

THE PHILOSOPHICAL LIMITS OF SCIENTIFIC ESSENTIALISM¹

George Bealer
Reed College

Part 1: The Autonomy of Philosophy

1. Will Science Eclipse Philosophy?

Philosophers are typically interested in such questions as the nature of substance, mind, intelligence, perception, knowledge, wisdom, truth, explanation, causation, freedom, purpose, goodness, duty, the virtues, love, life, and happiness. Typically, they are not interested in such questions as the nature of water, heat, lemon, arthritis, or the distinction between beech and elm. These questions belong to empirical science, and philosophers are not, as philosophers, typically interested them. In being interested in such things as the nature of mind, intelligence, the virtues, and life, philosophers do not want to know what those things just happen to be, but rather what those things *must* be. It is not enough that the virtue of piety happened to be what Euthyphro exhibited; a philosopher wants to know what piety *must* be. It is a requirement on a satisfactory philosophical theory that it should hold necessarily. A philosopher wishes to know what mind, life, the virtues, and so forth really *are* in a strong sense: what they *have* to be, not just what they happen to be.

Since the time of Plato, philosophers have nearly all agreed that, at least in most cases, the answer to a philosophical question should be a necessary not contingent truth. Plato and Aristotle in ancient philosophy and Hume and Kant in modern philosophy have taught

us that experience alone cannot tell us whether a given truth is necessary. Of course, by relying on the testimony of an authority, one can learn that a given truth is necessary. But how, ultimately, do the original authorities learn it? By examining experience, the original authorities can learn merely that something is so; they cannot learn that it is necessarily so. As Kant says in the introduction to his *Critique of Pure Reason*, "Experience tells us, indeed, what is, but not that it must necessarily be so, and not otherwise." No amount of experience (waiving testimony) can ever reveal whether something that is so is necessary. According to Plato, Aristotle, Kant, and most other philosophers, this kind of *modal* knowledge ultimately demands another source beyond experience, namely, reason.

Knowledge that a given truth is necessary demands that the original authorities use reason. But does their knowledge always rely on reason alone, without the use of any empirical evidence or justification? Kant seems to think so: "If, then, a judgment is thought with strict universality, that is, in such a manner that *no exception is allowed as possible*, it is not derived from experience, but is valid *absolutely a priori*." (B4, my emphasis) This Kantian thesis is certainly too strong. An original authority could come to know that a given proposition is necessary as follows. First, the person could use reason alone to know that propositions of that type (for example, mathematical propositions) are necessary if true. Second, the person could use experience (for example, counting, measuring, or using a computer) to discover that the proposition is true. Then, by combining the first bit of *a priori* knowledge and the second bit of *a posteriori* knowledge, the person can infer that the proposition is necessary. So, contrary to Kant's thesis, this person's knowledge that the proposition is necessary would not be *absolutely a priori*; its justification is at least partly empirical.

To avoid this kind of counterexample, a philosopher might be tempted to modify the strong Kantian thesis as follows: if it is possible to know that a given proposition is necessary, then it is *possible* to know this *absolutely a priori*; no empirical evidence is needed. For example, if it is possible to know that a certain mathematical proposition is necessary, then it is possible to know this without any empirical evidence (counting, measuring, testimony, etc.); reason alone suffices. (It is reasonable that this modified thesis is what Kant really wanted to defend.) The relevance of the modified Kantian thesis for philosophy is plain. Philosophers might from time to time rely

on empirical science to reach their conclusions that various philosophical truths are indeed necessary. However, such reliance on empirical science is inessential if the modified Kantian thesis is right. It is always possible to obtain our modal philosophical knowledge absolutely *a priori*. If the modified Kantian thesis is right, philosophy is in principle independent of the empirical sciences.

The modified Kantian thesis—and its corollary that philosophy is independent of the empirical sciences—has had nearly universal acceptance in the history of philosophy. (Certain Aristotelians have perhaps been dissenters.) In recent years, however, the modified Kantian thesis has come under severe attack from scientific essentialists (Matson, Putnam, Kripke, Shoemaker, Boyd, Devitt, and a great many others).² *Scientific essentialism* is the doctrine that there are truths that can be known to be necessary only if empirical evidence is used; more briefly, there are essences that can be known only with the aid of empirical science. For example, scientific essentialists hold that certain types of natural-kind identities (e.g., that water = H₂O, that heat = mean kinetic energy, that gold = the element with atomic number 79) are necessary if true, and they hold that it is impossible for a person to know that such identities are true without the aid of empirical science. So if scientific essentialists are right, the knowledge that such identities are necessary cannot be absolutely *a priori*, as the modified Kantian thesis requires. Instead, this knowledge must always be (at least) partially *a posteriori*; it is essentially *a posteriori*.³

Let us suppose scientific essentialism is correct; that is, let us suppose that there are genuine counterexamples (water, heat, gold, etc.) to the modified Kantian thesis. There is nevertheless a natural weakening of this thesis that preserves the view that empirical science cannot eclipse philosophy: for (most of) the central truths of philosophy, if it is possible to know these truths to be necessary, it is possible to know them to be necessary absolutely *a priori*, without the aid of empirical science. This is *the thesis of the autonomy of philosophy*.⁴ (This thesis is vague; later on I will replace it with something more precise.)

Although scientific essentialism contradicts the modified Kantian thesis, it does not contradict the thesis of the autonomy of philosophy. For scientific essentialism asserts only that there are *some* truths (e.g., that water = H₂O, etc.) that can be known to be necessary only with the aid of empirical science. It does not assert that *all* knowable

necessary truths can be known to be necessary only with the aid of empirical science. Indeed, I should think that most scientific essentialists recognize that this radical view is inconsistent with a fully elaborated presentation of their position. True enough, a person could come to know, say, that necessarily water = H₂O absolutely *a posteriori*, for example, exclusively on the basis of the testimony of others. However, the chain of testimony cannot go on forever; one of the authorities must have come to know that necessarily water = H₂O by another route. Scientific essentialists reconstruct this other route along the following lines. (Of course, there are variations, but none of them alters the point at issue, namely, that there are some truths that can be known to be necessary absolutely *a priori*.) First, the authority in question uses reason alone to establish that propositions of the type in question—namely, certain sorts of natural-kind identities—are necessary if true. Second this person uses experience (namely, various laboratory experiments) to establish that the specific proposition in question is true. Then, by combining the first bit of a *priori* knowledge and the second bit of a *posteriori* knowledge, the person infers that the proposition about the essence of water is necessary. As Kripke says, “*Philosophical analysis* tells us that they [i.e., propositions of the type in question] cannot be contingently true, so any empirical knowledge of their truth is automatically empirical knowledge that they are necessary. This characterization applies, in particular, to the cases of identity statements and of essence.” (P. 159, *Naming and Necessity*, emphasis added.)

Now consider the first bit of a *priori* knowledge, namely, the knowledge of the philosophical proposition that certain types of natural-kind identities are necessary if true. By performing a philosophical analysis very similar to that which led to this bit of a *priori* knowledge, we can obtain another bit of a *priori* knowledge, namely, that philosophical propositions like the original one are themselves necessary if true. So given that we can know *a priori* that the original philosophical proposition is true, we can know *a priori* that it is necessary. The upshot is that the radical generalization of scientific essentialism is inconsistent: at least one of the philosophical propositions used in the fully elaborated presentation of scientific essentialism can be known to be necessary without the aid of empirical science. (This little argument is not crucial to my larger argument for the autonomy of philosophy; I include it only

for heuristic purposes.)

There is, however, a less radical generalization of scientific essentialism that has enormous appeal to philosophers and philosophically inclined scientists. It is that *most* of the central truths of philosophy cannot be known to be necessary (and, indeed, cannot be known) without the aid of empirical science. That is, most philosophical knowledge must rely on empirical evidence. Accordingly, the thesis of the autonomy of philosophy would be false; philosophy would have no significant independence from empirical science. This doctrine is *global scientific essentialism*, and it is contrasted with *local scientific essentialism*, the far weaker doctrine that certain truths about naturalistic items (water, heat, gold, etc.) cannot be known to be necessary absolutely *a priori*.

(In fact, philosophers have advocated a spectrum of different positions between global and local scientific essentialism. Hilary Putnam once seemed inclined toward the global position. Most contemporary materialist philosophers of mind go well beyond local scientific essentialism, for they think empirical science is the authority in determining the essences of mental qualities and relations such as pain, perception, knowledge, and the like. On this score, at least, Kripke tends more toward local scientific essentialism, for he thinks that pain has no material essence and, therefore, that pain has no material essence that only the empirical sciences could discover.)

Global scientific essentialism impinges upon traditional philosophical inquiry not only in specific theoretical issues. It also significantly affects traditional philosophical method. For example, in the course of philosophical dialectic one would often advance hypothetical examples (usually counterexamples). One's *a priori* intuition that the example is possible (i.e., not necessarily not the case) would usually suffice to establish that it is possible. However, global scientific essentialism invalidates this traditional procedure. For example, suppose that in an argument in metaphysics (or philosophy of biology) you propose the hypothetical example of life forms (i.e., living beings) whose bodies are composed not of hydrocarbons but of something else. With the advent of scientific essentialism, your example might now be dismissed as follows: "You are not justified in accepting that example as genuinely possible. For all you know, you are mistaking a mere epistemic possibility for a genuine metaphysical possibility. Only a genuine metaphysical possibility will win your point in this metaphysical dialectic. But to determine that

your example is metaphysically possible you must first consult empirical science (specifically, biology), for empirical science might reveal that your example is not really possible." (According to Kripke's explication of the notion of epistemic possibility,⁵ when we have an intuition of a mere epistemic possibility that life forms could have nonhydrocarbon bodies, this amounts to the following intuition: that it is metaphysically possible that someone in qualitatively the same epistemic situation as we are in could assert something true by saying, "There are life forms that have nonhydrocarbon bodies." Now the point is supposed to be that this epistemic intuition tells us nothing about *life*, for the person uttering the word 'life' might not mean life at all but something else instead. However, in your original metaphysical dialectic, *life* is what you were interested in.) Notice, however, that this way of dismissing candidate hypothetical examples can be attempted at virtually every turn in philosophical inquiry. Thus, if global scientific essentialism were true, a massive revolution in philosophical method would be in order.

So will science eclipse philosophy? I hold that global scientific essentialism is demonstrably false and that the autonomy of philosophy is not threatened by scientific essentialism. My reason for thinking this is broadly "transcendental": local scientific essentialism actually implies the thesis of the autonomy of philosophy. In the idiom of Kant, the autonomy of philosophy is one of the "conditions of the possibility" of local scientific essentialism.

The arguments for these bold claims comprise a good part of a book, which is now underway. In the present paper my purpose is to present the arguments in broad outline. Before coming to the arguments proper, I will introduce some handy terminology and give some examples that help to motivate the view intuitively.

2. *Category and Content Concepts*

For convenience of exposition I will freely use the idiom of concepts; at most points we could paraphrase away talk of concepts. For example, rather than saying that I believe that the concept of a compositional stuff applies to a certain item in a given hypothetical case, we could say simply that I believe that the item in the case is a compositional stuff. And rather than saying that I possess the concept of being F, we could say that I stand in some basic intentional relation (belief, desire, decision, memory, perception, etc.) to

the proposition that ...F... (i.e., to some proposition involving the concept of being F).

Now to state my "transcendental" argument, I will find it convenient to restate my claim in a slightly more technical vocabulary. For the purposes at hand, we may divide our concepts into three general types: naturalistic concepts, category concepts, and content concepts. (At the initial stage of our investigation, these three types should be viewed as "dummy types": no precise definition of them need be ventured, and the boundaries between them may be left deliberately vague. Nevertheless, they can be characterized to some extent.) Naturalistic concepts are intended to be those concepts for which local scientific essentialism holds; they include, for example, the concepts of water, heat, gold, lemon, arthritis, beech, elm, and so forth. Category and content concepts are intended to include the concepts in terms of which (most of) the central questions of philosophy are formulated. Examples of category concepts are the concepts of stuff, compositional stuff, functional stuff, substance, quality, quantity, action, artificial, natural, cause, reason, person, etc. Examples of content concepts are familiar phenomenal qualities (pain, itchiness, tingling-sensation, etc.) and basic mental relations (knowing, perceiving, deciding, loving, etc.).

In this idiom my claim may be stated as follows. If scientific essentialism is true of various necessities involving naturalistic concepts, then it cannot hold for necessities involving *exclusively* category and content concepts. That is, if local scientific essentialism is true, then knowable necessities involving *exclusively* category and content concepts are knowable absolutely *a priori*. Given that (most of) the central questions of philosophy are formulated *exclusively* in terms of category and content concepts (this of course needs to be shown), it follows that, if local scientific essentialism is true, so is the thesis of the autonomy of philosophy. If a philosophical proposition can be known to be necessary, it is possible for this knowledge to be obtained absolutely *a priori*; empirical science is in principle never needed.

Before coming to my "transcendental" argument, I will give a few examples to motivate our position. The argument does not ride on these examples. But I think they show that something important is missing from the picture scientific essentialists have provided so far, and they suggest that what is missing is a satisfactory philosophical theory of our category and content concepts.

Let us recall Putnam's twin-Earth argument. Our chemists examine samples of water here on Earth and discover that all such samples are composed of H_2O . From this they conclude (correctly, we may suppose) that all and only samples of water here on Earth are composed of H_2O . Our astronauts visiting on twin Earth discover samples of a stuff macroscopically indistinguishable from water; however, these samples are composed of XYZ, not H_2O . The question is whether the samples on twin Earth are really samples of water or of something else. The scientific essentialists' intuition is that they are just *not* samples of water. Generalizing from examples like this, scientific essentialists conclude that, necessarily, all and only samples of water are composed of H_2O . This conclusion, to repeat, is based on the intuition that, although the samples on twin Earth are macroscopically like samples of water, they do not, strictly speaking, count as samples of water because they are not composed of H_2O .

Now I wish to consider an analogous example involving England, which, as you know, is well known for its cuisine. Suppose that, at a certain time, the only food in England was mutton stew,⁶ and all mutton stew in England was food. As you know, mutton stew is composed of boiled mutton, boiled potatoes, and boiled turnips: MPT. Now voyagers from England traveled to twin England down south, where they discovered fettucini, veal scallopini, and even bread, which are composed of things like wheat, cream, garlic, and so on; not MPT at all. Should they phone home and say, "Lots of tasty, nutritious things to eat here; but sorry, no food"? Because after all, no MPT. Of course not. But why is this case different from the twin-Earth case?

The answer must be something like this. Unlike water, which is a compositional stuff, food is a functional stuff.⁷ In the twin-Earth case we know that water is a compositional stuff. This knowledge together with our general categorial intuitions about compositional stuffs determines our intuition that, if all and only samples of water on Earth are composed of H_2O and if H_2O differs from XYZ, then the samples of XYZ on twin Earth are not samples of water. In the twin-England case we know that food is a functional stuff. Food has a function; food is *for* eating. (Or perhaps food is for nutritious eating or just for nutrition. The question is very subtle and complex, but we need not settle it here.) To study what food is, one does not investigate its composition directly; instead one investigates its func-

tion. (Of course, to study eating and/or nutrition one might eventually want to go to a chemist to determine how this function can be fulfilled in special cases.) It is for this reason that the twin-England example, unlike the twin-Earth example, does not lead one to conclude that, necessarily, all and only food is composed of mutton stew (MPT) whereas in the case of water the discovery that all and only water on Earth is composed of H_2O does lead one to conclude that, necessarily, all and only water is H_2O . In the twin-England example, fettucini, scallopini, bread, and so forth fulfill the function of food; therefore, even if their composition differs from that of mutton stew, they count as food. The difference, therefore, turns on knowledge of categories, specifically, that water is a compositional stuff and, by contrast, food is a functional stuff.

One can make the same point concerning food with a more serious example. For example, suppose our chemists discover that all and only samples of food of the kinds we eat are composed of certain complex hydrocarbons with various trace elements, and suppose our space cadets travel to twin Earth and find some stuffs, not composed of those chemicals C_1, \dots, C_n , but instead composed of alternate chemicals K_1, \dots, K_m . Suppose, finally, that these stuffs actually have great nutritional value and that they taste very good and are very digestible. We would not phone home and say, "No food here." There is no doubt that these stuffs would be food. They just would be different *kinds* of food than those we have on Earth. However, XYZ is not a different kind of water; it is not water at all. The explanation is that food is a functional stuff: the various twin-earth stuffs have the relevant function. By contrast, water is a compositional stuff: water is H_2O ; XYZ is something else.⁸

Now, examples like this can be multiplied indefinitely. Consider fuel. There was a time when the only available fuels were solid rapidly oxidizable hydrocarbons. But this fact did not lead our ancestors to discount petroleum as fuel when this *liquid* was discovered and burned, nor did it prevent twentieth century technologists from counting plutonium as fuel even though it is *not* rapidly oxidizable. The explanation is that fuel, unlike water, is a functional stuff; it is for something. The reason we count wood, petroleum, and plutonium as kinds of fuel is they have (or are essentially suited to achieving) this function.

Air and earth *might* be similar. (What I have to say about air and earth here is speculative; my other, less speculative comments do

not depend on it in any way.) The air on Earth is primarily nitrogen, carbon dioxide, and oxygen. On twin Earth, by contrast, the air might be composed of something different. Although twin Earthlings breathe the air (and fly airplanes in it), our astronauts tell us, "Do not breathe the air here because it is toxic to us." If this is right, air would seem not to be a straightforward compositional stuff. Instead, whether something is air would have to do with its having an appropriate role in a suitable kind of environmental system (a planet or other proto-ecosystem). If so, air would seem to be a kind of environmental (proto-ecological) stuff. It has a certain role in an environmental system as a whole. Now if air is like this, certainly earth is too: clay, loam, sandy loam, sand, humus, and so forth are all kinds of earth even though they are quite different compositionally. In this case, one wonders whether 'water' really denoted a compositional stuff all along; conceivably, 'water' has undergone a kind of meaning shift from a term for an environmental or functional stuff to term for a compositional stuff. If there has been a meaning shift, that would violate one of the doctrines espoused by scientific essentialists. (See the next section for what is meant by a meaning shift and why it violates scientific essentialist doctrine.) However, right now at least, we are inclined to treat 'water' as a compositional-stuff term. Accordingly, if it is true that all and only samples of water here on Earth are composed of H_2O , we would, using twin-Earth reasoning, seem entitled to conclude that, necessarily, all and only samples of water are samples of H_2O .

Here is another example, one that has particular philosophical importance, namely, the concept of person. Our doppelgänger on twin Earth (the twin Earthlings) certainly are persons. However, they are not *homo sapiens* according to scientific essentialist doctrine, for they lack both the genetics and evolutionary origins of *homo sapiens*. This seems to show that the concept of person is not the concept of a biological species. Rather, it is a moral-intellectual notion. Furthermore, unlike individual members of *homo sapiens*, persons on twin Earth are not composed mostly of water; they are composed mostly of XYZ. Person certainly is not any kind of compositional notion.

For a final example, consider life. All living things on Earth are composed mostly of H_2O plus certain macromolecules composed of various amino acids. This would not lead us to conclude that the beings on twin Earth macroscopically like the living beings on Earth were not alive. Surely, they would be alive. I suspect that, if we survey

an appropriately wide variety of examples, we would discover that the notion of life is a certain kind of Aristotelian proto-psychological notion.

