and third Antinomy stand apart from the others for the following reason: transcendental idealism points up a way in which both the thesis and antithesis can be re-interpreted as to come out true (mutually compatible or mutually consistent).

*(12.3) The Third Antinomy and its Transcendental Solution.*

(12.31) Some stage-setting.

The Third Antinomy is crucially constrained by two factors.

**First**, whatever Kant has to say about solving this version of the Antinomy, the three Analogies of Experience ("in all change of appearances substance persists, and its quantum is neither increased nor diminished in nature," "all alterations occur in accordance with the law of the connection of cause and effect" and "all substances, insofar as they can be perceived in space as simultaneous, are in thoroughgoing interaction"), which tell us about the nature of causation in the natural empirical world of possible experience, must all come out true.

**Second**, whatever Kant has to say about different types of causation, there must be a level of generality at which the concept of causation is univocal. In this connection we will remember that the general schematized pure concept of causation for Kant is that something *X* (the cause) necessitates something else *Y* (its effect) in time according to a necessary rule or law. Or equivalently, to say that *X* causes its effect *Y* is to say that *X* is nomologically sufficient for *Y* in time.

But this general schematized notion of causation allows for at least two distinct sub-concepts of causation. On the one hand, there is the concept of an *absolutely spontaneous* cause, and on the other hand there is the concept of a *naturally deterministic* cause.

(Strictly speaking, one could also postulate the notion of an *naturally indeterministic* cause, whose effects are brought about as the mathematical output of aggregated natural facts according to probabilistic or statistical laws. Since the very idea of a systematic or nomological science of probability is a 19th & 20th century invention, Kant would simply have assumed, I think, that the very notion of chance, as non-nomological, logically excludes the notion of a cause.)

In any case, the concept of an absolutely spontaneous cause depends on Kant's general notion of the spontaneity of a mental act or operation. For Kant, *X* is spontaneous if and only if *X* is a conscious mental event which expresses some acts or operations of a creature, and *X* is

> (i) causally and temporally *unprecedented*, in that (ia) those specific sorts of act or operation have never actually happened before, and (ib) antecedent events do not provide fully sufficient conditions for the existence or effects of those acts or operations,
> (ii) *underdetermined* by external sensory informational inputs, and also by prior desires, even though it may have been triggered by those very inputs or motivated by those very desires
> (iii) *creative* in the sense of being recursively constructive, or able to generate infinitely complex outputs from finite resources, and also

(iv) *self-guiding*.

Now there is also an important difference between *relative* and *absolute* spontaneity. A mental act or operation is relatively spontaneous if and only if it is spontaneous (as defined above) and it must always be triggered into activity or operation by *given* inputs. But a mental act or operation is absolutely spontaneous if and only if it can also generate its *own* inputs. This distinction is important because all *cognitive* or *theoretical* spontaneity for Kant (e.g., the spontaneity of the understanding, or apperception) is only relative, not absolute. But *practical* spontaneity can be absolute—Kant thinks, e.g., that practical freedom can generate its *own* motivating desire and feeling, which Kant calls "respect" (*Achtung*).

Now combining the notion of an absolute spontaneity with Kant's general schematized notion of a cause as a nomologically sufficient condition for its effect in time, it follows that according to him, *X* is an absolutely spontaneous cause of its effect *Y* if and only if (1) *X* is nomologically sufficient for *Y* in time, and (2) *X* is a mental act or operation that is absolutely unprecedented, underdetermined by external sensory inputs and desires, creative, and self-guiding.

In turn, absolutely spontaneous causation is the same as *transcendental freedom*:

> By freedom in the cosmological sense … I 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. Freedom in this signification is a pure transcendental idea. (*CPR* A533/B561)

It should also be noted here that the very idea of an absolutely spontaneous mental act or operation, and thus transcendental freedom, smoothly implies the existence of *a causally empowered substance* or *causally efficacious agent* or *agentive source* which or who acts or operates freely. So transcendental freedom for Kant, as applied to the human will, implies causally efficacious rational intentional agency or personhood.