Now an important qualification is in order. In suggesting that life is perhaps a proto-psychological notion or that person is a moral-intellectual notion, I might of course be making mistakes. However, at this point there is no need to be correct or certain about these philosophical matters. After all, ignorance about these matters would, on the view I will be defending, be only one more instance of the paradox of analysis. Knowledge about these matters arises at the end of the analytical endeavor; it need not be present at the beginning.

The above examples suggest that the intuitions that drive scientific essentialism are indexed to various categories. On this view, knowledge of category concepts and the necessary and sufficient conditions for an item's falling under a category concept must be matters that can be investigated antecedently to the researches of empirical scientists. This suggests that our questions concerning category concepts can be explored *a priori* in a wholly traditional way independently of empirical science. This, at least, is the sort of thing that will follow as a corollary from my "transcendental" argument that local scientific essentialism implies the autonomy of philosophy. We now move on to outline the argument.

3. Two Epistemological Problems for Scientific Essentialism: Intuitions and Determinateness

I have said that the argument is broadly "transcendental." The goal is to lay bare the conditions for a satisfactory scientific essentialism. (Hereafter I reserve the term 'scientific essentialism' to refer to local scientific essentialism; when I deal with global scientific essentialism, I will always say so explicitly.) When we think of early modern epistemology, we may isolate two central problems: the origin of ideas and the ground of knowledge of truths. Scientific essentialists inherit a version of each of these problems. They need a way of solving both of them simultaneously without contradiction.

On the surface at least, the two problems are independent: one can be a rationalist on one of them and an empiricist on the other, or an empiricist on one and a rationalist on the other. Locke, for example, gave an empiricist answer to the question of the origin of ideas but a moderate rationalist answer to the question of the ground

of our knowledge of truths. For even though he attacked the doctrine of innate ideas and defended the thesis that all ideas are derived from experience, he held that there was purely rational demonstrative knowledge of some truths; indeed, he made use of what he thought was absolutely a *priori* knowledge in his famous cosmological demonstration of the existence of God.

I have said that scientific essentialists inherit a version of each of these problems, the origin of ideas and the ground of knowledge of truths. The goal of my "transcendental" argument is to show that they cannot solve them both without contradiction unless they adopt a certain epistemological theory that implies the thesis of the autonomy of philosophy.

Consider first the ground of our knowledge of truths. According to the modified Kantian thesis, if a truth can be known to be necessary, that knowledge can be absolutely *priori*, it never must be a *posteriori*. Empirical evidence is never needed in order to know of a given necessary truth that it is necessary. According to scientific essentialism, however, there are various necessary truths, namely, those concerning naturalistic necessities, that can be known to be necessary only if empirical evidence is used. That is, our knowledge that certain naturalistic necessities are necessary is essentially a *posteriori*; it cannot be absolutely *priori*. The problem is this: empirical evidence just on its own cannot justify the conclusion that a given truth is necessary. Something else is needed in order to justify this *modal step*.⁹

In our earlier reconstruction of how scientific essentialists bridge the modal step, we saw that absolutely *priori* knowledge of a certain general philosophical proposition is invoked. (In our example the philosophical proposition was that certain types of natural-kind identities are necessary if true.) But what precisely is the ground of the scientific essentialist's knowledge of this sort of general philosophical proposition? The answer is that ultimately this knowledge is justified by *intuition*. Kripke tells us: "I think it [intuition] is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything."¹⁰ (For our purposes, we may identify intuitions with noninferential beliefs regarding the applicability of a concept to a hypothetical case.¹¹ Or if intuitions are not strictly identical to such beliefs, they are mental states having a strong modal tie with them.)

Here are some well-known examples of intuitions used in scientific essentialist arguments. The Kripkean intuitions that it was possible for Aristotle not to teach Alexander; that it was possible for Gödel (versus Schmidt) not to discover the incompleteness theorem; that it is possible that heat should not produce S ["sensations of heat"]; that if gold is in fact the element with atomic number 79, then in another possible world, if something were macroscopically like gold but did not have atomic number 79, it would not be gold; and so forth. And the Putnamian twin-Earth intuitions that, if all and only samples of water on Earth were composed of H_2O and if there were a twin Earth containing samples of XYZ ($\neq H_2O$) that are macroscopically just like water, then these samples of XYZ would not be water; and so forth.

Now for the problem. If scientific essentialists use intuitions to bridge the modal step, what gives these intuitions any evidential weight? Scientific essentialists must find a satisfactory answer to this question that is consistent with their denial of the modified Kantian thesis.

Some scientific essentialists would give us to believe the modal step can be bridged without relying on intuitions as evidence. They seem to think that the modal step can be bridged with logic alone (namely, Leibniz's law plus modal logic) or with a combination of logic and a theory of language (namely, the direct reference theory of names, the theory that names are rigid designators¹² having no Fregean sense). However, it is easy to show that this account just on its own is deficient on two counts. First, logic plus the theory that names are rigid designators not having Fregean senses does not by itself bridge the modal step at issue. A much richer logico-linguistic theory is needed: specifically, a theory to the effect that, in addition to names, certain *descriptions* are rigid designators. For example, descriptions such as 'the compound whose constituent molecules are constituted of two hydrogen ions and one oxygen ion', 'the element with atomic number 79', 'mean kinetic energy', and so forth. (Such descriptions are quite unlike the nonrigid descriptions 'the number of planets', 'the compositional stuff predominantly filling Lake Tahoe', etc. But why?) Second, the principles of logico-linguistic theory that are needed to justify the modal step are themselves controversial and, therefore, are in need of a justification of their own. What justifies these logico-linguistic principles? The data used to justify these principles are, once again, intuitions (for example, the sort of intuitions

Kripke appeals to in *Naming and Necessity* and the sort of intuitions that Putnam appeals to in "The Meaning of 'Meaning'"). In controversial areas of logic and linguistic theory the use of intuitions as data is crucial. (I argue for this assessment in detail later.)

Because scientific essentialists must ultimately use intuitions to justify the modal step, they are obliged to have (an overall theory that is consistent with) a satisfactory theory of why intuitions should have any evidential weight. This demand is dramatized when we remember that critics of scientific essentialism (e.g., neo-Humeans, Quineans, and skeptics) might hold that modal intuitions are mere beliefs having no evidential weight whatsoever. After all, some beliefs that pop into a person's mind (e.g., a gambler's hunches) have no evidential weight and should simply be disregarded in theoretical reasoning. What distinguishes intuitions from such beliefs? For example, Kripke's intuition that it was possible for Aristotle not to teach Alexander is utterly crucial to his argument for the conclusion that the name 'Aristotle' does not express, as a descriptive sense, the property of being the teacher of Alexander. *Without the intuition Kripke would have no argument.* But why should the mere belief that it was possible for Aristotle not to teach Alexander provide any evidential support for the proposition that it really was possible for Aristotle not to teach Alexander? Why is it legitimate to use that proposition—as opposed to a proposition about a given slot machine that a gambler might happen to fancy—in further theoretical reasoning? Similarly, Putnam's twin-Earth style intuitions are utterly crucial to his argument that, if water = H₂O, then necessarily water = H₂O. *Without these intuitions Putnam would have no argument.* But why should a mere belief—that, if all and only samples of water on Earth were composed of H₂O and if there were a twin Earth containing samples of XYZ (\neq H₂O) that are macroscopically just like water, then these samples of XYZ would not be water—provide any evidential support? Why is it legitimate to use this belief—as opposed to your hunches at the race track—in further theoretical reasoning?

There are versions of rationalism, which we will review below, that provide a coherent explanation of why intuitions (as opposed to hunches) have evidential weight. But scientific essentialism has called rationalism into question; indeed, its denial of the modified Kantian thesis in connection with naturalistic necessities is tantamount to a rejection of all the familiar unrestricted versions of rationalism. Consequently, a new theory of what gives intuitions their

evidential weight is needed if scientific essentialists are going to be able to explain why the modal step in their argument is justified. One such "theory" is the following: the intuitions that support scientific essentialism may legitimately be used as evidence; anti-scientific-essentialist intuitions may not legitimately be used as evidence. But this "theory" would be unacceptably *ad hoc*. If an acceptable theory cannot be found, we must take seriously the skeptical empiricist thesis that intuitions have no evidential weight, that their use in philosophical arguments is merely a hangover from the good-old rationalist days. If this were so, then the arguments given by scientific essentialists would not be justified, and scientific essentialism would have to be abandoned. Therefore, there is a crying demand for a new theory of why intuitions should have any evidential weight.

I will argue that the leading traditional theories—unrestricted rationalism, unrestricted Platonism, empiricism, and coherentism—are inconsistent with scientific essentialism or else they fail to explain the evidential status of intuitions. My conclusion is that the only acceptable theory of the evidential status of intuitions that is consistent with (local) scientific essentialism is a two-tier theory, one that segregates category and content concepts from naturalistic concepts. This theory incorporates a form of rationalist explanation of the evidential status of intuitions concerning the applicability of category and content concepts to hypothetical cases that are specified *exclusively* in terms of category and content concepts. This circumscribed rationalist explanation is consistent with scientific essentialism. At the same time, it can be used to explain why scientific essentialist intuitions about the applicability of naturalistic concepts to hypothetical cases have evidential status. These intuitions are determined by traditional rationalist intuitions regarding relevant category and content concepts together with our knowledge of the categories of the naturalistic items involved. What gives the naturalistic intuitions their evidential status is that they are determined in this way by category and content intuitions, intuitions whose evidential status can be explained independently by the circumscribed rationalist theory without contradicting scientific essentialism. (This explanation will be set forth more fully later.)

Consider an example, say, Putnam's twin-earth intuition about water: If all and only samples of water on Earth were composed of H_2O , then macroscopically similar samples of XYZ ($\neq H_2O$) on twin Earth would not be samples of water. This naturalistic intuition is

just one of infinitely many instances of the following rationalist intuition regarding the category concept of compositional stuff: If paradigm samples of a compositional stuff have a certain complex composition, then items lacking that composition would not qualify as samples of the compositional stuff. The specific Putnamian intuition about water is determined by this sort of categorial intuition together with our knowledge that water is a compositional stuff. However, given the circumscribed rationalist theory, the categorial intuition has evidential weight, and this explains why the specific naturalistic intuition has evidential weight.

(This two-tier theory also explains two other kinds of phenomena. First, it explains why our apparent anti-scientific-essentialist intuitions cannot really be about genuine metaphysical possibilities, but instead must be about mere epistemic possibilities: we do not really have intuitions about such metaphysical possibilities because the corresponding category and content principles that would be needed to determine such intuitions are themselves unintuitive. Second, with much the same techniques, the two-tier theory explains the patterns in our intuitions about such matters as food, fuel, person, life, and so forth, which we discussed in the previous section. These two kinds of phenomena remain unexplained mysteries on all other theories of intuition. These matters are discussed in some detail later.)

The two-tier theory thus provides an explanation—indeed, the only explanation—of why the pro-scientific-essentialist intuitions have evidential status, an explanation that does not contradict the scientific essentialists' denial of the modified Kantian thesis. Here then is one component of our "transcendental" argument. A condition on the justification of the scientific essentialists' modal step is that this two-tier theory of the evidential status of intuitions be true. On this theory, the concepts comprising one of the two tiers—namely, category and content concepts—are susceptible to traditional *a priori* philosophical investigation.

We come now to the second central problem of early modern epistemology, the origin of ideas. The early modern view was that certain ideas are gotten from experience or are possessed innately and, then, by means of what we might call "Lockean operations," a person combines these "given" ideas to obtain all other ideas. There is a continuity of thought from this early modern view to that of Russell (and probably Frege). All these philosophers held that a person has various given concepts—given either in experience or by

reason—and these given concepts are combined somehow (say, by various logical operations) to form complex descriptive concepts. Every concept is either a given concept or a complex descriptive concept formed from the given concepts. On this approach, concepts given in experience are concepts of phenomenal qualities and basic mental relations; they therefore would be cases of content concepts. And the concepts given by reason would be cases of category concepts. As a result, naturalistic concepts—for example, the concepts of water, heat, gold, lemon, arthritis, beech, elm—would have to be complex descriptive concepts formed from category and content concepts. But in this case, a person in command of any of these complex descriptive concepts would be in a position to analyze them absolutely *a priori*, and, in turn, the person should be able to define all the associated naturalistic items—water, heat, gold, and so forth—absolutely *a priori*. But this outcome contradicts (local) scientific essentialism, which repudiates the modified Kantian thesis. (This is the point contact between the repudiation of the modified Kantian thesis and the repudiation of the descriptivist theory of names.)

To avoid this problem, scientific essentialists need to find an alternate theory of what it takes to possess a naturalistic concept. To meet this demand, they turn to the *causal theory*. (I use the term liberally: historical-explanation theories and baptism-cum-naming-chain theories are to be counted as versions of the causal theory.) This theory provides an alternate explanation of the origin of (at least some of) our ideas: to possess a concept (of the right sort) a person need only be properly situated in the world; in particular, the person need only bear appropriate causal (historical, socio-linguistic) relations to items in the world. The strategy of our “transcendental” argument will be this. We will suppose that some version of the causal theory is right for various naturalistic concepts, and we will then try to determine the conditions required in order for it to be right in at least those cases needed for the truth of local scientific essentialism. If, as scientific essentialists, we were to throw out (part of) the traditional theory and to adopt instead a version of the causal theory, what would be required in order for the new view to be satisfactory for our purposes? Our answer will be this. On any causal theory that is satisfactory for our purposes, a person with a naturalistic concept must possess background category and content concepts, concepts whose possession must be explained independently by a traditional theory that insures that these concepts may be investigated absolutely

a priori.

Before we go any further, an important point of clarification is in order. In the present context, when we speak of a theory (explanation, account) of what it is to possess a concept, we mean a theory that specifies the general conditions that, as a matter of logical or metaphysical necessity, are necessary and/or sufficient for possessing the concept. We do not mean a theory that gives a causal explanation of why in the course of one's life one has come to possess a certain concept. Thus, we are seeking "formal causes" not "efficient causes." It is especially important to keep this distinction in mind when we speak of the causal theory or the historical-explanation theory, for there is a temptation to think that such a theory is giving an efficient-causal explanation (perhaps in the deductive-nomological format) of a specific mental occurrence. On the contrary, such a theory aims to provide a general statement of conditions that are, as a logical or metaphysical necessity, necessary and/or sufficient for the possession of a concept. It just so happens that these conditions involve causal (historical, socio-linguistic) features. (The causal theory of perception is a helpful analogy. This theory aims to provide conditions that are necessary and/or sufficient for someone *x* perceiving that *A*; the key condition is that the fact that *A* should cause it to appear to *x* that *A*. This is a causal theory simply because the notion of causation figures into this condition.)

Now if scientific essentialism is right, a primary requirement is that the relevant naturalistic concepts be *determinate*, that our relevant naturalistic concepts be possessed *determinately*. To see what this requirement comes to, consider an example of indeterminateness. Suppose I say, "What is the weight of that ship?" Do I mean with or without the mast, the deck fittings, the sails, the engine, the reserve engine, the radar equipment, the stateroom furnishings, the galley stove? Do I mean down to the ton, pound, ounce, gram, or what? Well, I just do not have an answer these questions. When I said, "What is the weight of that ship?" I did not ask a determinate question. A plurality of questions have equal claim to being what I asked, and there seems to be no fact of the matter as to which one I did ask.¹³ Consider another example. Suppose that you introduce a new term 'multigon' simply by applying it to various closed plane figures having several sides. You have never consciously and explicitly considered the question of whether triangles and rectangles are multigons. When someone asks you this question, you do not know

what to answer. You have no inclination to favor one answer or the other (even though you have all the pertinent empirical information, you are fully attentive, and you are more than intelligent enough to identify elementary geometric figures). Your term 'multigon' might mean closed plane figure with more than four sides, or it might simply mean closed plane figure. (In my dictionary each of these is entered as a meaning of 'polygon'.) Each has equal claim to being the meaning of your newly introduced term 'multigon'. But there seems to be no fact of the matter regarding which one it is. Rather, it seems that you have no determinate concept.

As I have said, if scientific essentialism is right, the relevant naturalistic concepts must be determinate; it will not do for them to be indeterminate. To see why, recall the once-popular logical-positivist theory of "scientific definitions." According to this theory, when scientists offer a "scientific definition" (e.g., that water = H_2O , that heat = mean kinetic energy, that gold = the element with atomic number 79, etc.), they are merely stipulating a new concept or "precisifying" a previously indeterminate concept. Either way, the "scientific definition" produces a mere "meaning shift"; it does not report a genuine empirical discovery. Consider the well-worn example of the term 'fish'. On this logical-positivist view, 'fish' once literally applied to whales, but scientists realized that it would be more useful to have a term that excluded these fish-like mammals. They accordingly redefined 'fish' so that thereafter it expressed a concept with this new, circumscribed extension. Or, on a common variant of this view, 'fish' previously did not express a single determinate concept at all, and the "scientific definition" had the effect of finally singling out, from a plurality of candidate concepts, a particular concept as the determinate meaning of the term. According to scientific essentialists, by contrast, when scientists give a "scientific definition," they are not redefining or precisifying. Right from the start, it was necessary that whales were not fish. Biologists made an empirical discovery about the essence of the natural kind fish; they made an empirical discovery about the essence a single determinate "it" that was under investigation all along.

The reason that scientific essentialism is inconsistent with these "meaning-shift" theories is this. On either the redefinition view or the precisification view, a "scientific definition" is a form of stipulative definition. Accordingly, the person doing the stipulating is in a position to know *a priori* that what is said is true (e.g., that water =

H₂O, that heat = mean kinetic energy, that gold = the element with atomic number 79, etc.).¹⁴ It would be just be a case of old-fashioned analytic *a priori* knowledge. But this contradicts the scientific essentialist doctrine that what is said in such cases must be essentially *a posteriori*. If it were not essentially *a posteriori*, the modified Kantian thesis would be true; in turn, scientific essentialism would be false.

(To get a better picture of what is going on here, consider Kripke's meter-stick example. The term 'meter' was introduced by a person who stipulated that one meter was the length of a certain stick. According to Kripke, at the moment of the stipulation the person knows *a priori* that one meter is the length of that stick, and this is so despite the fact that the proposition that one meter is the length of that stick is contingent. Now if the "meaning-shift" view were correct, then our scientists would have stipulated, not discovered, that water = H₂O, that heat = mean kinetic energy, that gold = the element with atomic number 79, and so forth. Accordingly, these things would be known *a priori* by the scientist making the stipulation. This *a priori* knowledge would differ from the kind in the meter-stick example inasmuch as what is known by the scientist is necessary, not contingent. The upshot would be that all the supposedly *a posteriori* necessities associated with scientific essentialism would turn out to be knowable *a priori*, and thus the modified Kantian thesis would be right after all. To avoid this outcome, scientific essentialists must therefore reject "meaning-shift" theories.)

Summing up, the second part of our "transcendental" argument concerns the need to give a nontraditional, causal explanation of the determinateness of our naturalistic concepts. When scientific essentialists try to work out such a theory in detail, they are, I will try to show, unable to account for determinateness by purely causal (or purely "externalist") mechanisms. The reason is that the formation of a new naturalistic concept (or the formation of a new *de re* thought or the introduction of a new name) runs into a fatal problem of underdetermination (ambiguity) if exclusively causal resources are invoked. When a person appears to form a new naturalistic concept (or to form a *de re* thought of a new naturalistic item or to introduce a name for the item), the act will always be underdetermined (ambiguous) if exclusively causal or "externalist" resources are involved. The thought or reference is determinate only if the person employs background category and content concepts. So just as in

the case of the problem of the evidential status of our intuitions, so in the case of the problem of the determinateness of our concepts, scientific essentialists will be forced to adopt a two-tier theory that separates naturalistic concepts from background category and content concepts.

To explain the determinateness of the concepts occupying the lower tier, namely, the naturalistic concepts, scientific essentialists must supplement the purely causal theory with another theory that emphasizes our category and content concepts. But then the question arises, how are we to explain the determinate possession of these category and content concepts? An attempt to use a causal theory to explain the determinate possession of those concepts would lead to a vicious explanatory regress: the explanation would have to invoke antecedently possessed category and content concepts whose possession would in turn be in need of explanation, and so on *ad infinitum*. A noncircular condition for what it takes to possess a concept determinately could thus never be stated. It follows that some other model, besides the causal model, will (at least at some stage) be needed to explain the determinate possession of (at least some of) our category and content concepts.¹⁵ I will argue that the only satisfactory model is a traditional, noncausal one that meshes exactly with the two-tier theory of the evidential status of intuition that is already needed by scientific essentialists. Otherwise, there would be no satisfactory explanation of the determinate possession of our naturalistic concepts and there would be no explanation of the evidential status of the scientific essentialists' intuitions about naturalistic items. Scientific essentialists would be unable to solve the two fundamental epistemological problems confronting their theory.

The overall conclusion will be this. There must be an autonomous level of category and content concepts, determinate possession of which must be explained in a traditional, noncausal way and theoretical knowledge of which may be obtained absolutely *a priori*. Only then can scientific essentialists have both a satisfactory causal theory of the determinateness of our naturalistic concepts and also a satisfactory explanation of what gives intuitions their evidential weight and, in turn, of what justifies the modal step in the arguments for essentially *a posteriori* necessities. However, (most of) the central questions of philosophy are formulated exclusively in terms of category and content concepts. It will follow, therefore, that insofar

as the answers to these questions can be known at all, they can be known absolutely *a priori*, without the aid of empirical science. Hence, the autonomy of philosophy. In the book on the autonomy of philosophy that is now underway, I plan to show how all this sheds new light both on philosophical method and on a great many specific philosophical problems ranging from metaphysics and epistemology to philosophical psychology and moral philosophy.

In the remainder of the paper I will present in some detail the first component of our "transcendental" argument, which deals with the evidential status of intuitions. Following that, I will sketch the second component, which deals with the determinateness of our concepts. I will close by sketching how to argue "transcendentally" that (most of) the central questions of philosophy are formulated exclusively in terms of category and content concepts. These three "transcendental" arguments yield the thesis of the autonomy of philosophy.

Part II: The Evidential Status of Intuitions

4. *The Dependency of Scientific Essentialism on Intuitions*

"I think it [intuition] is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have about anything."

Saul Kripke, *Naming and Necessity*.

The first thing to show is that the modal step made by scientific essentialists cannot be bridged by uncontroversial principles of logic or linguistic theory. Any principles of logic or linguistic theory that would bridge the modal step are themselves controversial and, hence, in need of justification. After all, many alleged principles of logic and linguistic theory are false. How does one tell which ones are true? Ultimately, by using intuitions as evidence. I will begin by arguing that this assessment holds even for a principle as sacrosanct as Leibniz's law. The point is not to challenge Leibniz's law but rather to lay bare its theoretical status. After that, I will turn to principles that are controversial by anyone's standards.

A principle of logic that is frequently used in scientific essentialist arguments is the principle of the necessity of identicals. That is, for all x and y , if x is identical to y , then necessarily x is identical to

y. This principle follows from Leibniz's law together with the principle of the reflexivity of identity and the necessitation rule from quantified modal logic.¹⁶ However, there are a number of *prima facie* counterexamples to Leibniz's law, such as the ship of Theseus. At time t_1 , the ship is wood₁. Over time, however, wood₁ is slowly replaced by a new quantity of wood, wood₂. At time t_2 , the ship is wood₂. But, manifestly, wood₁ is not wood₂. However, Leibniz's law implies that, if at t_1 the ship is wood₁ and at t_2 the ship is wood₂, then wood₁ is wood₂. Because the consequent is manifestly false, we have a *prima facie* counterexample to Leibniz's law.

One approach is just to conclude that Leibniz's law, in fact, does not hold in its unqualified form. Another approach would save Leibniz's law from candidate counterexamples such as this by multiplying the senses of 'is': not only is there an 'is' of identity, but also there is an 'is' of constitution. Correlatively, one multiplies the objects floating there in the water. Not only is there the ship, but also the wood. These two items have exactly the same size, shape, location, and weight. If the ship weighs ten tons, then when Hercules lifts it, he is actually lifting *two* ten-ton items simultaneously. And there are many more items floating there besides the ship and the wood: for example, the collection (or quantity) of molecules, the collection of atoms, and the collection of subatomic particles. Moreover, besides ship-of-Theseus style identity puzzles, there are other problems for Leibniz's law; for example, identity puzzles involving office-holders. To solve these problems one must multiply the senses of 'is' even further, positing next an 'is' of function or role, and, in turn, one must multiply entities even further.

Are there really so many entities? It is not *obvious* that saving unqualified Leibniz's law yields the best logico-linguistic theory. Following an idea of Paul Grice, George Myro has constructed a very elegant, rigorously formulated time-relativized logic that might yield a simpler overall theory.¹⁷ After all, nearly everyone already holds that *most* statements must be time-relativized; only then can we consistently assert, for example, that the ship is red at t_1 and not red at t_2 . A fully time-relativized logic is simply one in which this is generalized to *all* statements, including identity statements. (Generality, you will recall, is a theoretical virtue.) This generalization opens up the possibility that, just as the ship can be red at t but not red at t_2 , so the ship can be identical to the wood at t_1 but not identical to it at t_2 . In this way, such a logic permits one to give up Leibniz's

law in its unqualified form and, thus, to avoid the ontological and linguistic complications needed to save the law from the *prima facie* counterexamples. (Nevertheless, one can keep Leibniz's law in an unproblematic time-relativized form.) Now just as there is a time-relative logic, there is a corresponding "world"-relative logic.¹⁸ In this logic one finds the prospect of contingent, as opposed to necessary, identities. For example, consider a "possible world" in which no wood was replaced in the ship of Theseus during the interval between t_1 and t_2 . Then in that "possible world" the ship, which was identical to wood₁ at t_1 , would still have been identical to wood₁ at t_2 . So there is a "possible world" in which the ship would be identical to wood₁ at t_2 . However, in the actual "world" the ship is identical to wood₂ at t_2 . It follows that the identity of the ship to wood₂ at t_2 is relative to this "world." Hence, this identity is only contingent. Since in a "world"-relative logic contingent identities are permitted, the principle of the necessity of identicals would not hold. So the relativized theory would go.

Now the point here is not to deny or even to challenge Leibniz's law. The point is to show the need for justifying it over the time-relativized or "world"-relativized theory. But how can that justification be given? To save Leibniz's law from the wealth of candidate counterexamples, one needs to complicate one's ontology and one's logico-linguistic theory. However, there are other theories that also handle those examples. How does one adjudicate between the competing theories—the classical Leibnizian theory and the competing time-relativized or "world"-relativized theory? There is no alternative but to compare them as whole theories. But theories of what? What data or evidence are these theories designed to systematize and to explain? The data are just our intuitions concerning the various relevant examples. To adjudicate between the two theories, one must gather together a whole body of intuitions concerning these examples, and then one must determine which of the competing theories best systematizes and explains them.

So, if the way scientific essentialists justify the modal step is to invoke Leibniz's law—and, in turn, the principle of the necessity of identicals—they must justify these principles. But the only way we have of justifying these principles is to construct a theory that systematizes the body of our intuitions regarding the relevant examples. Thus, *at bottom* the justification of the modal step would rely on intuitions. Consequently, scientific essentialists are required

to have an epistemological theory that explains why such intuitions should have any evidential weight.

Besides the principle of the necessity of identicals, the theory that names are rigid designators is also often used in the attempt to justify the modal step. But what justifies the theory that names are rigid designators? The arguments for this theory are plainly intuitive, for example, the argument against the theory, attributed to Frege, that names are synonymous to definite descriptions. But this argument makes liberal use of intuitions regarding hypothetical cases: the possibility that Aristotle did not teach Alexander, the possibility that Gödel did not discover the incompleteness theorem, the possibility that water is not the liquid filling the rivers and lakes, that heat is not the cause of S [i.e., "sensations-of-heat"], and so forth. But what is it that makes these possibilities evident? The answer is that we find them intuitive. But what is it that gives an intuition any evidential weight? If the rigid-designator theory of names is used to justify the modal step, the scientific essentialist once again needs a theory of what it is that gives the supporting intuitions their evidential weight.

Some people might challenge our conclusion by holding that the Mill-Kripke-Putnam theory that names are not synonymous to definite descriptions is really just an ordinary empirical theory, namely, an empirical theory of language. According to this view, scientific essentialists do not rely on intuitions as evidence, for example, the intuition that it was possible for Aristotle not to teach Alexander. Rather, scientific essentialists merely invoke commonplace empirical knowledge about our native language, for example, the knowledge that the string 'It was possible for Aristotle not to teach Alexander' is a true sentence in our native language (or the knowledge that the string 'Aristotle did not teach Alexander' is a possibly true sentence in our native language).¹⁹ Now it is true that this linguistic knowledge is *partly* empirical; it is based in part on our empirical knowledge of the syntax and semantics of our native language. Specifically, our syntactic knowledge that the string 'It was possible for Aristotle not to teach Alexander' is a sentence in our native language. Plus our semantical knowledge that in our native language 'It was possible for Aristotle not to teach Alexander' is true if and only if it was possible for Aristotle not to teach Alexander.²⁰ (This semantical knowledge derives from the underlying semantical knowledge that in our native language 'possible' means possible, 'Aristotle' means Aristotle, 'not' means not, 'teach' means teach,

'Alexander' means Alexander, and our semantical knowledge that the truth conditions of a whole sentence are determined by certain compositional rules and the meanings of the constituent expressions of the sentence.) But how does empirical linguistic knowledge of this biconditional help us to get to knowledge of the left hand side, that is, to knowledge that 'It was possible for Aristotle not to teach Alexander' is a true sentence in our native language? We first need to have knowledge of the right hand side, namely, the *extralinguistic* knowledge that it was possible for Aristotle not to teach Alexander. So we are right back where we started; the idea of relying on empirical linguistic knowledge was just a run-around. Antecedent extralinguistic knowledge that it was possible for Aristotle not to teach Alexander is required in order to get the empirical linguistic knowledge that 'It was possible for Aristotle not to teach Alexander' is a true sentence in our native language. How do we get this antecedent extralinguistic knowledge? Ultimately, either we or the authorities upon whom we rely must use intuitions as evidence.

Those who disagree with this conclusion have an inflated view of what knowledge of the syntax and semantics of our language amounts to. Consider an analogy. How do we know that the strings 'Stock markets crashed in 1929' and 'The square root of two is irrational' are true sentences in our native language? To be sure, this knowledge is based in part on empirical knowledge of the syntax and semantics of our native language: 'Stock markets crashed in 1929' is a true sentence in our native language if and only if stock markets crashed in 1929; 'The square root of two is irrational' is a true sentence in our native language if and only if the square root of two is irrational. (Knowledge of these biconditionals derives from the underlying semantical knowledge that 'stock' means stock, 'markets' means markets, 'crashed' means crashed, and so forth, plus the semantical knowledge that the truth conditions of a whole sentence are determined by certain compositional rules and the meanings of the constituent expressions of the sentence.) But one's knowledge of these biconditionals is hardly sufficient. One must in addition have knowledge of the right hand sides. That is, one must have the extralinguistic knowledge that stock markets crashed in 1929 and that the square root of two is irrational. In the former case, this extralinguistic knowledge is a *posterori*; in the latter, it is a *priori*. And this *priori* knowledge is arrived at by a proof from axioms of number theory resting ultimately on intuitions. By analogy, our

knowledge that the string 'It was possible for Aristotle not to teach Alexander' is a true sentence in our native language rests partly on our empirical knowledge of the syntax and semantics of our native language. But this empirical linguistic knowledge does not suffice; extralinguistic knowledge—namely, that it was possible for Aristotle not to teach Alexander—is also required. And this extralinguistic knowledge is based on intuition.

A third point where intuitions are used in justifying the modal step is this. Even if one takes the principle of necessity of identicals and the rigid-designator theory of names to be evident, those two theories do not by themselves justify the modal step needed in order to reach *scientifically interesting* necessities concerning natural kinds. For example, consider the following argument: water is identical to H_2O ; therefore, necessarily water is identical to H_2O . Notice that ' H_2O ' is not a name at all, but rather a description.²¹ (Or at least it is a disguised description, short for something like 'the compound molecules of which consist of two hydrogen ions and one oxygen ion'. If you doubt this, then throughout the present paragraph just replace the term ' H_2O ' with the description I just mentioned.) Because ' H_2O ' is a *description* rather than a name, Leibniz's law plus the theory that *names* are rigid designators do not license the above argument. To dramatize this, notice that this modal argument has the same form as the following: water is identical to the compound predominantly filling Lake Tahoe; therefore, necessarily, water is identical to the compound predominantly filling Lake Tahoe. If Leibniz's law and the rigid-designator theory of names licensed the earlier argument, they would also license the latter argument. But the latter argument is plainly invalid. So it will not do to hold that Leibniz's law and the rigid-designator theory of names licenses the earlier argument. Some further logico-linguistic principle is needed, one that distinguishes between descriptions like ' H_2O ' and descriptions like 'the compound predominantly filling Lake Tahoe'.

The missing principle evidently is that the former are *rigid descriptions* whereas the latter are not. (A rigid description is one that designates the same item in every "possible world" in which the item exists.) However, the rigid-designator theory of names tells us nothing about this. Descriptions are not names. So how do we know that descriptions like ' H_2O ' are rigid and descriptions like 'the compound predominantly filling Lake Tahoe' are not? Evidently, scientific essentialists need a theory according to which the former descriptions iden-

tify the *essence* of items whereas the latter do not. However, the formulation and justification of this theory will require a traditional investigation in metaphysics. This investigation will make liberal use of intuitions as evidence for the theoretical conclusion that 'H₂O' and kindred expressions are rigid descriptions because they identify the essences of the items they denote. So, once again, the justification of the modal step is at bottom dependent on the use of intuitions as evidence.

There is a way of trying (at least for a while) to avoid the conclusion that intuitions are needed to show that the property of being H₂O is essential to water whereas the property of being the compound predominantly filling Lake Tahoe is not. Namely, one could adopt the following extremely strong philosophical thesis about *pure* scientific theories: if T is a pure scientific theory (as opposed to an *applied* scientific theory) and if T is true, then T is necessary. The true proposition that water = H₂O is a pure scientific theory; however, the true proposition that water = the compound predominantly filling Lake Tahoe is only an applied scientific theory. Consequently, from the strong philosophical thesis, one can infer that the former proposition is necessary if true, but one cannot infer this in the case of the latter proposition. In this way, the strong philosophical thesis suffices to bridge the modal step in the scientific essentialists' argument. However, invoking this thesis hardly allows scientific essentialists to avoid relying on intuitions as evidence in the long run. After all, this thesis is really just a very strong version of scientific essentialism. Indeed, in our earlier reconstruction of the scientific essentialists' argument, this strong thesis could just be substituted for the weaker modal thesis that certain types of natural-kind identities are necessary if true. So the real question facing advocates of the strong thesis is how to justify it. But there is evidently no way to do so except by consulting intuitions about individual hypothetical cases, including in particular the standard sort of hypothetical cases invoked by scientific essentialists.

In addition to the problem of justifying the strong thesis, there is also the problem of determining in individual cases whether an hypothesis is *purely scientific* in the sense invoked in this thesis. Any descriptive identity statement that is purely scientific in this sense will in effect single out an essence. This fact would place an extraordinary burden on every pure scientist, a burden that might force scientists, as a matter of daily practice, to go well beyond empirical

investigations and to engage in *a priori* philosophical investigations of essence. Accordingly, the strong thesis might radically alter empiricist conceptions of science: science might be dependent upon philosophy (although philosophy would not be dependent upon science).

The inescapable conclusion is that scientific essentialists must ultimately use intuitions as evidence to justify the modal step in their arguments. But why is one ever entitled to use intuitions to justify a theoretical conclusion? After all, a person could have all sorts of crazy noninferential beliefs that could not legitimately be used to justify theoretical conclusions. What makes our intuitions about hypothetical cases different? Perhaps, just as with hunches, these intuitions carry no evidential weight. (This, of course, is what radical Quineans and skeptics hold. Why should we not be Quineans or skeptics?) If this were so, scientific essentialism could not be justified. So plainly scientific essentialists need an epistemological theory that explains why intuitions have evidential status.

My strategy will be to examine critically four promising traditional theories of intuition: rationalism, empiricism, coherentism, and Platonism. (This survey must for reasons of space be briefer than one would wish. I hope that the main lines of argument can be made clear.) My conclusion is that the latter three theories fail unless they presuppose rationalism. But that will leave us with a problem, namely, that full rationalism is inconsistent with scientific essentialism. There will be only one satisfactory way out. Namely, the two-tier theory that separates category and content concepts from naturalistic concepts. The theory of intuition that will serve scientific essentialists will be a circumscribed rationalist theory that holds for category and content concepts and a special form of derived theory that holds for naturalistic concepts.

But now for the survey of the promising traditional theories of why intuitions have evidential weight.

5. Rationalism

We begin with an example. Suppose that a sincere, normal, attentive person introduces the new term 'multigon' by applying it to various closed plane figures having several sides. Suppose that we know that the concept he expresses with this new term is either the concept of being a closed plane figure or the more restrictive con-

cept of being a closed plane figure with more than four sides. (As I mentioned earlier, these two concepts are listed in the dictionary as alternate meanings of the common noun 'polygon'.) But we do not know which of these two concepts is his concept of multigon. What determines the answer? What is required for his concept of multigon to be that of a closed plane figure rather than that of a closed plane figure with more than four sides? Intuitively, the matter would be settled by how the person applies his concept to decisive hypothetical cases. Suppose that, upon attentively considering the case in normal cognitive circumstances, the person confidently makes the noninferential judgment that it is possible for a three-sided closed plane figure to be a multigon. Then, intuitively, his concept of multigon is that of a closed plane figure rather than that of a closed plane figure having more than four sides. If this normal, attentive person confidently judges that a three-sided figure could be a multigon, how could his concept be that of a closed plane figure with more than four sides? This makes no sense. Or suppose that, upon attentively considering the case in normal cognitive circumstances, the person confidently makes the noninferential judgment that, if a figure were three-sided or four-sided, it would not be a multigon. Then, intuitively, the person's concept of multigon is that of a closed plane figure with more than four sides. This would settle our question. For if in normal cognitive circumstances this normal, attentive person is confident about this extremely simple question, how could he be mistaken (assuming that he really possesses the concept determinately)? Intuitively, he cannot. If this assessment is right, the answer to our question concerning which of the two concepts is the person's concept resides in the person's capacity in normal cognitive circumstances to judge correctly the decisive hypothetical cases. The person possesses the concept determinately only if he has this cognitive capacity.