On the other hand, the concept of naturally deterministic causation is sharply distinct from the concept of transcendental freedom. According to Kant, *X* is a naturally deterministic cause of its effect *Y* if and only if

> (1) *X* is nomologically sufficient for *Y* in time,
>
> (2) the law under which *X* and *Y* both fall is a causal-dynamic natural law,
>
> (3) *X* and *Y* are either simple events or complex events in asymmetric time,
>
> (4) *Y* cannot precede *X* in time (hence either *Y* follows *X* in time or *Y* is simultaneous with *X*),
>
> (5) *X* and *Y* are material substances or parts of material substances,

(6) *X* is itself the effect of an earlier cause *Z1*, which in turn is the effect of an earlier cause *Z2*, and so-on indefinitely backwards in time, and

(7) from the existence of the causdal-dynamic natural laws together with the actual existence of all the simple or complex events prior to *X*, not only *Y* but also every other future simple or complex event in nature follows with metaphysical necessity from *X*.

In short, the causal metaphysical framework described by the three Analogies of Experience is precisely that of naturally deterministic causation. So much for the conceptual stage-setting. We are now in a position to reconstruct the Third Antinomy.

(12.32) Reconstruction of the Third Antinomy

**Thesis:** "Causality in accordance with laws of nature is not the only one from which all the appearances of the world can be derived. It is also necessary to assume another causality through freedom in order to explain them" (*CPR* A444/B472). In other words, naturally deterministic causation is not the only kind of causation and transcendental freedom therefore exists.

(1) Suppose that there is only naturally deterministic causation.
(2) If (1) is true, then every simple or complex event is necessitated according to a natural law by some earlier simple or complex event, and that earlier simple or complex event is in turn nomologically necessitated by an earlier one, and so on ad infinitum.
(3) But if (2) is true, then there is never a first beginning to the series of causes of a given simple or complex event, hence never a complete nomologically sufficient condition for that event. But that is absurd, since the very idea of a naturally deterministic cause is that it is the nomologically *sufficient* condition of the simple or complex event which is its effect.
(4) Therefore, by reductio, (1) is false, and we must assume the existence of an absolutely spontaneous cause, transcendental freedom, as the nomologically sufficient condition of every naturally deterministic causal series. **QED**

In other words, the argument for the **Thesis** says that naturally deterministic causation violates the sufficiency condition of the general schematized concept of a cause (= a nomologically *sufficient* condition in time).

**Antithesis:** "There is no freedom, but everything in the world happens solely in accordance with the laws of nature" (*CPR* A445/B473). In other words, there is only naturally deterministic causation and transcendental freedom does not exist.

(1) Suppose that transcendental freedom exists.
(2) If (1) is true, then the nomologically sufficient condition of every naturally deterministic causal series itself has no cause.
(3) But if (2) is true, then transcendental freedom does not itself fall under any

> laws of nature and is a law unto itself (that is, a miracle) operating by totally uncognizable (that is, occult) means. But that is absurd, since the very idea of a naturally deterministic cause is that it is the *nomologically* sufficient condition of the simple or complex event which is its effect.
> (4) Therefore, by reductio, (1) is false: there is only naturally deterministic causation and transcendental freedom does not exist. **QED**

In other words, the argument for the **Antithesis** says that transcendental freedom violates the nomological condition of the general schematized concept of a cause (= a *nomologically* sufficient condition in time).

(12.33) Kant's Transcendental Solution for the Third Antinomy

We will remember that according to Kant every Antinomy is diagnosed and dissolved by distinguishing sharply between noumena and phenomena. In this light, as I mentioned above, the shared error of Thesis and Antithesis in each case is that both fail to distinguish between noumena and phenomena and falsely assume that their principles apply to a single undifferentiated domain of substances or properties.

As I also mentioned above, the Third Antinomy has a reconciliation phase in which the recognition of the distinction between noumena and phenomena allows for a re-interpretation according to which the Thesis and the Antithesis both come out true: the Thesis applies to noumena only (hence transcendental freedom is a noumenal cause), and the Antithesis applies to phenomena only (hence the phenomenal world is naturally determined).