This sort of insight (perhaps qualified in one way or another) provides the key to the rationalist theory of the evidential status of intuitions. According to rationalists, if a noninferential judgment is of a type that, by its nature, cannot go wrong, then it may legitimately be used in the justification of other judgments. If, by its nature, a type of judgment has this kind of strong modal tie to the truth, then judgments of that type count as evidence (or data). On the rationalist theory, however, a person's noninferential judgments about the applicability of concepts to elementary hypothetical cases do indeed

have this kind of strong modal tie to the truth. Indeed, a necessary condition of having the concept is that the person's noninferential judgments about the applicability of the concept to elementary hypothetical cases must be true (at least when relevant auxiliary conditions are met). But an intuition is (either identical to or at least intimately bound up with) a noninferential judgment about the applicability of a concept to a hypothetical case. It follows, therefore, that intuitions count as evidence.

This in broad outline is what I mean by a rationalist theory of the evidential status of intuitions. Although it plainly needs to be qualified in various ways, this type of theory does have considerable intuitive appeal. Naturally, there are several distinguishable versions of the theory, some of which are reminiscent of one important historical figure or another. For illustrative purposes I will now lay out in a bit more detail the *moderate rationalist* version. The larger line of our argument would not be lost by skipping ahead to the next section at this point.

As I have indicated, the rationalist theory of the evidential status of intuitions is a kind of *truth-based* theory. According to a truth-based theory of evidence, certain types of beliefs by their nature have a strong *modal* tie to the truth (i.e., *necessarily*, they are always true, mostly true, probably true, or normally true). The idea is that this strong modal tie to the truth makes beliefs of this type ones that a person could not fail to be justified in using in subsequent reasoning. Accordingly, such beliefs would count as evidence. On a *moderate* truth-based theory, a proposition is evidence (data) for a person if the person believes the proposition noninferentially and, necessarily, *most* of the person's noninferential beliefs in propositions of that type are true.

We will call the following the *moderate rationalist thesis*: necessarily, *most* of a person's noninferential beliefs about the applicability of a concept to elementary hypothetical cases are true.²² From this thesis and the moderate truth-based theory of evidence, it follows that a person's noninferential beliefs about the applicability of a concept to elementary hypothetical cases count as evidence for the person. But intuitions (at least the intuitions used in justifying scientific essentialism) are noninferential beliefs about the applicability of concepts to hypothetical cases. (Or if intuitions are not strictly identical to such noninferential beliefs, the contents of one's intuitions nevertheless must by and large coincide with the contents of such

noninferential beliefs.) It follows, therefore, that a person's intuitions about the applicability of a concept to elementary hypothetical cases count as evidence for the person. This, then, is a moderate rationalist explanation of why these intuitions have evidential weight.

This explanation of why intuitions have evidential weight depends on the moderate rationalist thesis (the thesis that, necessarily, most of a person's noninferential beliefs about the applicability of a concept to elementary hypothetical cases are true). This thesis is related to a version of the *principle of charity*, the principle that in interpreting the speech and action of others we must assume (at least at the outset) that most of their beliefs are true. However, the rationalist thesis is far more cautious. It does not require that most of a person's beliefs—including empirical beliefs—be true. Rather, it requires only that most of a person's beliefs about the applicability of his concepts to hypothetical cases be true. The reason for this caution is that it seems in principle possible for a person not to have mostly true empirical beliefs. (For example, a person could have a vast array of superstitious and quack-scientific beliefs. Or a person could be subjected to systematic trickery—perceptual illusions, drugs, hypnosis, brainwashing. And many philosophers hold that a person could suffer from one of the traditional skeptical possibilities—the-brain-in-the-vat, systematic hallucination, or dream.) Moderate rationalists avoid this problem by requiring only that most of a person's beliefs about the applicability of a concept to elementary hypothetical cases must be true. Beliefs of this special type are far safer, for they hold no matter what contingent situation the person or the world is actually in. Indeed, these beliefs are as weak as possible: they concern only whether a concept applies to a situation whose auxiliary features are built right into the case. Therefore, there is no dependency on the person's auxiliary empirical information. These are the beliefs that any normal person would have if the person truly has the concept. If a person seems to us not to make mostly true judgments about the applicability of the concept to elementary hypothetical cases, that would indicate, not that the person is mistaken, but rather that the person does not really have the concept to begin with. Or so moderate rationalists would hold.

This brings us to the moderate rationalist theory of what it is to possess a concept determinately, a theory that yields the above moderate rationalist thesis as an immediate consequence. Having a determinate concept is not like having a stamp on your forehead

or having (an inscription of) a predicate in the “belief box” in your brain. The rationalist view of determinate concept possession is broadly Kantian: it is not possible for there to be a disconnected “piece of a mind” (e.g., a disconnected possession of a concept without an associated capacity for applying the concept correctly to elementary hypothetical cases). A disconnected piece of a mind is no more possible than a disconnected piece of space; like space, a mind comes as an integrated, synthesized whole. According to moderate rationalists, having a determinate concept requires having a mental capacity—broadly akin a vivid imagination or perfect pitch—except that it is a cognitive capacity. Specifically, having a determinate concept requires having a cognitive capacity for necessarily making mostly true judgments regarding the applicability of the concept to elementary hypothetical cases that the person might consider.

Let us be more precise. Suppose that a person has some concept *k* but we do not know which concept it is. Let *c* be some familiar concept. (For example, *k* might be the concept of being a multigon, which we discussed at the outset of this section, and *c* might be the concept of being a closed plane figure or the concept of being a closed plane figure with more than four sides.) What is required for the person’s concept *k* is to be this familiar concept *c*? The moderate rationalist capacity theory provides a necessary condition. Let the proposition that ...*k*... be an elementary hypothetical case. Suppose that the person were carefully to sort propositions of this form into two classes, those he believes to be true and those he believes not to be true. Then the person’s concept *k* would be *c* only if most of the propositions that ...*c*... associated with the first class are really true and most of the propositions that ...*c*... are really not true (where the proposition that ...*c*... is just like the proposition that ...*k*... except that *c* takes *k*’s place).²³

Summing up, the moderate rationalist capacity theory of what is required for a person to possess a concept determinately has the moderate rationalist thesis as an immediate consequence. But, as we saw earlier, this thesis yields a moderate rationalist explanation of why intuitions have evidential status. Thus, if correct, moderate rationalism would be able to solve our first problem, namely, the evidential status of intuitions, just by giving a (partial) solution to our second problem, namely, the problem of determinate concept possession. It must be acknowledged that, despite its eventual need for significant qualifications, moderate rationalism is economical and at

least initially plausible.

Now I believe that all rationalists would be wise to modify moderate rationalism in various ways. A preferable position would be either *dialectical rationalism* or *holistic rationalism*. For present purposes there is no need to spell out these positions. One advantage of them is that they put rationalists in a position to state a condition that, on their view at least, would be not only necessary but sufficient for the determinate possession of a concept. On this approach a person possesses a concept determinately if and only if the person possesses certain intellectual capacities: specifically, the capacity in *cognitively ideal circumstances* to apply the concept correctly to most elementary hypothetical cases and to correctly subject the concept to most elementary logical manipulations. Cognitively ideal circumstances are those achieved at the end (versus the beginning) of philosophical dialectic or at the end of the theoretical systematization of one's intuitions, where throughout the process of dialectic or theoretical systematization there is sufficient distinctness, clarity (attentiveness), intelligence, memory (and perhaps desire). By elementary logical manipulations, we mean to include decompositions under the inverses of fundamental logical operations like conjunction, negation, existential generalization, singular predication, and so forth.²⁴ (These forms of rationalism are examined in my projected book on the autonomy of philosophy.)

What is it for a person to possess determinately one concept rather than another? A desideratum for any acceptable epistemology is that there should be some objectively describable intellectual feature of the person that marks this distinction. That is, ideally, possessing one concept rather than another should consist in some objectively describable intellectual feature of the person. Rationalism strives to provide such an intellectual feature, namely, the person's capacity in cognitively ideal circumstances to get at modal and logical truths that hold of one but not the other concept. On this view, a concept is not a free-floating thing that one can just happen upon independently of one's correlative cognitive capacities. The determinate possession of a concept is not an "atomic" phenomenon: a person cannot have a disconnected piece of a mind; a mind comes as whole cloth. Or so rationalists would have us believe.

Let us now return to our larger line of argument and the problem of the evidential status of intuitions. According to rationalists, a person cannot possess a concept determinately unless the person has

an associated capacity to make judgments about the applicability of the concept to hypothetical cases, which judgments have a strong modal tie to the truth. For this reason, the judgments that arise from the exercise of this capacity are the sort of judgment that a person could not fail to be justified in using in subsequent theoretical inferences. Therefore, such beliefs would qualify as evidence (data) for the person. Thus the rationalists' theory of determinateness implies their theory of the evidential status of intuitions.

In the next sections, we will survey two alternate theories of the evidential status of intuitions, empiricism and coherentism. Our conclusion will be that either they are unsatisfactory or they reduce to a rationalist theory. Following that, we will show why rationalism cannot be satisfactory as it stands; it requires significant modification if it is to be of any help in the effort to save scientific essentialism. We will turn finally to Platonism; our conclusion will be that the only version of Platonism that is acceptable is a special case of this modified rationalism.

6. Empiricism

Empiricists are those who wish to count experiences, and only experiences, as evidence. Elementary propositions about one's immediate experiences are the only sort of propositions that may just on their own be used to justify other propositions. Virtually everyone would agree that simple versions of empiricism will not provide an evidential basis for reaching rich modal conclusions of the sort scientific essentialists espouse. There is, however, a sophisticated version of empiricism that is more promising.

Scientific essentialists who wish to maintain sophisticated empiricism might try to adapt a strategy used by Quine and others. This strategy is grounded on the empiricists' thesis that elementary propositions about one's experiences constitute one's *basic data*. The idea is to show that perhaps there are other propositions that can acquire the status of *derived data*. One could reach this conclusion by showing that there is a positive correlation between the truth of these derived data propositions and a person's disposition to believe those propositions.

A leading version of this strategy relies on evolutionary theory. One hopes to find an evolutionary basis for concluding that there is a positive correlation between the truth of a certain type of pro-

position and a disposition to believe propositions of that type. So, for example, if we take sense experiences (construed as inner states) to be our basic data, we might by this strategy be able to show why observational beliefs (that is, beliefs about immediately present external objects) should acquire the status of derived data. The argument is that there are good evolutionary grounds for its being probable that a person will believe elementary propositions about immediately present external object if and only if those propositions are true. For example, we evolved in such a way that it is probable that a person would believe a tiger is running directly at him from the front if and only if there is indeed a tiger running directly at him from the front. The evolutionary advantage of this correlation is plain. A similar account can be attempted for testimony. Our beliefs in the content of what other people tell us do not count as basic data for empiricists. After all, the basic data in the case of testimony are simply certain sense experiences (e.g., the hearing of sounds and the seeing of colors, shapes, and motions). Nevertheless, testimony of others might acquire the status of derived data if one could establish a correlation between our beliefs in the content of the testimony of others and the truth of that content. If an empiricist can establish such a correlation empirically, perhaps by invoking evolutionary theory (e.g., an appropriate doctrine in sociobiology) or a game-theoretic explanation of cooperative activities (specifically, the cooperative activity of linguistic communication), then the empiricist might be able to conclude that testimony indeed has the status of derived data.

Quine has attempted the same general strategy with regard to our elementary logical beliefs. He does not accord to such beliefs the status of basic data. However, he produces an evolutionary argument for the thesis that there is a general correlation between our elementary logical beliefs and the truth of such beliefs. After all, having true elementary logical beliefs enhances our means/ends reasoning and, in turn, our capacity to choose actions that enhance our survival. If this were right, one would be entitled to conclude that, although elementary logical beliefs are not basic data, nevertheless they acquire the status of derived data and, hence, may be used in justifying other propositions.

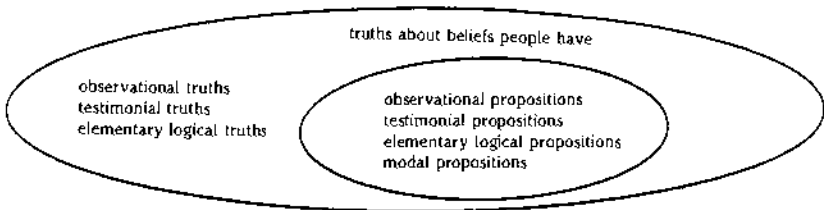
In summary, on the sophisticated empiricist strategy, one first constructs one's best empirical scientific theory, and then one uses the theory to see whether or not there is a correlation between a certain category of beliefs and the truth of those beliefs. If one's best

empirical scientific theory does establish such a correlation, then beliefs in that category are accorded the status of derived data.

The question facing scientific essentialists who wish to be empiricists is this. Can this sophisticated empiricist strategy be used to accord modal intuitions the status of derived data (at least in the case of the modal intuitions used by scientific essentialists to justify the modal step in their arguments)? The answer is that it cannot. (Indeed, implicit in Quine's general attack on modality is the thesis that modal intuitions cannot be accorded evidential status by this or any other empiricist strategy.)

To see what the problem is, let us try to follow the sophisticated empiricist strategy. We are told to consult our best comprehensive empirical theory for the purpose of comparing our theories about the beliefs people are liable to have and our corresponding theories about the way the world actually is. We might represent our best comprehensive empirical theory with the following diagram:

COMPREHENSIVE EMPIRICAL THEORY



The circle within the circle represents our theories about the beliefs people are liable to have. The entries outside that circle represent our corresponding theories about the way the world actually is. Because the comprehensive empirical theory establishes a significant positive correlation between observational beliefs and observational truths, between beliefs in others' testimony and the truth of others' testimony, and between elementary logical beliefs and elementary logical truths, each of these three types of beliefs may be accorded the status of derived data. However, when sophisticated empiricists seek an analogous correlation between the modal beliefs people are liable to have and modal truths affirmed by our best comprehensive empirical theory, they hold that they will find none. Why?

Because they hold that our best comprehensive empirical theory can do its job—namely, predicting and explaining experiences—perfectly well without taking a stand on any modals. That is, they hold that our best empirical theory can do its job perfectly well while just eliminating all modal statements from the theory. (Echoing Hume, Quine holds that this is exactly what we must do if we are really consistent empiricists.) They hold that the resulting comprehensive empirical theory is *simpler* than its competitors and is just as good in all other respects honored by empiricists (namely, in the prediction and explanation of experiences). If this is right, the comprehensive empirical theory that just eliminates modal propositions is better than all the competing theories that take stands on modal propositions. The upshot would be that our best comprehensive empirical theory does not establish any positive correlation between the modal beliefs a person is liable to have and the truth of those beliefs. Unlike observational beliefs, beliefs in the testimony of others, and elementary logical beliefs, one's modal beliefs could not be accorded derived evidential status by the sophisticated empiricist strategy. Consequently, empiricism would be of no help to the scientific essentialist in meeting the demand to explain the evidential status of the intuitions that are used to justify scientific essentialism.

Some sophisticated empiricists might try to avoid this conclusion by including among a person's experiences the "inner experience" of *having a modal intuition*. (Having a modal intuition would thus be counted as a Lockean experience of reflection or a Russellian experience of introspection). What explains the occurrence of these "inner experiences"? Most contemporary empiricists would attempt to explain why people have their modal intuitions by means of a naturalistic cybernetic theory according to which our disposition to have these intuitions is just "hard-wired" or is just a causal byproduct of an efficient brain organization. However, on this kind of empiricist theory, there is no *independent* reason to hold that intuitions that are just "hard-wired" or that are just byproducts of efficient brain organization are tied to the truth. Indeed, even if "hard-wired" modal intuitions once upon a time enhanced the biological fitness of human beings, this would not be evidence that these intuitions had any positive correlation to the truth. (Consider an analogy: even if it were to turn out that there is a "hard-wired" disposition to have racist or sexist beliefs and that once upon a time this disposition somehow enhanced human survival, this would not be evidence that these

beliefs are true.) So if this naturalistic causal explanation is the simplest explanation of why people have their modal intuitions, then modal intuitions cannot be accorded the status of derived evidence on the sophisticated empiricist strategy.

Suppose, however, that the naturalistic, causal explanation is not simplest. Suppose instead that the simplest explanation is based on a rationalist capacity theory of determinate concept possession, for example, the moderate rationalist theory (described earlier). On this theory, possessing a concept requires having a cognitive capacity, namely, the capacity to make mostly true judgments about the applicability of the concept to elementary hypothetical cases. So if a person has a given concept and if the person considers the question of whether the concept applies to a given elementary hypothetical case, the person's capacity to judge the question will be activated and the person will (in most instances) judge that it does (or does not) apply just in case it really does (does not) apply. This style of explanation establishes a positive correlation between the elementary modal intuitions a person is liable to have and the truth of those intuitions. Therefore, if this were the simplest explanation of why people have their elementary modal intuitions, then these intuitions would be accorded derived evidential status on the sophisticated empiricist strategy.

(Platonists also have an explanation of why people have their modal intuitions. Intuitions are a kind of "direct perception" of abstract truths: when intelligent people consider the question of whether a concept applies to an elementary hypothetical case, in most instances they just "see directly" that it applies if indeed it does apply and they just "see directly" that it does not apply if indeed it does not apply. Thus, like the rationalist capacity theory, this "direct perception" theory establishes a positive correlation between elementary modal intuitions and the truth. So if this theory provided the simplest explanation of why our modal intuitions occur, then they would be accorded derived evidential status on the sophisticated empiricist strategy. In section 9 we show that, insofar as this theory is defensible, it may be viewed as a special variant of the circumscribed rationalist theory we will advance. To simplify the discussion, therefore, we will defer separate consideration of the Platonist theory until later.)

Notice, however, that the rationalist capacity theory implies the thesis of the autonomy of philosophy. (I will elaborate on this later.) So if this theory provides the only way in which empiricists can ex-

plain the evidential status of intuitions, empiricists may not adopt global scientific essentialism. Indeed, the situation is probably much worse for empiricists. If the rationalist capacity theory of determinate concept possession were true, there would be a strong modal tie between these intuitions and the truth. Now suppose that a truth-based approach to the theory of evidence is correct. (There is great *intuitive* support for this approach, and it is a component of an attractive explanation of a basic tenet of empiricism, namely, that all elementary propositions about one's current experiences are basic data. See notes 22 and 25.) Then, on the rationalist capacity theory, intuitions would have the status of *basic data* not derived data, as sophisticated empiricists would have us believe. Thus, sophisticated empiricists would no longer be able to hold that only elementary propositions about one's current experiences are basic data propositions. A vast array of nonexperiential propositions must be admitted as basic data, propositions about the applicability of concepts to all manner of hypothetical cases. Thus, a basic tenet of empiricism—that only elementary propositions about one's current experiences are basic data propositions—would be contradicted.

Summing up, if the naturalistic causal explanation of why we have our elementary modal intuitions is the simplest explanation, then empiricism cannot explain the evidential status of our intuitions. On the other hand, if the rationalist capacity theory is the simplest explanation of why we have our intuitions, then empiricism does not provide scientific essentialists with a way to avoid the larger claims we are trying to defend, namely, that the thesis of the autonomy of philosophy is true and global scientific essentialism is false. Moreover, given a truth-based approach to the theory of evidence, this way of trying to explain the evidential status of intuitions actually contradicts empiricism: elementary modal intuitions would be among our basic data, as *rationalists claimed all along*. Thus, empiricism promises to be no help in meeting the demand to explain the evidential status of the intuitions that are used in justifying local scientific essentialism.

7. Coherentism

According to coherentism, there are no privileged beliefs that, independently of their belonging to a person's *theoretical system* of beliefs, have the status of being evident (justified) to the person.

Only when systematized theoretically do a person's beliefs have the status of being evident to a person. A belief that belongs to a person's theoretical system is evident to the person; one that does not is not.

But let us look more carefully at what theoretical systematization must be. According to one account, theoretical systematization consists simply in finding a maximally consistent body of one's beliefs (i.e., a consistent body of one's beliefs that would be rendered inconsistent by adding any of one's other beliefs to it). This account is plainly inadequate. For a person could have such a body of consistent beliefs but nevertheless quite unjustified beliefs. For example, consider a gambler with a host of beliefs about what numbers will come up on roulette wheels, on dice, in lotteries, at the races, and so forth. At the time he has them, these beliefs do not, we may suppose, contradict his other beliefs. (Or if they do, suppose that he just disregards the beliefs contradicted by his hunches and that the beliefs he regards, including his hunches, do form a maximally consistent body. Surely this is possible.) But plainly the gambler's beliefs are not justified.

For a more extreme example, suppose some perverse scientists have made an effort to produce a wide variety of false beliefs in some poor soul. They subject him to drugs, behavioral modification, and hypnosis. The effect of all this is that the person has a host of detailed beliefs about the exploits of presumed creatures in very distant places (e.g., Mars, Alpha Centauri, etc.). He also has detailed beliefs about other "worlds" that he believes to be actually existing although not spatially connected to our "world" (i.e., not connected to the three dimensional space in which we are located). Finally, he has beliefs about a vast array of undecidable mathematical propositions, and he has beliefs about new domains of mathematical objects. (For example, he believes that there are abstract objects, bizarrons, that have exactly the structure of the natural numbers but are not natural numbers, nor are they set-theoretical or property-theoretic constructs. They are different from all other sorts of abstract objects we know of, and yet they are otherwise exactly like natural numbers.) Now the perverse scientists do all of this very carefully so that the resulting beliefs are consistent with (most of) the person's other beliefs. Consequently, these far-out beliefs belong to a maximally consistent body of the person's beliefs. But clearly the person is not justified in believing these things.

A rational agent must do more than merely have a maximally consistent body of beliefs. In addition to having (most of) his beliefs be consistent, the person must assemble his beliefs into a *theory*, that is, a system of propositions that includes general, law-like propositions that serve to *explain* (at least schematically if not in full detail) the various specific propositions in the system. But let us suppose that, knowing of this requirement, the perverse scientists have already supplied the person with a body of far-out law-like beliefs that do just that. He has far-out law-like beliefs from which he could, if he tried, deduce his specific beliefs about what is going on in distant places and in other "worlds" and about, say, the continuum hypothesis and bizarrons. He just finds himself believing all these things, both general and specific; but he has no idea why. Would the beliefs induced by the perverse scientists be justified? Hardly.

What is missing is that the person must have a theory that explains (or at least provides a schematic explanation of) why he himself has come to believe the various propositions that he believes. But suppose that the perverse scientists are aware of this requirement, too, and that they have accordingly supplied the person with far-out theoretical beliefs that do just this. For example, suppose that the person has been hypnotized by the scientists to believe that his various beliefs about distant places and other "worlds" are produced by oracular contact with these places and that his various beliefs about undecidable parts of mathematics and bizarre new types of abstract objects result from the Muse of Mathematics whispering associated mathematical truths to him during his sleep.

Plainly the person would still not be justified. A person cannot justify the propositions he believes just by supplying this kind of far-out explanation of why he has come to believe these propositions. The problem is that not any old explanation will do. To be justified, the person must, minimally, have the *simplest* explanation. (Or, at least, the person must have attempted to find the simplest explanation.) But surely there is a simpler explanation available to the fellow than the one he has offered.

Now let us suppose that the person does seek the simplest explanation. And let us suppose that the resulting explanation identifies exactly the source of all his far-out beliefs. For example, it states that perverse scientists have subjected him to drugs, behavioral modification, and hypnosis that caused him to believe all these far-out things. But let us suppose that, after having arrived at this simplest explana-

tion, the person persists in believing all the far-out things he previously believed—specific and general propositions about distant places, other “worlds,” undecidable parts of mathematics, bizarre new types of abstract objects, and magical explanations of why he believes these propositions. This would clearly be unsatisfactory. After one has arrived at the simplest explanation of the origin of one’s beliefs, one is obliged by reason to abandon the beliefs whose *truth* is not affirmed by that very explanation. (Call this the *truth-affirming-explanations requirement*.) Consider perceptual beliefs, for example. The simplest explanation of why I presently believe that there is a table here might be something like this: there is a table here; the table’s being here causes it to appear to me that there is a table here, and this appearance produces my belief that there is a table here. This explanation affirms the truth of my belief inasmuch as the truth that the table is here is part of the explanation of why I have this belief.

Only when coherentism is supplemented with the truth-affirming-explanations requirement does it stand a chance of being a satisfactory epistemology. With this conclusion in mind, let us return to the question of how the coherence theory might help account for the use of the modal intuitions that are invoked to support scientific essentialism. We have discovered that, if the coherence theory is to be of help, our simplest explanation of why we have these modal intuitions must be one that affirms the truth of these intuitions.

Broadly speaking, two sorts of explanation are available—rationalist capacity theories and causal theories. According to rationalist capacity theories, a necessary condition for possessing a concept is that a person’s intuitions about the applicability of the concept to hypothetical cases must have a strong tie to the truth. So explanations based on such theories serve to affirm the truth of (at least most of) the modal intuitions used in support of scientific essentialism. So the truth-affirming-explanations requirement is met if the rationalist capacity theory is the simplest explanation of why a person has his or her modal intuitions. (Explanations based on Platonist “direct perception” theories are also like this. According to Platonist theories, intuitions are a kind of “direct perception” of abstract truths: when normal, intelligent people consider the question of whether a concept applies to an elementary hypothetical case, in most instances they just “see directly” that it applies if and only if it truly does apply. Of course, misfires can occur, but they are the result of some defi-

ciency in the cognitive circumstances. As the cognitive circumstances approach the ideal in relevant respects, the occurrence of an intuition is nothing but the unmediated "seeing" of an abstract truth. So the truth-affirming-explanations requirement would be met if the Platonist "direct perception" theory were simplest. However, I believe that, insofar as this theory is defensible, it may be viewed as a special variant of the circumscribed rationalist theory we will soon advance. To simplify things, let us defer consideration of this Platonist theory until section 9, where it will be examined in detail.)

Besides rationalist capacity theories, there are various causal theories. The leading causal theories are naturalistic cybernetic theories to the effect that modal intuitions are just "hard-wired" or that they are just a byproduct of economical brain organization. The problem is that these naturalistic explanations do not affirm the truth of one's modal intuitions. These explanations do not provide any reason to think that modal intuitions that are just "hard-wired" or that are just byproducts of efficient brain organization are in any way tied to the truth. Even if in our evolutionary history a certain "hard-wiring" or efficient brain organization happened to enhance biological fitness, that would not by itself indicate that associated beliefs are true. (Recall the examples of racism and sexism mentioned in the previous section.) So if a naturalistic causal explanation is truly satisfactory and if it is the simplest explanation of why we have our modal intuitions, the truth-affirming-explanations requirement would not be met. On this alternative, therefore, one would be obliged to abandon one's modal intuitions. The upshot is that on this alternative scientific essentialists would be unable to justify their theory.

Suppose instead that the rationalist capacity-theory explanation of the origin of our modal intuitions is simpler than the various causal explanations. Since this explanation affirms the truth of (most of) these modal intuitions, scientific essentialists would have succeeded in finding an account of the evidential status of the intuitions they use in support of their position. Notice, however, that this account differs little from the original rationalist theory of *a priori* evidence. Indeed, this account may be viewed as a kind of "transcendental deduction" of the rationalist theory of *a priori* evidence from the coherence theory of justification. If, because of their necessary tie to the truth, (most) beliefs of a certain type are guaranteed necessarily to survive in a person's best theoretical systematization, we can know in advance that beliefs of this type will always be justified regardless

of the other properties that the person's best theoretical systematization might turn out to have. Beliefs that are like this are privileged and, accordingly, may be accorded the status of *data* independently of other considerations. So if the rationalist capacity theory indeed provides the simplest explanation, a person's intuitions about the applicability of a concept to elementary hypothetical cases will be like this. Accordingly, they would have a privileged status: they would count as *data*.²⁵ This would be the inevitable conclusion even if one's starting point were coherentism rather than a foundationalist theory of *data* (whether truth-based or rule-based).

The above conclusion depends on the assumption that the rationalist capacity theory provides an explanation of the origin of our modal intuitions that is simpler than the various causal explanations. But suppose this not so; suppose instead that some naturalistic causal explanation is simplest. This kind of explanation does not affirm the truth of our modal intuitions, specifically, the modal intuitions used in support of scientific essentialism. It follows that on the coherence theory (as explicated above) these intuitions would not be justified and so in turn scientific essentialism would not be justified.²⁶

Our overall conclusion, therefore, is this. Either coherentism does not validate scientific essentialism, or if it does, it does not provide a genuine alternative to a traditional rationalist theory of *a priori* evidence. Thus, in the quest for a satisfactory theory of the evidential status of intuitions, scientific essentialists have yet to find a way to avoid rationalism.

8. Natural Rationalism

These considerations give scientific essentialists reason for taking rationalism very seriously. But there is a problem: as they stand, the various versions of rationalism are inconsistent with scientific essentialism. They imply that pretty much any elementary necessity could in principle be justified absolutely *a priori*. However, scientific essentialism maintains that a great many elementary necessities cannot be justified except by *a posteriori* means. Let me explain this conflict.

Consider moderate rationalism. Take the following central, clear-cut hypothetical cases: a puddle of water with hydrogen in it; a puddle of water with oxygen in it; a puddle of water with H₂O in it; a puddle of water with pH of approximately 7; a puddle of water that, when subjected to hydrolysis, yields two H₂ molecules for every O₂

molecule; and so on. Are these cases possible? If scientific essentialism is true and if water = H_2O , they are indeed possible. Therefore, according to moderate rationalism, you would have to judge that most such cases are possible. After all, you have the concept of water, and having the concept of water requires having the capacity for making mostly true judgments about whether the concept applies to elementary hypothetical cases such as these. Now take the following central, clear-cut hypothetical cases: a puddle of water with only HO_3 in it; a puddle of water with only H_5O_2 in it; a puddle of water with only $NaCl$ in it; a puddle of water with only XYZ ($\neq H_2O$) in it; a puddle of water with $ph\ 1$; a puddle of water that, when subjected to hydrolysis, yields one H_2 molecule for every two O_2 molecule; and so on. Given the truth of scientific essentialism and given that water = H_2O , these cases are not possible. Therefore, according to moderate rationalism, you would not judge that they are possible. For, as before, you have the concept of water, and having the concept of water requires having the capacity to make mostly true judgments about whether the concept applies to elementary hypothetical cases such as these. Now let us generalize. As you sort through all the possible cases judging whether the concept of water applies, you will make mostly true judgments about possible H_2O items being water and possible non- H_2O things being non-water. But you will be unable to find any, or any but a few, cases in which you judge possible non- H_2O items to be water or possible samples of water to be non- H_2O . Your best account of this would be that necessarily all and only samples of water are samples of H_2O . Since, according to the moderate rationalist thesis, your judgments about hypothetical cases must have a strong tie to the truth, the moderate rationalist counts them as evidence (data). Consequently, your conclusion that necessarily all and only samples of water are samples of H_2O would be justified according to moderate rationalists. What is more, all this evidence and the theoretical inferences you base upon it would be absolutely *a priori*. However, this outcome contradicts scientific essentialism, for according to scientific essentialism, a person could be justified in believing that necessarily all and only samples of water are samples of H_2O only *a posteriori*; it is impossible to have an absolutely *a priori* justification for believing this.

For much the same reason, scientific essentialism is also inconsistent with all the other familiar versions of rationalism, including in particular dialectical and holistic rationalism.²⁷

So here is our dilemma. The only satisfactory theory of the evidential status of intuitions that we have found so far is the rationalist theory, which consists of the rationalists' capacity-theory of determinate concept possession and their truth-based theory of evidence. And in the next section we shall see that Platonism is also unsatisfactory unless it presupposes a certain version of rationalism. As it stands, however, the full rationalist theory contradicts scientific essentialism. So how can scientific essentialism be saved?

A radical answer is to give up all forms of rationalist theory of determinate concept possession (including the new, circumscribed form of rationalism I will propose below) and to put in its place a purely causal theory of determinateness for all concepts. (Examples of this kind of purely causal theory would be a radical externalist baptism-cum-naming-chain theory and a radical externalist historical-explanation theory. Later on I will indicate why such purely causal theories fail as theories of determinateness. For now I will concentrate on how poorly they mesh with a satisfactory theory of the evidential status of intuitions.) Unlike a rationalist capacity theory, a purely causal theory of determinateness makes the occurrence of mostly true intuitions about elementary hypothetical cases (rather than, say, mostly false intuitions about such cases) look like a nonrational, chance phenomenon. The sort of noncognitive causal connections specified in a purely causal theory do not rationally determine a person's judgments about the applicability of the concept to hypothetical cases any more than being confronted with, say, a pair of dice rationally determines a gambler's hunches about what numbers will come up on the dice during a craps game. If the gambler gets it right, it is a matter of pure luck. Likewise, as far as a purely causal theory of determinate concept possession is concerned, if your noninferential judgments about hypothetical cases are right, that it is a matter of pure luck. For this reason, the purely causal theory contributes nothing to the explanation of why a person's intuitions should have any evidential status at all.

So what can an advocate of a purely causal theory of determinateness say about the evidential status of intuitions? Given the failure of empiricism and coherentism and given that purely causal theorists reject all forms of rationalist theory of determinateness, they have no alternative but dogmatism. (There are still Platonist theories—"direct perception" Platonism and rule-of-evidence Platonism. But in the next section I will show that these theories are

of no help to purely causal theorists.) That is, purely causal theorists must just declare dogmatically that some modal intuitions—or perhaps all modal intuitions—simply do have the status of evidence and that this is a brute fact having no explanation. This response is unacceptable. After all, various opponents of scientific essentialism—for example, Quinean empiricists and skeptics—hold that modal intuitions are mere beliefs that have no evidential weight whatsoever. Some beliefs that pop into a person's mind (e.g., gambler's hunches) have no evidential weight and should simply be disregarded in rationally conducted theory justification. What distinguishes modal intuitions from such beliefs? If it is declared that there is no relevant distinction, scientific essentialists are not free to base their strong modal conclusions on them. Scientific essentialism would be seen ultimately to be a mere dogma. (Furthermore, the circumscribed rationalist theory I will propose below provides an *explanation* of the evidential status of intuitions that is consistent with scientific essentialism. So this theory will prove to be superior to the dogmatic causal theory in explanatory power. See the discussion of rule-of-evidence Platonism in the next section for more on this point.)

There is a further problem with adopting the purely causal theory of determinateness. Indeed, it is a problem for any scientific essentialist who, like the purely causal theorist, rejects all forms of rationalist capacity theory of determinateness (including the new, circumscribed form I will propose in a moment). As we have seen, scientific essentialists must hold that some intuitions—or perhaps all intuitions—have the status of evidence. Suppose that only some intuitions have the status of evidence. Which ones? Scientific essentialists cannot very well just declare that, as an unexplainable brute fact, intuitions consistent with scientific essentialism are evidential and intuitions inconsistent with scientific essentialism are not. For example, consider the Kripkean intuitions (that possibly Aristotle did not teach Alexander; that possibly Gödel did not discover the incompleteness theorem; that possibly heat does not produce S ["sensations-of-heat"]; and so forth) and the Putnamian intuitions (that, if all and only samples of water on Earth were composed of H_2O and if there were a twin Earth containing samples of XYZ [$\neq H_2O$] that are macroscopically just like water, then these samples of XYZ would not be water; and so forth.) Contrast these pro-scientific-essentialist intuitions with our old-fashioned anti-scientific-essentialist intuition (e.g., that possibly there is a puddle of water with no hydrogen in

it, that possibly there is a hot thing with no rapidly moving microscopic parts, etc.). Suppose dogmatic causal theorists just declare that, as an unexplainable brute fact, the Kripkean and Putmanian intuitions have the status of evidence but the old-fashioned anti-scientific-essentialist intuitions do not. This declaration would appear to be based on the following principle: admit as evidence those intuitions that get your favorite theory to come out true and dismiss as nonevidential all intuitions that go against your favorite theory. If this is how scientific essentialists proceed, their theory is wholly unjustified.

The alternative is to claim that all—not just certain favored—intuitions have the status of evidence (at least *prima facie*). But in this case, what is one to say about the multitude of our apparent anti-scientific-essentialist intuitions (e.g., that possibly there is a puddle of water with no hydrogen in it or that possibly there is a hot thing with no rapidly moving microscopic parts). If these intuitions are taken at face value and if all intuitions have evidential status (at least *prima facie*), then scientific essentialism would turn out not to be justified, for “half” of our *prima facie* evidence would contradict it. The upshot would be a stalemate between scientific essentialists and their opponents. To avoid this, scientific essentialists must hold that, although, like all other intuitions, our apparent anti-scientific-essentialist intuitions have the status of evidence (at least *prima facie*), they only *seem* to contradict scientific essentialism. In reality, they are *misreported* intuitions that are actually consistent with scientific essentialism. This is the standard line taken by scientific essentialists. The usual diagnosis given by these scientific essentialists is that, unlike our pro-scientific-essentialist intuitions, our apparent anti-scientific-essentialist intuitions are really intuitions of mere epistemic possibility rather than of genuine metaphysical possibility.