Why does Kant undertake this reconciliation? One important reason is that Kant is assuming for the purposes of the Third Antinomy that the three Analogies *are* true. And as I noted earlier, the concept of causation contained in the Analogies is equivalent to the concept of causation contained in the Antithesis, the concept of naturally deterministic causation. The only salient difference between the two presentations of that concept is that in the Analogies, it is specifically restricted to the domain of phenomena or appearances, whereas in the Antithesis of the Third Antinomy, at least initially, it is allowed to range ambiguously over the domains of phenomena and noumena alike.

But another even more important reason for the reconciliation phase is that Kant thinks that morality is impossible without the concept of practical freedom, which is negatively defined as the ability to choose independently of all sensory impulses or empirical desires, and positively defined as autonomy or self-legislation according to the moral law or Categorical Imperative.

Now as Kant argues in the third section of the *Grounding for the Metaphysics of Morals*, the concept of the moral law or Categorical Imperative reciprocally entails the concept of practical freedom or autonomy, and practical freedom presupposes transcendental freedom. So without the reconciliation phase, morality itself would be undermined. Here are two crucial texts.

> It is this transcendental idea of freedom on which the practical concept of freedom is grounded …. 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)

> 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)

But here is where things get (even) murkier. He wants to claim not only that transcendental freedom and naturally deterministic causation are formally or logically-analytically consistent with one another, *but also that it's really or metaphysically possible for them to apply to the very same natural phenomenal events*, considered as the effects of *each* cause individually and *both* causes together. Here is a text along those lines:

> It is easy to see that if all causality in the world of sense were mere nature, then every occurrence would be determined in time by another in accord with necessary laws, and hence—since appearances, insofar as they determine the power of choice, would have to render every action necessary as their natural consequence—the abolition of transcendental freedom would simultaneously eliminate all practical freedom. For the latter presupposes that although something has not happened, it nevertheless **ought** to have happened, and its cause in appearance was thus not so determining that there is not a causality in our power of choice such that, independently of those natural causes and even opposed to their powre and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of appearances **entirely from itself.** (CPR: A534/B562, underlining added)

The idea is this. One and the same thing—e.g., a rational human animal intentionally acting in a certain way—can be *both* transcendentally freely caused to do what it does and *also* naturally deterministically caused to do what it does. If the transcendentally free

cause had not existed, then the naturally deterministic cause would still have brought the action about; if the naturally deterministic cause had not existed, then the transcendentally free cause would still have brought the action about. So K. seems to be asserting the real or metaphysical-synthetic possibility of *the systematic causal overdetermination of all human intentional actions*.

K. regards this proof of the formal consistency + real or metaphysical possibility of transcendental freedom as absolutely necessary for the possibility of morality. Morality involves transcendentally free causation from moral ideas and laws of pure practical reason alone. If every phenomenon of rational human action can be consistently regarded as both the result of some noumenal transcendentally free cause and also as the result of some phenomenal naturally deterministic cause, then at least some of our real willings and actions can be regarded as produced by our pure practical reason according to the law of the categorical imperative and not merely by natural determined causes. So the transcendental solution for the third Antinomy provides a necessary segue to Kant's practical philosophy.

Here is a *very* quick sketch of the concept of freedom in Kant's practical philosophy.

In *Groundwork for the Metaphysics of Morals*, section III, Kant argues for four basic theses about freedom:

> **(i)** that the concept of positive freedom is necessarily equivalent with the concept of autonomy, and both of them analytically entail the Categorical Imperative or CI,
>
> **(ii)** that the thesis (which K also calls "the principle of morality") that a good will is volition from (for the sake of ) duty, which is the same as to obey the moral law or CI, is a necessary synthetic proposition whose necessity can be explained only by appealing to positive freedom or autonomy,
>
> **(iii)** that persons or rational agents necessarily act only under the pure rational concept or Idea of their own positive freedom or autonomy, and
>
> **(iv)** that positive freedom is logically and metaphysically possible, although scientifically inexplicable.

In the *Critique of Practical Reason*, Kant also adds two crucial factors to his theory of freedom:

(1) the notion of a "fact of reason," which is an agent's direct conscious experience of her practical freedom, and

(2) a distinction between (i) *psychological freedom*, which is a necessary but not sufficient condition of all the other types of freedom, (ii) transcendental freedom (agentive sourcehood), and (iii) practical freedom (both negative = choice that is

independent of alien causes & also from determination by the impulses of sensibility, and also positive = autonomy). Here are the relevant texts.

> 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)

> If these determining representations [i.e., instincts or motives] themselves have the ground of their existence in time and, more particularly, in the antecedent state and these again in a preceding state, and so on…; and if they are without exception internal; and if they do not have mechanical causality but a psychological causality through representations instead of through bodily movements: they are nonetheless determining grounds of the causality of a being insofar as his existence is determinable in time…. Thus these conceptions do indeed imply psychological freedom (if one wishes to use this word for a merely internal concatenation of representations in the mind), but nonetheless they also imply natural necessity leaving no room for transcendental freedom which must be thought of as independence from everything empirical and hence from nature generally, whether regarded as an object of inner sense merely in time or also as an object of outer sense in both space and time…. [A]ll necessity of events in time according to natural law can be called the "mechanism of nature," even though it is not to be supposed that things which are subject to it must really be material machines. Here reference is made only to the necessity of the connection of events in a temporal series as they develop according to natural law, whether the subject in which this development occurs be called *automaton materiale* when the machinery is impelled by matter, or, with Leibniz, *automaton spirituale* when it is impelled by representations. And if the freedom of our will were nothing else than the latter, 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)

# Metaphysics with a Human Face: Lectures on Kant's *Critique of Pure Reason*

**Fall 2014**
**Robert Hanna**
**Lecture Notes**
## LECTURE 13

Covers: CPR: 551-569 = A567-602/B595-630 (The Transcendental Ideal, and the Impossibility of an Ontological Proof of God's Existence).

## The Ideal of Pure Reason, the Impossibility of Ontological Arguments, & How to Deal with the Unprovability of God's Existence (or Non-Existence)

*(13.1) The context of the Ideal*.

Kant's 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 (i) that there cannot be an ontological proof, (ii) that there cannot be a cosmological proof, (iii) that there cannot be a physico-theological proof (i.e., a sound design argument), and (iv) that 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 existence. The cosmological proof, if sound, would yield God's existence as a synthetic a priori truth; and the physico-theological proof or design argument, if sound, would yield God's 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*.

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.

It's also crucial to note that by the same basic Kantian reasoning, it's also impossible to prove that God does *not* exist, since authentic scientific knowledge either of God's existence or *non*-existence is impossible.

So this is Kant's radical agnosticism, as applied to theology: it is a priori known that it is impossible to know whether God exists or does not exist. Theism and atheism are *equally* mistaken.

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 soul (Paralogisms); (2) the Idea of an absolute object of cognition, or nature as a cosmological totality (Antinomies); and (3) the Idea of an absolute ground of both the subject and the object of cognition, or God.

The dialectical error in the Paralogisms was 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 was failing to draw distinction between phenomenal entities and 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.

*(13.2) Ideals, Concepts, Ontology, and God.*

Ideals, according to Kant, are the Ideas of Pure Reason incarnate or reified or hypostatized: they are ideal 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 possibilities." 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 K. 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 K. 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 K., 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 logical Principle of Non-Contradiction (PNC): For all predicates *P* & all objects *x*, necessarily ~ (*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 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 being that contains within its essence all of actuality or reality: hence Kant calls the concept of God the concept of the *ens realissimum*.

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 existence of what is described by that concept (false existence proof). The Ontological Argument falls under the fallacy of false existence proof.

*(13.3) What is the Ontological Argument?*

As I've mentioned already, the ontological argument (OA) is the analytic a priori argument from the concept of God to God's 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' fifth *Meditation*. Here are quick glosses of those two arguments:

*(13.31) 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.

*(13.32) 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.

*(13.4) Kant on the Impossibility of the OA*.

Kant's critique of the OA consists of three distinct parts:

(i) "exists" is a logical and not a determining (a.k.a."real") 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,

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

(iii) objectively valid and true existence-judgments (e.g., "Socrates exists") are synthetic (hence their meaning & 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 separately and then re-combine them into a single thesis about the OA.