Suppose that this is right. Nevertheless, this use of the distinction between metaphysical and epistemic possibility leaves scientific essentialists with a puzzling question. Do we have any intuitions one way or another about the *metaphysical possibility* of a puddle of water with no hydrogen in it or the *metaphysical possibility* of a hot thing with no rapidly moving microscopic parts? The answer, it seems, must be this. When one suppresses all auxiliary empirical beliefs that might contaminate one’s intuitions (e.g., the empirical beliefs that water = H₂O and that heat = mean kinetic energy), one does not really have an intuition one way or the other about such

matters. Most scientific essentialists I have asked take this line about their own intuitions. At any rate, this is what they are forced to say if they claim that all genuine intuitions have the status of evidence at least *prima facie*. After all, when we scrupulously suppress all auxiliary empirical beliefs that might contaminate our intuitions (in particular, our empirical beliefs that water = H₂O and that heat = mean kinetic energy), we certainly do not have the intuition that it is metaphysically *impossible* for there to be a puddle of water with no hydrogen in it or that it is metaphysically *impossible* for there to be a hot thing with no rapidly moving microscopic parts. So if we do have intuitions one way or the other about these things, our intuitions would have to be that they are metaphysically *possible*. But in this case, we would find ourselves right back in the sort of stalemate mentioned above, and scientific essentialism would not be justified.

So if, as they must, scientific essentialists grant *prima facie* evidential status to all correctly reported intuitions, they must acknowledge the following phenomenon. When it comes to the standard anti-scientific-essentialist examples, we lack real intuitions one way or the other about whether they are metaphysically possible; but when it comes to the standard pro-scientific-essentialist examples, we do have real intuitions about whether they are metaphysically possible. But this poses a problem. Why is the cut just here? This phenomenon is a complete mystery on the purely causal theory.

So those who abandon the full rationalist theory of determinateness in favor of a purely causal theory are left with these questions. First, why not treat intuitions on a par with, say, gamblers' hunches; what gives intuitions any evidential status whatsoever? Second, given that some or all intuitions count as evidence, why are the intuitions that support scientific essentialism genuinely evidential whereas those that go against it are either nonevidential or nonexistent? A purely causal theory is entirely silent on these questions; they are complete mysteries. The result is that scientific essentialism cannot be integrated into an acceptable epistemology by anyone who adopts a purely causal theory.

Summing up, scientific essentialism is unacceptable both on a purely causal theory and on the full rationalist theory. The only way to answer the above questions—and thereby to save scientific essentialism—is with a synthesis of rationalism and the causal theory. That is, we should not be trying to “naturalize epistemology”; we

should be trying to "rationalize naturalism." This synthesis of rationalism and naturalism is a new, circumscribed form of rationalism, which we will call *natural rationalism*. It goes as follows.

If the determinate possession of *all* concepts is explained by a rationalist theory, then scientific essentialism is false. If instead the determinate possession of *all* concepts is explained by a purely causal theory, the evidential basis of scientific essentialism is a mystery. However, if the possession of *some* concepts, namely, category and content concepts, is explained by a rationalist theory and the possession of naturalistic concepts is explained by a moderate causal theory designed to mesh with this limited rationalist theory, the dilemma is avoided and the problems will be solved.

According to natural rationalism, our concepts divide into two tiers, one comprised of category and content concepts and the other, of naturalistic concepts. On the first tier are concepts that are by nature intelligible, concepts whose essential relations to one another are accessible to reason at least in principle (i.e., given sufficient attentiveness, intelligence, and memory). On the second tier, by contrast, are concepts that are not by nature intelligible; a great many of their essential relations to one another (and to category and content concepts) are open only to empirical hypothesis. Now on this theory, the determinate possession of a category or a content concept consists in a cognitive capacity, namely, the capacity in cognitively ideal circumstances to correctly apply the concept to hypothetical cases characterized exclusively in terms of other category and content concepts and to correctly manipulate the concept logically with respect to other category and content concepts. (See the remarks at the close of section 10 for an important point of clarification.) Since the identity of a category or content concept is uniquely fixed by the totality of such relations to other category and content concepts, the pattern, in cognitively ideal circumstances, of the person's application and logical manipulation of the concept with respect to other category and content concepts uniquely determines the identity of the person's concept. Because in cognitively ideal circumstances these applications and manipulations must be correct, merely possessing category and content concepts insures that reason in principle (i.e., when exercised in cognitively ideal circumstances) must have a pathway to the truth about essential relations among them. Category and content concepts thus are by nature intelligible at least in the ideal (i.e., at the end of the ideally conducted philosophical dialectic

or theoretical systematization of intuition given sufficient attentiveness, intelligence, and memory). Of course, there is no supposition that human beings could ever fully achieve cognitively ideal circumstances in all respects. However, in a number of cases we could in principle approximate cognitively ideal circumstances in relevant respects, enough to insure the determinateness of a central core of our category and content concepts.

This rationalist theory of determinateness for category and content concepts does not contradict scientific essentialism. Scientific essentialism would be contradicted only if the conditions required for the possession of naturalistic concepts were also provided by a rationalist theory. But on natural rationalism they are not. Instead, natural rationalism offers a moderate causal theory of determinateness for naturalistic concepts. But why do intuitions about the applicability of naturalistic concepts to hypothetical cases (in particular, pro-scientific-essentialist intuitions) count as evidence? Causal elements on their own cannot explain this. The key to the explanation is our intuitions about the applicability of our category and content concepts, whose evidential status is independently insured by the circumscribed rationalist theory of concept possession and accompanying truth-based theory of evidence. These category and content intuitions, together with our knowledge of the naturalistic items involved, determine all our intuitions about the applicability of naturalistic concepts to hypothetical cases. For example, we have such categorial intuitions regarding the applicability of the category concept of a compositional stuff (as opposed to the category concept of a functional stuff). This sort of intuition together with our knowledge that water is a compositional stuff yields exactly the sort of intuitions invoked by scientific essentialists. The intuition about water in Putnam's twin-earth example is just one of infinitely many instances of the following categorial intuition about compositional stuffs W : If all and only samples of W at some place P were samples of a complex compositional stuff $U_n V$ and there were a place P' just like P except that there are samples of a complex compositional stuff XYZ ($\neq U_n V$) that are macroscopically just like the samples of W at P , then these samples of XYZ would not be samples of W . Similarly, the Kripkean intuition that, say, it is metaphysically possible that lakes should not contain water is one of infinitely many instances of the following general categorial intuition about compositional stuffs W and kinds L of natural fluid containers: It is metaphysically possible

that Ls should not contain W.²⁸ And this generalizes.

All our intuitions regarding the applicability of naturalistic concepts to hypothetical cases are determined in this way by our category and content intuitions together with our knowledge of the categories of the naturalistic items involved. Given that our category and content intuitions have the status of evidence, so do the associated intuitions about the applicability of naturalistic concepts. Here then is an explanation of the evidential status of our pro-scientific-essentialist intuitions.

But what about the old-fashioned anti-scientific-essentialist intuitions. According to natural rationalism, all our intuitions about the applicability of naturalistic concepts to hypothetical cases are determined in the indicated way by our category and content intuitions together with our knowledge of the categories of the naturalistic items involved. This provides the key to the explanation of why the anti-scientific-essentialist intuitions are never really about genuine metaphysical possibilities: *The general categorial principles that would determine such naturalistic intuitions intuitively do not hold.* Consider, for example, the proposition that it is metaphysically possible for there to be a puddle of water containing no hydrogen. This is an instance of the following general categorial principle for compositional stuffs W and U: It is metaphysically possible for there to be sample of W containing no U. But, intuitively, this does not hold. (On the contrary, the following general categorial principle holds intuitively: It is metaphysically possible for there to be compositional stuffs W and U such that, necessarily, every sample of W contains U.) For this reason, we have no corresponding naturalistic intuition to the effect that it is metaphysically possible for there to be a puddle of water containing no hydrogen. The general categorial principle that would determine this naturalistic intuition intuitively does not hold. Consequently, our old-fashioned anti-scientific-essentialist intuition that possibly there is a puddle of water containing no hydrogen must be about the mere epistemic possibility rather than a genuine metaphysical possibility.

We have just explained why we do not really have an anti-scientific-essentialist intuition to the effect that it is metaphysically possible for there to be a puddle of water containing no hydrogen. But it is also true that we do not really have the counter-intuition either. That is, when we suppress all auxiliary empirical beliefs that might contaminate our intuitions (e.g., the empirical belief that water = H₂O),

we do not really have an intuition to the effect that it is metaphysically impossible for there to be a puddle of water containing no hydrogen. The explanation parallels that just given. The proposition at issue is an instance of the following categorial principle for compositional stuffs *W* and *U*: It is metaphysically impossible for there to be a sample of *W* that contains no *U*. But, intuitively, this categorial principle does not hold. (Indeed, it is metaphysically possible for there to be compositional stuffs *W* and *U* such that, possibly, there is a sample of *W* that contains no *U*.) However, to have the naturalistic intuition that it is metaphysically possible for there to be a puddle of water containing no hydrogen, this categorial principle about compositional stuffs would have to be intuitive. Because it is not, we do not really have the associated naturalistic intuition about water. Those who claim to have this naturalistic intuition have just failed to suppress fully their contaminating auxiliary empirical beliefs; or perhaps they are still misreporting their intuitions, say, by confusing epistemic possibility with metaphysical possibility.

The upshot is that we do not really have any intuition one way or the other concerning the metaphysical possibility of a puddle of water containing no hydrogen. And this generalizes to other issues of metaphysical possibility like this. The only intuitions we have concerning genuine metaphysical possibilities involving naturalistic items are those determined in the indicated way by category and content intuitions together with our knowledge of the categories of the naturalistic items. Although pro-scientific intuitions are like this, none of the old-fashioned anti-scientific intuitions are. (Incidentally, natural rationalism can be used in analogous ways to explain all the fascinating patterns in our intuitions, uncovered in section two, concerning food, fuel, person, life, and so forth.)

This, then, is a intuitive, economical, and non-*ad hoc* way of explaining why the cut in our intuitions falls right where it does. Natural rationalism does exactly what is needed to save scientific essentialism. It provides an epistemological account of the evidential status of the intuitions that are used to defend scientific essentialism, and it explains why the other, anti-scientific-essentialist intuitions that we seem to have cannot be intuitions about genuine metaphysical possibilities and so must be intuitions about mere epistemic possibilities. Accordingly, the otherwise mysterious cut between intuitions that count as evidence regarding genuine metaphysical possibility and our other, only apparent intuitions about metaphysical possibility is not an ar-

bitrary or *ad hoc* cut at all. And thus we have found a way to make scientific essentialism epistemologically acceptable, a fully satisfactory explanation of the evidential status of its supporting intuitions.²⁹

9. Platonism

Is there no alternative to natural rationalism? What about some kind of Platonist epistemology? I will now show that Platonists who accept scientific essentialism must at least implicitly accept natural rationalism. Platonist theories of the evidential status of intuitions come in two varieties, "direct perception" theories and rule-of-evidence theories. (Both varieties are suggested by Kurt Gödel's philosophical work on mathematical knowledge although Gödel seemed not to distinguish between them.³⁰) According to the "direct perception" theory, intuitions are a kind of "direct perception" of abstract truths: when normal, intelligent people consider the question of whether a concept applies to an elementary hypothetical case, in most instances they just "see directly" that it applies if and only if it truly does apply. (Of course, misfires can occur, but they are the result of some deficiency in the cognitive circumstances. As the cognitive circumstances approach the ideal in relevant respects, the occurrence of an intuition is nothing but the unmediated "seeing" of an abstract truth.) If the Platonist "direct perception" theory were true, it would certainly be necessarily true; that is, if intuitions are a kind of "direct perception," then necessarily they are a kind of "direct perception." Accordingly, on the Platonist "direct perception" theory, intuitions have a strong modal tie to the truth. Now suppose that a truth-based theory of evidence is correct. (On a truth-based theory, a type of belief qualifies as evidence if beliefs of that type have a strong modal tie to the truth. There is great *intuitive* support for this sort of theory, and it is a component of an attractive explanation of the empiricist tenet that elementary propositions about one's current experiences are basic data.) Then, given the "direct perception" theory, it would follow that intuitions have the status of evidence. In this way "direct perception" Platonism, in tandem with a truth-based theory of evidence, yields an explanation of the evidential status of intuitions.

The other variety of Platonism is a rule-of-evidence theory. On this theory there is a close epistemological analogy between empirical

science and a *priori* science (logic, mathematics, and philosophy): the former is the result of theoretical systematization of our sense perceptions; the latter, the result of theoretical systematization of our intuitions. In empirical science all sense-perceptual propositions are taken as *prima face* evident (justified) and they remain so throughout the formulation of one's best comprehensive empirical theory. Once this theory is formulated, it too acquires the status of being evident. Of course, some sense perceptions are illusory. The best theory will identify the associated sense-perceptual propositions, and once these propositions have been isolated, they lose their status of being evident. According to rule-of-evidence Platonism, a *priori* theorizing is wholly analogous to empirical theorizing except that intuitions play the role of sense perceptions. All intuitions are taken as *prima face* evident (justified), and they remain so throughout the formulation of one's best a *priori* theory. Once this theory is formulated, it too acquires the status of being evident. Of course, some intuitions are mistaken. One's best a *priori* theory will identify which these are, and once they have been isolated, they lose their status as being evident.

Now "direct perception" Platonism attributes to intuitions a strong modal tie to the truth, and because of this, it is able to give an *explanation* of the evidential status of intuitions by means of a truth-based theory of evidence. Unlike a truth-based approach, a rule-of-evidence approach is content merely to *describe* our rules of evidence, where these rules are to be thought of as functioning rather like the rules of evidence used in a courtroom trial. (This approach is thus like that taken by Chisholm on the question of perceptual knowledge and by Strawson on the question of inductive knowledge.) Such rules identify certain types of proposition that can unproblematically be used to justify other propositions. Specifically, our rules of evidence tell us that perceptual beliefs are evident and, analogously, that intuitions concerning the applicability of a concept to elementary hypothetical cases are evident. Furthermore, these rules tell us that, by systematizing one's perceptual beliefs, one arrives at a physical theory that is evident. Similarly, these rules tell us that, by systematizing one's intuitions, one arrives at an a *priori* theory (logical, mathematical, or philosophical) that is evident.

I do not wish to challenge these two versions of Platonism. What I wish to do is to show their relationship to rationalism.

I begin with rule-of-evidence Platonism. Let us suppose that intui-

tions are *prima facie* evident. Someone might dogmatically declare this to be brute fact that is in principle unexplainable. (The dogmatic causal theorist from the previous section held this, too.) However, there would be a considerable gain in theoretical economy and explanatory power if, using antecedent principles, we could derive the principle that intuitions are *prima facie* evident. Ockham's razor deems a theory superior if, other things being equal, it posits fewer principles than a competing theory. This is one important advantage of natural rationalism. It relies on a theory of evidence (namely, a truth-based theory) that is supported by a wealth of intuitive data, and using this theory plus the thesis that there is a strong modal tie between a person's intuitions and the truth, one can then derive rule-of-evidence Platonism as a consequence. Moreover, this thesis is not itself a first principle. It is a consequence of the natural rationalist theory of what it is for a person to possess a concept determinately, a theory that is also supported by a wealth of intuitive data. What theory of determinate concept possession would rule-of-evidence Platonists adopt in a comprehensive theory? Presumably, they too would be forced in the direction of the natural rationalist answer, for their theory obliges them to honor and to systematize the very intuitions that support this theory. The conclusion is that natural rationalism does not contradict rule-of-evidence Platonism; rather, it is a simpler, explanatorily more powerful theory that yields rule-of-evidence Platonism as a consequence. Other things being equal, such a theory should be adopted as superior. Indeed, this conclusion follows by *the very standards* advocated by rule-of-evidence Platonists, namely, that we should adopt the theory that is the simplest comprehensive systematization of our intuitions.

But this is not all. We have seen that, if scientific essentialism is justified, there must be a curious "cut" in our intuitions. When it comes to the standard anti-scientific-essentialist examples, we lack real intuitions one way or the other about whether they are metaphysically possible; but when it comes to the standard pro-scientific-essentialist examples, we do have real intuitions about whether they are metaphysically possible. Why is the cut just here? In their best comprehensive *a priori* theory, what explanation would be adopted by rule-of-evidence Platonists who are scientific essentialists? As we have seen, purely causal (radical externalist) theories make this phenomenon a complete mystery. The simplest explanation is that provided by natural rationalism. (Or perhaps "direct perception"

Platonism; see below.) Therefore, *by their very own standards* rule-of-evidence Platonists who are scientific essentialists have this decisive reason for supplementing their theory with natural rationalism (or perhaps "direct perception" Platonism). The conclusion is that rule-of-evidence Platonism does not provide a serious alternative to natural rationalism. By their very own standards, rule-of-evidence Platonists are forced to accept natural rationalism (or perhaps "direct perception" Platonism).

So let us now turn to "direct perception" Platonism. I have two observations to make. The first concerns the question of how on this theory the aforementioned "cut" in our intuitions is to be explained? Suppose that a person could (at least in principle) have "direct perceptions" of all types of metaphysical possibilities, including those concerning the applicability of naturalistic concepts to hypothetical cases characterized in terms of naturalistic concepts. Then, by the inference route characterized at the outset of the previous section, the person could (at least in principle) know absolutely *a priori* that necessarily water = H₂O, thereby contradicting scientific essentialism. Thus, like unrestricted rationalism, unrestricted "direct perception" Platonism is inconsistent with scientific essentialism.

Short of giving up scientific essentialism, the only plausible way for "direct perception" Platonists to solve this problem would be to mimic the natural rationalist solution. Accordingly, they would hold that a person cannot have "direct perceptions" of all types of metaphysical possibilities but only of those that concern the applicability of category and content concepts to hypothetical cases characterized exclusively in terms of category and content concepts. However, like natural rationalism, this circumscribed "direct perception" Platonism implies the autonomy of philosophy thesis. All that is needed to show this is the premise, which will be discussed below, that (most of) the central questions of philosophy are formulated in terms of category and content concepts. The conclusion is that "direct perception" Platonism provides no hope for global scientific essentialists; on this Platonist theory, local scientific essentialism implies the autonomy of philosophy.

My second observation about "direct perception" Platonism concerns the question of how on this theory errors in intuition are possible. For the answer let us turn to the analogy Platonists draw between sense perception and intuition. What makes an erroneous sense perceptual belief possible? First, the perceived object itself could

be somehow defective (for example, it might be a fake or a cleverly constructed physical illusion or something of the sort). Second, there could be an aberration in the observation conditions. Third, there could be an aberration in the person's sense organs. Fourth, the person's perceptual belief could be formed under the influence of a strongly countersuggestive false opinion or desire. Now if no such problematic factor were present, would errors in perceptual belief be metaphysically possible? Intuitively, the answer is yes. For intuitively the laws governing the transmission of the physical signal from the object to the sense organs could, as a metaphysical possibility, be different from what they in fact are. (My argument does not depend on this claim. However, it is heuristically useful in setting up the contrast between ordinary sense perception and Platonist "direct perception.")

Let us now see how errors in one's intuitions are possible on "direct perception" Platonism. Suppose we are interested in the applicability of a given category or content concept to a hypothetical case characterized exclusively in terms of category and content concepts. First, suppose that all distinctions relevant to the case are explicitly marked and that the case is not borderline but central, clear-cut, and vivid. This distinctness condition is analogous to the above condition that the object of sense perception should not be defective in any way. Second, suppose that the person gives the case full attention, noting clearly all of the distinctions marked in the case. This clarity condition is more or less analogous to the condition that in sense-perception the observation conditions should be good. Third, suppose that the person has ample intelligence. So, for example, the situation is not like one in which the person fails to "get" the Barber Paradox: with ample intelligence, a person can tell straight off that a barber who shaves all and only people who do not shave themselves is not metaphysically possible. This intelligence condition is more or less analogous to the condition that in sense perception a person's sense organs should be well-functioning. Fourth, suppose that our person has gone through all relevant philosophical dialectic and theoretical systematization of his intuitions, and suppose that throughout the process all of the other cognitive conditions (distinctness, clarity, intelligence) have been met and that throughout there has been neither memory lapse nor distracting desire. This condition plays more or less the same role as the condition that in sense perception there should be no countersuggestive

beliefs or desires. If dialectic and theoretical systematization of one's intuitions were conducted in such cognitively ideal conditions, the effect would be that impure intuitions resulting from distorting philosophical "pictures" or other mere opinions (*doxa*) would be identified and discharged.

If all these cognitive conditions are met, is it metaphysically possible that at the end of the process the person could be in error concerning his intuition about the application and logical manipulation of category and content concepts *vis à vis* other category and content concepts? In the sense perception case, error still seemed metaphysically possible because it seemed metaphysically possible that the relevant laws governing the external causation of the sense experience could have been different. But in the case of intuition there simply is no counterpart of the transmission of the signal from the object to the sense organs. So the question of the contingency of the associated causal laws cannot come up: there simply are no such laws at work in the case of intuition. If "direct perception" Platonism is right, there is no "external" medium. Indeed, the whole point of "direct perception" Platonism is that intuition is *direct*. In cognitively ideal circumstances all potential "internal" deficiencies are removed; in such circumstances it is impossible not to "see directly" the relevant necessary relations among category and content concepts. Or at least this is what the "direct perception" Platonists are committed to holding. But this capacity for necessarily error-free "direct perception" in cognitively ideal circumstances is just a strong form of the kind of cognitive capacity posited by natural rationalism. The upshot, therefore, is that "direct perception" Platonism implies the truth of natural rationalism. (We need take no stand on whether the converse implication holds.)

So in our quest for a genuine alternative to natural rationalism that can also serve the needs of scientific essentialism, we have come up short. Our overall conclusion is that (some form of) natural rationalism is needed for an adequate scientific essentialism. It provides the only satisfactory epistemological account of the evidential status of the intuitions that are used to defend scientific essentialism and the only plausible explanation of the otherwise mysterious "cut" between our pro-scientific-essentialist intuitions and our other, apparent anti-scientific-essentialist intuitions. Only with the aid of natural rationalism can scientific essentialism be made epistemologically acceptable.

10. The Determinateness of Our Naturalistic Concepts

Although natural rationalism solves the problem of the evidential status of pro-scientific intuitions, there is still the question of what is required to possess a naturalistic concept determinately. After all, two problems confront scientific essentialism—the evidential status of our intuitions and the determinate possession of our naturalistic concepts. Natural rationalism and the accompanying truth-based theory of evidence provides the only satisfactory solution to the first problem. This theory incorporates a circumscribed rationalist account of the determinate possession of our category and content concepts and a moderate causal account of the determinate possession of our naturalistic concepts. But what are the details of this causal theory? My thesis is that causal theories are unsatisfactory unless supplemented with the theory that our thought and talk about naturalistic items is mediated by a background of appropriate category and content concepts for which the circumscribed rationalist theory of determinateness holds.³¹ If this thesis can be defended, the natural rationalist solutions to the two problems facing scientific essentialism will mesh perfectly. The perfectness of the fit will be the final confirmation that scientific essentialism depends on natural rationalism.

I have said that a satisfactory causal theory must be supplemented with the theory that our thought and talk about naturalistic items is mediated by background category and content concepts for which a rationalist theory of determinateness holds. In the sequel to this essay, which is planned for the second volume of this series, I will give the argument for this thesis in detail. But it is possible to indicate some parts of the argument here. As we have seen, there are two versions of the causal theory—the radical, purely causal version and the moderate, natural rationalist version. On the radical version, a person has a concept of a naturalistic item if the person is suitably situated with respect to the item; the person need not employ mediating category and content concepts. On the moderate version, the person not only must be suitably situated with respect to the item but also must employ mediating category and content concepts (or logical combinations of category and content concepts). Our argument is designed to show that the moderate, natural rationalist theory is acceptable but the radical theory is not.

The problem with the radical theory is underdetermination: “external” features of the situation do not uniquely fix the object of a

person's thought. For example, merely looking, and pointing, in the direction of a quantity of water does not suffice for thinking of water, the compositional stuff. The object of thought, if there is one, could just as well be H_2O -with-exactly-those-impurities, monoxide, dihydrogen compound, oxygen compound, hydrogen compound, and so forth. Besides these compositional stuffs, various functional stuffs might be the object of thought, for example, drink, washing liquid, solvent, and so forth. And various institutional stuffs might be the object of thought, for example, sacred water or National Drink (if the National Drink of a certain country were changed from, say, whisky to water during Prohibition). Or very general stuff notions—compound, compositional stuff, functional stuff, liquid, or even matter itself. Finally, the object of thought might not even be a stuff at all; it could be the particular quantity of water or some individual substance that is constituted by that quantity of water (e.g., a particular lake) or a natural kind to which that individual substance belongs (e.g., lake) or the place where the quantity of water is (i.e., there). I claim that the only way that a determinate object of thought can be fixed is by employing relevant mediating category and content concepts. What alternative is there?

Merely uttering the word 'water' does not settle the question. After all, the person could be a fluent speaker of two languages, English (in which 'water' refers to H_2O) and twin-Earth English (in which 'water' refers to XYZ). Is the person thinking of H_2O or of XYZ? The case is underdetermined. When this bilingual speaker utters the term 'water', what determines whether he is thinking of water rather than XYZ? Perhaps this is determined by his intention to refer to water. But what determines whether he has this intention as opposed to the intention to refer to the other stuff? Perhaps this is determined by his intention to speak English rather than twin-Earth English. If so, what is involved in having this intention? Surely the answer is that he must be employing relevant category and content concepts. For example, a broadly Gricean analysis of this intention would attribute to the person a rich array of philosophically important category and content concepts—for example, the concepts of intending, thinking, wanting, perceiving, cause, reason, and so forth. (Of course, the paradox of analysis shows us that the person need not have an articulated conscious awareness of using these concepts.)

The same conclusion would hold even if our person is a monolingual English speaker. Merely uttering 'water' would not guarantee

that a monolingual English speaker is thinking of water. The person might not be paying attention; instead, he might just be mouthing words and staring off into space.

Or the person might have made a slip of the tongue. What further conditions must be met in order to pay attention to what one is doing while speaking a language? What further conditions must be met in order to intend to refer by uttering the term 'water'? These mental acts cannot occur in a cognitive vacuum. The person must be employing relevant category and content concepts. Any plausible analysis of this intention (for example, any broadly Gricean analysis) would attribute to the speaker a rich array of category and content concepts.

Consider another example. Merely looking (and even pointing) in the direction of the ship of Theseus does not suffice for thinking of the ship. You might instead be thinking of the wood. What is the difference? It does not lie in the external environment; the external environment is the same whether you are thinking of the ship or the wood. The difference must lie in the mediating category and content concepts you are employing.

Now sophisticated radical externalists might concede that, to think of the wood, you must employ mediating category and content concepts. But they might also hold that, to think of the ship, no such mediating concepts are required because the ship is the *metaphysically basic* (or otherwise most *salient*) item in the context. This sophisticated response has a wealth of defects. I have time to mention only one.

Consider any case in which you are thinking of the ship (including one in which you are allegedly not employing any mediating category or content concepts). There could always be an externally similar case in which you are thinking of the wood instead and in which your use of mediating category and content concepts is neither more nor less in evidence to you phenomenologically than it is in the case in which you are thinking of the ship (where allegedly you are using no such concepts). Just try it. The sophisticated externalist claim (i.e., that thinking of the wood requires the use of mediating category and content concepts but thinking of the ship does not) simply does not match the phenomenological reality, for in each case, the use of mediating category and content concepts is *equally* in evidence phenomenologically, no more in the one case than in the other. But sophisticated radical externalists have already been forced to agree

that mediating category and content concepts *must* be used in thinking of the wood. Therefore, the faithful—and theoretically uniform—way to characterize the phenomenological reality is to say that in *both* cases (thinking of the ship and thinking of the wood) you are employing relevant category and content concepts.

Of course, typically you have no articulated, conscious awareness of employing such category and content concepts. However, using concepts without having an articulated, conscious awareness of doing so is an extremely widespread phenomenon which radical externalists must countenance for independent reasons. (This phenomenon has close affinities to the paradox of analysis, for example. For another analogy, consider the duck/rabbit example. The situation is the same externally whether you see a duck or a rabbit; the difference between seeing a duck and seeing a rabbit is tied, at least in part, to differences in the way you conceptualize the situation. Of course, you need not have any articulated, conscious awareness of employing any concepts. But this hardly shows that you are not employing concepts. You must be. Otherwise, there would be no satisfactory explanation of the difference between seeing a duck and seeing a rabbit.) So our method of characterizing the ship/wood example requires no special theoretical apparatus that would not already be needed by radical externalists themselves. Thus, our method is just as simple as the sophisticated radical externalist method. At the same time, our method is theoretically more uniform than the sophisticated radical externalist method inasmuch as it does not hypothesize that a fundamentally different mechanism is at work in thinking of the ship than the one at work in thinking of the wood. And unlike the sophisticated radical externalist method, our method is fully faithful to the phenomenological reality inasmuch as the use of mediating category and content concepts is *equally* in evidence phenomenologically in the case of thinking of the ship as it is in the case of thinking of the wood. (I should emphasize again that these phenomenological considerations provide only one of many lines of argument against sophisticated radical externalism.)

But now we are left with a question. How on our moderate theory can the determinate possession of our mediating category and content concepts be explained? If, as in the case of the determinate possession of naturalistic concepts, we tried to explain the determinate possession of these concepts by means of a moderate causal theory, we would get caught in a vicious explanatory regress: the

determinate possession of mediating category and content concepts would require the determinate possession of other mediating category and content concepts, and so on *ad infinitum*. We thus would have failed to give a satisfactory explanation of the determinate possession of *any* of these concepts, the original naturalistic concept or the associated category and content concepts. For this reason, there is no reasonable alternative but to explain the determinate possession of the category and content concepts by means of another theory. But there is none except the circumscribed rationalist theory of determinateness that is incorporated in natural rationalism. This is exactly what we set out to show.

An important point of clarification is now in order. If scientific essentialism is right (and, relatedly, if the recent critique of individualism in philosophy of mind is right), a person could be deeply ignorant about whether a given naturalistic concept applies to relevant hypothetical cases, and nevertheless the person could have numerous beliefs involving the concept. For example, a person could be wholly in the dark about whether it is metaphysically possible for there to be a puddle of water containing no hydrogen, and yet the person could have countless beliefs about water, for example, that water is wet, clear, potable, thirst-quenching, and so forth. Similarly, a person who is deeply ignorant about what arthritis is could nevertheless have beliefs about arthritis, for example, that he has arthritis in his own thigh. (At least this is what we would say in everyday conversations about a person's beliefs. Similar cases explored in the literature include Putnam's beech/elm example, Fodor's brisquit example, Kripke's Londres/London example, and the Castaneda-Perry examples.) Notice, however, that in a somewhat analogous way a person could be deeply ignorant about a given category or content concept and nevertheless seem to have various beliefs about it. For example, having overheard philosophers talking about Aristotelian primary substances, someone might say, "I think that substances are stuffs." At least in everyday conversation, we would say on this basis that the person does believe that substances are stuffs. Or having overheard artists talking about uncommon colors, a person might say, "I think that mauve is a shade of yellow." At least in everyday conversation we would say on the basis of this that the person believes that mauve is a shade of yellow. Let us suppose that this way of talking about beliefs is literally correct. (We may call beliefs like these *cognitive commitments*.³² In this terminology, a person

can be cognitively committed to a proposition while being deeply ignorant about essential properties and relations of the concepts involved, including even category and content concepts.) The above examples show that there is a mode of possessing category and content concepts that does not require the person to have the associated cognitive capacity required by natural rationalism (i.e., the capacity in cognitively ideal circumstances to correctly apply and logically manipulate the concept *vis à vis* other category and content concepts). In fact, this mode of possessing category and content concepts is exactly like that in which *all* naturalistic concepts *must* be possessed, the mode explained by our moderate causal theory. But this fact is entirely consistent with natural rationalism. Indeed, the main point of natural rationalism is that there must be *another* mode in which category and content concepts (but not naturalistic concepts) can be possessed, namely, the mode explained by the circumscribed rationalist model. Here determinate possession of a concept consists in a cognitive capacity, namely, the capacity in cognitively ideal circumstances to correctly apply the concept to hypothetical cases characterized exclusively in terms of other category and content concepts and to correctly manipulate the concept logically with respect to other category and content concepts.³³

The determinate possession of no concept—whether a naturalistic concept, a category concept, or a content concept—can ever be explained just by “external” causal mechanisms. Causal factors must be accompanied by the employment of relevant category and content concepts whose determinate possession is explained instead by the circumscribed rationalist model. For without the employment of such background category and content concepts, our thought would always be undetermined, and unless the determinate possession of these category and content concepts were explained by the circumscribed rationalist theory, there would be a vicious explanatory regress. Furthermore, if the possession of relevant category and content concepts could not be explained in this fashion, there would be no way to explain the evidential status of our naturalistic intuitions (including our pro-scientific-essentialist intuitions). For, as we have seen, the evidential status of these intuitions depends on that of relevant category and content intuitions, and the evidential status of the latter intuitions derives from their strong modal tie to the truth, which can be insured only by the circumscribed rationalist theory.

This is exactly what we set out to show. Both in the explanation

of the evidential status of pro-scientific essentialist intuitions and in the explanation of the determinateness of our naturalistic concepts, the circumscribed rationalist theory of the determinateness for our category and content concepts seems to be unavoidable. This perfectness of fit—that the same theory seems to be unavoidable in *both* explanations—is the final confirmation of the theory.

11. The Autonomy of Philosophy

From the rationalist theory of determinateness for our category and content concepts, however, the thesis of the autonomy of philosophy follows directly. All that is required is the premise that (most of) the central questions of philosophy are formulated in terms of category and content concepts. The full argument for this premise must be deferred to the projected book on the autonomy of philosophy. However, I can make a few suggestive remarks to indicate the direction of the argument.

The natural rationalist explanation of the evidential status of pro-scientific-essentialist intuitions (and the associated explanation of why we lack genuine anti-scientific-essentialist intuitions) provides two kinds of “transcendental” test—one for category concepts and one for content concepts. One component of the natural rationalist explanation makes use of general categorial principles that determine (or fail to determine) corresponding naturalistic intuitions; another component makes use of the distinction between epistemic possibility and metaphysical possibility. The former component provides a “transcendental” test for category concepts; the latter component provides one for content concepts.

Consider the system of general categorial principles that would be invoked in the best natural rationalist explanation of our pro-scientific-essentialist intuitions. A catalogue of category concepts can be extracted from this system of general categorial principles. Our thesis is that the category concepts used in formulating (most of) the central questions of philosophy will belong to this catalogue. Strong support for this thesis is already at hand. By constructing our best provisional explanation of the pro-scientific-essentialist intuitions, we can extract a provisional catalogue of category concepts. When we do this, we already find an impressive list of philosophically important category concepts: matter, composition, function, purpose, cause, sensible quality, living being, person, and so forth. If we press ahead

with this method, I believe that (nearly) every philosophically important category concept will eventually be added to our catalogue.

The explanation of our apparent anti-scientific-essentialist intuitions depends on the distinction between epistemic possibility and metaphysical possibility. But what is epistemic possibility? The explication of this notion turns on the notion of *qualitatively identical epistemic situation*. Epistemic situations are qualitatively the same only if they involve the same content concepts (phenomenal qualities and conscious operations of mind). So as we sort through the diverse epistemic possibilities that must be invoked to explain away our apparent anti-scientific-essentialist intuitions, we will obtain a catalogue of content concepts. Even a provisional catalogue already contains an impressive list of philosophically important content concepts: sensing, feeling, loving, knowing, acquaintance, consciousness, and so forth.

Another style of "transcendental" argument exploits the dependence of scientific essentialism on our moderate causal explanation of the determinateness of our naturalistic concepts. To explain how a person can determinately possess a specific naturalistic concept, one must invoke further concepts whose determinate possession must be explained by the circumscribed rationalist model. These further concepts are category and content concepts. Therefore, if we can show that a concept is of this type, it may be added to our catalogue of category and content concepts. By this alternate route, then, we might be able to get additional "transcendental" confirmation that a number of philosophically important concepts belong in our catalogue.

By pursuing these three "transcendental" strategies, we can, I think, show that the concepts used in formulating (most of) the central questions of philosophy are category and content concepts. We have seen that local scientific essentialism is committed to a limited rationalist theory of the determinate possession of these concepts. But this rationalist theory implies that, if a proposition involving exclusively category and content concepts can be known to be necessary, then in principle it can be known to necessary absolutely *a priori*, without the aid of empirical science. Therefore, global scientific essentialism is untenable. Local scientific essentialism implies the autonomy of philosophy.

Notes

1. This paper is a revised and expanded transcription from tapes of the talk I gave at the spring 1986 Greensboro Conference on Identity and Existence. I also spoke on this material that spring at Dartmouth College and CUNY Graduate Center. This is work in progress. A sequel on the problem of the determinateness of our concepts is planned for the next volume of this series. And a book on the general topic of the philosophical limits of science is now underway. My warmest thanks to Mark Bedau, Tyler Burge, Peter van Inwagen, Jerrold J. Katz, George Myro, C.D.C. Reeve, Steve Rieber, David Rosenthal, Nathan Salmon, Ernest Sosa, Neil Thomason, Adam Thompson, and James Tomberlin.
2. Wallace I. Matson, "Against Induction and Empiricism," *Proceedings of the Aristotelian Society*, 1962, 143-158. Saul Kripke, *Naming and Necessity*, Cambridge, Mass., 1980 (originally published as "Naming and Necessity," in *Semantics of Natural Language*, D. Davidson and G. Harman, eds., Dordrecht, 1972). Hilary Putnam, "The Meaning of 'Meaning'," *Philosophical Papers II: Mind, Language, and Reality*, Cambridge, 215-271 (originally published in *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge*, K. Gunderson, ed., Minneapolis, 1975).
3. Saul Kripke states that theses like the modified Kantian thesis "involve the obscure notion of the possibility of a *priori* knowledge, but to the extent that the notion is clarified by restricting it to a *priori* knowledge of a standard human sort, I argue against [such theses] in the text." (P. 159f., *Naming and Necessity*) However, *Naming and Necessity* attempts to clarify the notions of possibility and of a *priori* knowledge. If it is successful, the notion of the possibility of a *priori* knowledge cannot be obscure, for it is by definition simply a *priori* knowledge that is possible. As I read *Naming and Necessity*, its primary thesis is just the denial of the modified Kantian thesis. That is, its primary thesis is that there are some necessary truths—e.g., that water = H₂O—that it is impossible for a person to know a *priori* to be necessary. The restriction to human beings plays no role in the thesis or in the arguments; if it did, a good bit of the philosophical interest would be lost. So when Kripke speaks of knowledge of the standard human sort, he probably intends to rule out knowledge obtainable by an infinite (versus finite) intelligence or by what Kant calls "original intuition" (B72, B138-139, B145).
4. This thesis is (pretty much) equivalent to the following: for (most of) the central truths of philosophy, if it is possible to know them to be true, it is possible to know them to be true absolutely a *priori*. A "proof" of this (near) equivalence is this. For (most) central philosophical propositions *p*, we know a *priori* that the following principle holds: *p* is true if and only if *p* is necessary. Accordingly, it is possible to know a *priori* that *p* is true if and only if it is possible to know a *priori* that *p* is necessary. Similarly, it is possible to know that *p* is true if and only if it is possible to know that *p* is necessary. From these two biconditionals, the (near) equivalence follows directly.