Re (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. E.g., 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.") 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*, & 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, *determining (real) predicates* or *determining (real) concepts* are those concepts whose application to another concept does indeed change and augment the semantic content of the second concept. E.g., RED is a determining (real) concept whose application to the concept ROSE modifies the latter's content by further specifying it and also correspondingly narrowing its extension.

Now EXISTS is merely a logical predicate in that applying it to the concept of, say, ONE HUNDRED DOLLARS, doesn't in any way change or augment the latter's semantic content.

Notice that K. 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 a predicate: on the contrary EXISTS *is* a predicate. It is just that it is a logical predicate and not a determining (real) predicate.

*Q*: What more precisely does the logical predicate or 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. So EXISTS is a second-order predicate that functions in essentially the same way as the existential quantifier of first-order predicate logic.

Re (ii): For K., the concept EXISTS is empirically meaningful or objectively valid when it is schematized by the representations of time and space, and 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 (iii): If EXISTS is a logical predicate but not a determining (real) predicate, and if the concept EXISTS is a second-order concept meaning that the concept to which it is applied has instances, 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. Hence "*X* exists" is true if and only if something falling under the concept *X* has empirically intuitable real or actual instances. Any judgment whose whose meaning and truth depend on empirical intuition is synthetic. Hence every objectively valid and true existential judgment is synthetic.

How does this all apply to the OA? In two ways. First, the OA errs by treating the concept EXISTS as if it were a determining (real) predicate. But EXISTS is neither a determining (real) predicate nor is it ever contained analytically in any other determining (real) predicate. Therefore all arguments purporting to show that the concept EXISTS is analytically contained in the concept GOD are bogus and fallacious.

Second, consider the judgment "God exists." It is true just in case GOD has empirically intuitable instances. Hence even if "God exists" were true, that judgment could only ever be synthetic, not analytic.

*(13.5) Two important logico-semantic consequences of Kant's critique of the OA.*

There are two important logico-semantic 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 or non-existence. Both God's existence & God's non-existence are a priori knowably unprovable. Theism and atheism are equally mistaken. Again, this is Kant's radical agnosticism.

**Second**, Kant's critique of the OA also suggests a general solution to a longstanding problem in philsophical 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 has seriously worried Frege, Russell, and many other major philosophical logicians. 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 existential predications are made, the subject-term of the proposition stands for a concept, not an object. And some concepts have a null real-world or actual-world extension, e.g., the 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 the subject concept of the proposition has some real or actual instances; and a negative existential proposition is true just in case the subject concept of the proposition has no real or actual instances.

*(13.6) How to deal with the unprovability of God's existence (or non-existence)*

Kant's critique of the ontological proof also has direct implications for ethics, metaphysics, and the philosophy of religion. We know from the Paralogisms and the Ideal of Pure Reason, both the idea of the human soul and the idea of God are *unknowable Ideas of pure reason*. Correspondingly, both the immortality of the soul and the existence/non-existence of God are *logically unprovable propositions*. Neither their truth nor their falsity can be demonstrated. Hence the correct philosophical attitude to take towards them is *radical agnosticism*.

Generally speaking, subject S is radically agnostic about a proposition *P* if and only if S a priori knows that it is impossible to know *P* and impossible to know not-*P*.

But the ideas of immortality and of God's existence still can have regulative, practical significance as *postulates of pure practical reason*. Here is what Kant says:

> 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 anu 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)

In other and fewer words, what Kant is saying is that our immortality, as an *endless human personal existence*, if it were true, would make a life of moral virtue much easier to pursue. As we all know, in this life, stuff happens all the time and all over the place, including radically wicked acts, no good deed ever goes unpunished, and this world is a vale of tears (see, e.g., Schopenhauer on "the sufferings of the world"); but in an endless life we would always have time enough to choose and do all the right things, and also for everyone to take responsibility for all the wrong things they had chosen and done, to undergo what the *Religion* he calls a "revolution of the heart" or a "revolution of the will," and to change their lives for the morally better.

And God's existence, if it were true, would guarantee that all and only the morally virtuous people would be happy sooner or later.

So we must presuppose immortality and God's existence insofar as we are striving to be autonomous moral agents in this finite human-all-too-human life.