5. Pp. 103-5, 141-4, 150-3, especially p. 104, *Naming and Necessity*.
6. Of course, this is not quite accurate, for a mutton stew has ingredients (e.g., pieces of turnip) that count as food even though they do not in isolation count as (quantities of) mutton stew. This kind of inaccuracy can be avoided by modifying the example slightly. For example, suppose that at some time the only food on Earth is chocolate. If our space cadets find carob rather than chocolate on twin Earth, plainly they should count it as food even though it is microscopically (and macroscopically) different from chocolate. Why? For another example, consider perfume. Suppose all and only perfume here on Earth is composed of P_1, \dots, P_n . On twin Earth, they do not have those chemicals. Nevertheless, they have stuffs composed of M_1, \dots, M_m . These stuffs smell very nice; indeed, let us suppose that they are observationally indistinguishable from P_1, \dots, P_n . Twin Earthlings apply these stuffs to their necks and wrists for the purpose of smelling nice. In fact, these stuffs are subsequently brought back to Earth to be used for the same purpose. Certainly when our space cadets discovered these stuffs on twin Earth, they were correct in counting them as perfume. Later on, these stuffs are (nonfraudtently) marketed as perfumes back on Earth. How can this be given that their chemical constitution differs from P_1, \dots, P_n ?
7. At this stage, we need not venture a precise analysis of the concepts of compositional stuff and functional stuff. As the paradox of analysis teaches us, a precise analysis is called for at the end, not the beginning, of an analytical inquiry. However, the following *provisional* principles might be illuminating for heuristic purposes: a stuff is purely compositional if and only if, necessarily, samples of it all have the same composition; a stuff is compositional if and only if, for some fixed family of purely compositional stuffs, necessarily, each sample of the original stuff is composed of some purely compositional stuffs in that fixed family. Intuitively, functional stuffs (or at least most functional stuffs) are not like this. Functional stuffs are defined by their function. A functional stuff is essentially *for* something; the composition of the various samples of a functional stuff makes no difference as long as they are *for* the right thing.
8. Of course, like food, drink is a functional stuff. Both water and XYZ are kinds of drink.
9. I am suppressing the possibility of learning that a given truth is necessary on the basis of the testimony of an authority. The point is that ultimately someone, you or the original authority, must go beyond empirical evidence in order to learn that a given truth is necessary. In any event, scientific essentialists do in fact go beyond empirical evidence to justify the modal step. They make liberal use intuitions as evidence, and without this use of intuitions they would have no justification of the modal step.
10. P. 42, *Naming and Necessity*.
11. One type of hypothetical case is just that of a (metaphysically) possible case, that is, a concretely characterized (metaphysically) possible situation of the sort one might invoke in giving a counterexample to a candidate definition. Kripke's hypothetical cases—that it was possible for

Aristotle not to teach Alexander, etc.—are examples of this type of case. Another type of hypothetical case is expressed in language by means of a subjunctive conditional: if such and such were the case, then so and so would be also be the case in that situation. Putnam's twin-Earth case is an example of this type of hypothetical case.

12. A term is, by definition, a rigid designator if it denotes the same item in every "possible world" in which the item exists.
13. See Arne Naess, "Toward a Theory of Interpretation and Preciseness," in *Semantics and the Philosophy of Language*, L. Linsky, ed., Urbana, 1952, 248-269.
14. Giving such a stipulative definition may thus be viewed as the linguistic counterpart of forming a complex descriptive concept (on the traditional theory of the origin of ideas); in each case the person would be in a position to know *a priori* an associated *analytic* proposition.
15. Let me be more precise about what a theory of determinateness should do. Suppose that a person has some concept *k* but we do not know which concept it is. Let *c* be some familiar concept. (For example, *k* might be the concept of being a multigon, which we just discussed, and *c* might be the concept of being a closed plane figure or the concept of being a closed plane figure with more than four sides.) What is required for the person's concept *k* is to be this familiar concept *c*? A theory of determinateness is designed to answer this question: the person's concept *k* is identical to the concept *c* if and/or only if It is understood that the right hand side does not trivially identify the person's concept *k* and the familiar concept *c*; rather, the right hand side specifies, in a noncircular fashion, features of the person and the circumstances that are sufficient and/or necessary for the person's concept *k* to be the familiar concept *c*.

An analogy will be helpful. In metaphysics we often seek identity criteria for items of a given ontological type, for example, criteria for the identity of physical objects. What does it take for a physical object *s* to be some familiar physical object *s'*? Our identity criteria should tell us: *s* is identical to *s'* if and/or only if It is understood that the right hand side does not trivially identify *s* and *s'*; rather, the right hand side specifies, in a noncircular fashion, features of *s* and *s'* that are sufficient and/or necessary for *s* to be *s'*. (For example, the right hand side might specify that *s* be the closest-continuer of *s'*.) A theory of determinateness is quite analogous except that it is answering a question in epistemology rather than metaphysics.

Incidentally, suppose our argument is correct that some other theory, besides the causal theory, is needed to explain the determinate possession of (at least some of) our category and content concepts. Then we will also have produced an argument against *radical* anti-individualism, the doctrine that *no* concepts (not even category and content concepts) are "in the head."

16. To go from ' $x = x$ ' to 'Necessarily $x = x$ ' in this derivation of the necessity of identicals, we need to apply the necessitation rule to the open-sentence ' $x = x$ '. However, in "Semantical Considerations on Modal Logic," Kripke

restricts application of the necessitation rule to closed sentences in order to block the derivation of the converse of the Barcan formula. Those advocating the indicated derivation of the necessity of identicals from Leibniz's law need to patch up this detail.

17. George Myro, "Identity and Time," in *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, R.E. Grandy and R. Warner, eds., Oxford, 1986, 383-409.
18. This theme is mentioned by Myro in "Identity and Time," p. 403 n.
19. Quine holds that there are no purely behavioral tests for synonymy or analytic equivalence. Consider a widely believed, contingently true identity sentence (e.g., 'Aristotle = the teacher of Alexander', 'Gold = the metal used for stabilizing international currencies') or a widely believed, contingently true biconditional (e.g., 'All and only aqueous items are composed predominantly of the compound that fills the rivers and lakes'). Relative to *extensional* contexts, the relevant linguistic behavior in connection with these expressions will resemble that observed in connection with genuine synonyms or genuine analytic equivalents. For this reason, the nonsynonymy or nonanalytic-equivalence of the expressions will be reliably revealed only by examining speech behavior relative to appropriate *intensional* contexts, specifically, modal and counterfactual contexts. But why rely on this speech behavior? What reason is there to think that ordinary speakers (scientific essentialists, for example) are using their words *truly* in these special contexts? Knowledge of the semantics of the language (e.g., that 'A' means that A and that 'A' is true in the language if and only if what 'A' means is true) is not enough to guarantee that the words will be applied truly in modal and counterfactual contexts. The reason for thinking that they will be is that their application is based on underlying judgments (e.g., the judgment that it was possible for Aristotle not to teach Alexander, etc.) that are intuitive. But why trust a person's intuitive judgments? We have now come full circle. The behavioral strategy merely masks an implicit reliance on the intuitions of speakers regarding relevant metaphysical possibilities.

The paradox of analysis compounds the problem of using purely behavioral tests for synonymy or analytical equivalence. Consider expressions whose synonymy or analytical equivalence is highly nontrivial. Relative to *extensional* contexts, the relevant linguistic behavior in connection with these expressions will resemble that observed in connection with nonsynonyms and nonanalytic-equivalents. The synonymy or analytical equivalence of the expressions can be uncovered only by examining usage in appropriate *intensional* contexts, specifically, the modal and counterfactual contexts that would arise in high quality philosophical dialectic. But why trust that ordinary speakers apply these expressions *truly* in these contexts? Answer: their linguistic behavior is based on their intuitive judgments about the hypothetical cases involved. But why trust their intuitions? We are back to our original question.
20. Alternatively, it is in part based on our *intensional* semantical knowledge that in our native language 'It was possible for Aristotle not to teach

Alexander' means that it was possible for Aristotle not to teach Alexander and our semantical knowledge that a sentence is true in a language if and only if what it means is true. These two bits of semantical knowledge yield the following derived semantical knowledge: in our native language 'It was possible for Aristotle not to teach Alexander' is true if and only if it is true that it was possible for Aristotle not to teach Alexander. But how do we go from knowledge of this biconditional to knowledge of its left hand side? Presumably by invoking our antecedent extralinguistic knowledge of the right hand side, that is, our knowledge that it is true that it was possible for Aristotle not to teach Alexander. But on what is this antecedent extralinguistic knowledge based? This is where we started. The empirical linguistic run-around adds nothing.

21. In this paragraph the definite descriptions 'H₂O' and 'the compound molecules of which consist of two hydrogen ions and one oxygen ion' are understood as having narrow scope. This is the only reading that would make scientific essentialism interesting; otherwise, the conclusion would amount simply to asserting the necessary identity of x to x . The scientifically interesting conclusion is one that makes a substantive assertion, and to get the substantive conclusion one must interpret the definite descriptions as having narrow scope.
22. The moderate classical empiricist analogue of this thesis is that, necessarily, most of a person's beliefs about the applicability of a sensible concept (i.e., a concept of a sensible quality) or a reflective concept (i.e., a concept of an "inner feeling" or of a conscious operation of mind) to actual cases of the person's current experiences are true.

To feel the intuitive pull of this theory, consider the following example. Suppose that a normal, attentive person has a certain concept that he dubs *chromaticity*. Let c be the concept of being a color other than black, white, or gray, and let c' be the concept of being a color other than black, white, gray, or red. Suppose that the person's concept *chromaticity* is identical either to c or to c' , but we do not know which. Then, intuitively, if the person were to apply his concept *chromaticity* to a vivid shade of red he is sensing, then his concept would be c , not c' . On the other hand, if he were to hold that it does not apply to this shade, *chromaticity* would be c' , not c . Now we have here a case (i.e., a vivid shade of red) that is encountered in sensation. The situation, nevertheless, is quite the same with hypothetical cases except that only some hypothetical cases can be encountered in sensation; most of them can be encountered only in thought. In either situation, however, the rationalist capacity theory of determinate concept possession (see below) would provide a simple, unified explanation why judgments about the applicability of concepts to these elementary cases must be mostly true. This, together with the moderate truth-based theory of evidence, yields a simple, unified explanation of the evidential status of intuitions and of one's noninferential beliefs about one's current conscious conditions.

23. For example, suppose that k is the concept of being a multigon, and

suppose that *c* is the concept of being a closed plane figure with more than four sides. Suppose that, upon considering the matter, the person were to make the noninferential judgment that possibly there is a circular multigon, an elliptical multigon, a spiral multigon, and so on. Then the person's concept *k* would not be identical to *c*, for it is in fact not possible that closed plane figures with more than four sides should be circular, elliptical, spiral, and so on. On the other hand, suppose that, upon considering the matter, the person were to make the noninferential judgments that possibly there is a multigon with five sides, that possibly there is a multigon with six sides, that possibly there is a multigon with seven sides, and so forth. Then the person's concept *k* could be *c*, for it is in fact possible that closed plane figures with more than four sides should have five sides, six sides, seven sides, and so forth.

24. These fundamental logical operations are characterized in my book *Quality and Concept*, Oxford, 1982.

A point of clarification is in order here. To determine that a person has the indicated cognitive capacity to sort hypothetical cases, we must already be able to determine that the person have determinate concepts, namely, the *other* concepts involved in the hypothetical cases and in the logical manipulations. The rationalist capacity theory should not therefore be viewed as a vestige of verificationism. However, determinate possession of a concept could nevertheless *consist in* having the indicated kind of cognitive capacity. At least this is a theoretical possibility in the logical theory of *Quality and Concept*. On this theory, what would be required in order for possessing a concept to consist in having the indicated cognitive capacity is that the relation of possessing a concept be a *connection* (that is, a logically and metaphysically basic relation).

25. A corollary of the rationalist capacity theory is that a person's judgments about his immediate experiences (sensory or introspective) will also be like this. Accordingly, they too will count as data independently of other considerations. This is just the traditional empiricist theory that our experiences count as data. Thus, we would also have a "transcendental deduction" of this traditional empiricist theory of a *posteriori* evidence given within the framework of the coherence theory of justification.
26. Besides holding that the rationalist capacity theory provides a simpler explanation than do naturalistic causal explanations, is there no other way in which coherentists might try to save scientific essentialism? There is. Namely, to strengthen the formal requirements on satisfactory comprehensive theoretical systems. This requirement is one that might be needed in order to insure that the coherence theory itself is counted as justified by its own standards. According to the coherence theory, what justifies the belief in the coherence theory itself? Given the above truth-affirming-explanations requirement, this belief should be thrown out unless the simplest explanation of its origin affirms its truth. But naturalistic causal explanations of the origin of this belief seem not to affirm its truth. So on the assumption that a naturalistic causal theory provides the simplest explanation, coherence theorists are obliged to abandon their belief in the coherence theory. To avoid this self-defeating

outcome, coherentists might have no choice but to strengthen the truth-affirming-explanations requirement as follows. (In a full development, special steps would have to be taken to avoid difficulties that arise in connection with Gödel's second incompleteness theorem and the Montague-Kaplan paradox of the knower.) Consider various explanations of the origin of your belief in a particular theory of justification T and of the origin of the specific intuitions you would ordinarily use in support of this theory T . The truth of the theory T and the T -supporting intuition plays no role in some of these explanations; in other explanations, the truth of the theory T and the T -supporting intuitions does play a role. We may call the latter explanations *T-affirming*. And we may call a comprehensive theoretical system *T-affirming* if it includes a T -affirming explanation. (For example, let T be coherentism. Then an explanation will be coherentism-affirming if the truth of coherentism and of your coherentism-supporting intuitions plays a role the explanation of the origin of your belief in coherentism and of the origin of your coherentism-supporting intuitions. And your comprehensive theory will be coherentism-affirming if it includes a coherentism-affirming explanation.) The revised requirement would be this: a person should drop all beliefs that are not affirmed by the person's simplest coherentism-affirming comprehensive theory. Now the question is what is your simplest coherentism-affirming comprehensive theory? Whatever its other features, it includes the simplest coherentism-affirming explanation of the origin of your belief in coherentism and of the origin of your coherentism-supporting intuitions. So which explanation is this? Naturalistic causal explanations are not coherentism-affirming, so they are out of the running. In view of this, your simplest coherentism-affirming explanation, and in turn your simplest coherentism-affirming comprehensive theory, would, I conjecture, be that based on the circumscribed rationalist theory I will sketch in the next section. In this case, however, you are once again back to a rationalist theory of *a priori* evidence. For you now may implement the above "transcendental deduction" of a rationalist theory within the coherentist framework. The upshot is that, even within coherentism, intuitions about the applicability of a concept to hypothetical cases necessarily will have a privileged status as data independently of the other details of your best theoretical systematization of your beliefs.

27. There is another way to bring out the problem with the familiar unrestricted versions of rationalism. Consider twin Earth. My doppelgänger on twin Earth has a concept w_{te} that corresponds to my water concept w_e . My doppelgänger's intuitions regarding the applicability of his concept w_{te} to hypothetical cases fit the same pattern as do my intuitions regarding the applicability of my concept w_e to hypothetical cases. That is, the pattern of my doppelgänger's intuitions mimics exactly the pattern of my intuitions. So there would be no difference in the sorting procedures that we would go through concerning the applicability of w_e and w_{te} , respectively, to hypothetical cases, and there would be no difference in the logical manipulations to which

we would submit w_e and w_{te} , respectively. So according to the unrestricted rationalist theories of determinate concept possession, my doppelgänger and I must have the same concept. However, according to the causal theory of determinate concept possession that is espoused by scientific essentialists, my doppelgänger and I have different concepts.

Strictly speaking, in this example we should consider not a single pair of our naturalistic concepts (e.g., w_e and w_{te}); we should consider all pairs of our naturalistic concepts simultaneously. The philosophical point, however, is the same: the pattern in my application and logical manipulations of my family of naturalistic concepts is indistinguishable from the pattern in my doppelgänger's application and logical manipulations of his corresponding family of naturalistic concepts. So, according to the unrestricted rationalist theories of determinate concept possession, his naturalistic concepts would be the same as mine. But this contradicts the causal theory of determinate concept possession advocated by scientific essentialists.

28. By the same token, the Kripkean intuition that it is metaphysically possible for something to be hot and yet not cause S ["sensations-of-heat"] is determined by of the following categorial intuition about internal non-relational physical qualities F and sensible qualia G: It is metaphysically possible for something to be F and yet not cause members of some species of sentient being to sense G.
29. Notice that, if natural rationalism is to be successful, we must be able to explain how we could know the categories of the naturalistic items of which we have a *posteriori* modal knowledge. How, for example, do we know that an item such as water is a compositional stuff and that an item such as perfume or food is instead a functional stuff? It turns out that the answer to this question is an easy corollary of our moderate form of causal theory of what is required for having a determinate concept of a naturalistic item. For reasons of space, the explanation will have to await the sequel to the present paper.
30. Kurt Gödel, "What Is Cantor's Continuum Problem?", in *Philosophy of Mathematics*, second edition, P. Benacerraf and H. Putnam, eds., Cambridge, 1983, 470-485 (revised and expanded version of a paper of the same title that appeared in *The American Mathematical Monthly*, vol. 54, 1947, 515-525).
31. Thus, as I mentioned earlier, the argument doubles as an argument against *radical* anti-individualism, which is the doctrine that *no* concepts (not even category and content concepts) are "in the head."
32. This notion of cognitive commitment was introduced by me in "Pragmatics" (section 39, *Quality and Concept*) to assist in the solution of puzzles posed by Mates, Burge, Kripke, Castaneda, and Perry.
33. The judgments the person would make here are examples of *convictions in acquaintance*. A person has cognitive commitments to all of his convictions in acquaintance, but not conversely. Typically, a person has a great many cognitive commitments that are not among his convictions in acquaintance. This distinction plays a pivotal role in solving the puzzles mentioned in the previous note. For example, Kripke's Pierre

has a cognitive commitment to the logical falsehood that x is pretty and x is not pretty (where London = x = Londres), but this proposition is not one of his convictions in acquaintance. The amnesiac has a cognitive commitment to the logical falsehood that $x \neq x$ (where x = the amnesiac), but this proposition is not one of his convictions in acquaintance. When I fail to recognize myself in the mirror, I have a cognitive commitment to the logical falsehood that $x \neq x$ (where x = me), but this is not one of my convictions in acquaintance. And so on. For more on this, see section 39, *Quality and Concept*.