What, more precisely, does Kant mean by 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 a 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 Kant means: He is *not* saying that we are obligated to believe that either personal immortality or God's existence is true. But even though we can never know or even have any adequate logical or scientific reasons to believe that either personal immortality or the existence of God is true, we are nevertheless obliged *to act as if we believed that they were true*. Again: we are *not* obligated to believe—rather, we are obligated *to act as if* we believed, even though we *cannot* rationally believe.

This, I think, is the best way of cashing out the notion of *moral commitment* or *believing-in*, as opposed to *believing-that*. In the Canon of Pure Reason, Kant calls this "**moral belief**" or "**moral faith**" (*moralischen Glauben*) (*CPR* A828-830/B856-858). He also says that this provides us with "moral certainty," as opposed to "logical certainty," i.e., epistemic necessity.

This idea of a thoroughly rational moral commitment to acting as if one believed in personal immortality and God's existence, via the full recognition that neither personal immortality nor God's existence nor God's non-existence can be logically proved or scientifically known, is also closely connected to Kant's famous remark in the B Preface, with direct reference to the Idea of freedom, that "I had to deny **knowledge** (*Wissen*) in order to make room for **faith"** (*Glauben*) (*CPR* Bxxx), in the Guyer-Wood translation.

But a philosophically more accurate (if less colloquial) translation of that remark would have been: "I had to deny **scientific knowing** in order to make room for **moral faith**."

This interpretation of the implications of Kant's doctrine of the unprovability of God's existence leads to a doctrine I will call *Existential Kantian Theology* or EKT. EKT can be most clearly defined in relation to another view I will call *Hard Secularism*.

*Hard Secularism* says that morality and political principles have mechanistic and reductive materialist foundations, knowable by means of the natural sciences, and that all attempts to hold some alternative thesis about the foundations of morality or political principles must imply some or another version of a wholly implausible theological metaphysics, and some or another version of the wholly implausible Divine Command Theory of Morality—which says that moral principles are dictated by God, and are true only because God dictates them.

EKT by sharp contast, as I am understanding it, says that morality has irreducibly rational human foundations, but *not* theological foundations, and correspondingly that political principles have irreducibly moral foundations, *not* theological foundations. Therefore EKT is as far from any theological metaphysics and the Divine Command Theory of Morality as Hard Secularism is. At the same time, however, EKT is fully rationally open to a positive and sympathetic reading of *the moral import of religion for autonomous human moral agents or persons*, which is essentially the same as what Kierkegaard also called "the ethical" when it is fused with "the religious":

> 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.[1]

Very briefly put, according to EKT, for someone to believe-in God's existence *is for her to believe that her life has a meaning by virtue of its categorically normative moral content*, via her 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. Whether God actually exists nor not is completely irrelevant to this. Indeed, it is strictly logically unprovable and scientifically unknowable whether God exists or not. So precisely the right attitude to take towards the question of God's existence or non-existence is radical agnosticism—which means not only *believing* that God's existence or non-existence is strictly logically unprovable and scientifically unknowable, but also *acting as if you believed that* God's existence or non-existence is strictly logically unprovable and scientifically unknowable. This attitude begins as a fundamental "loss of faith" or anxiety (*Angst*), 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,[2] and also as in Kant's equally sublime "I had to deny **knowledge** (*Wissen)* in order to make room for **faith** (*Glauben*)," when interpreted in the way I did just a few paragraphs above.

According to EKT, then, what ultimately matters, then, is actively believing-in the real possibility that your life has a meaning and categorically normative moral content, via radical agnosticism about about God's existence or non-existence. If a rational human agent or real human 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, 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. By acting and living well under the pure practical postulates of God's existence and personal immortality, you carry out a *constructive counterfactual proof* of God's existence—but not a logically *demonstrative proof*, which is impossible: You construct for yourself exactly the same sort of life you *would have* if God really *were* to exist. But a priori you know that you cannot know whether God exists or does not exist. So it's ultimately all up to you, to each one of us, and to all of us together. In my opinion, this is actually the same basic philosophical message as Pascal's so-called "Wager." But that's another story for another day.

---

[1] 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.

[2] Kierkegaard, "Fear and Trembling," in *The Essential Kierkegaard*, pp. 93-101.