Concepts and Conceptual Analysis

STEPHEN LAURENCE
University of Sheffield

ERIC MARGOLIS[1]
Rice University

Conceptual analysis is undergoing a revival in philosophy, and much of the credit goes to Frank Jackson. Jackson argues that conceptual analysis is needed as an integral component of so-called serious metaphysics and that it also does explanatory work in accounting for such phenomena as categorization, meaning change, communication, and linguistic understanding. He even goes so far as to argue that opponents of conceptual analysis are implicitly committed to it in practice. We show that he is wrong on all of these points and that his case for conceptual analysis doesn't succeed. At the same time, we argue that the sorts of intuitions that figure in conceptual analysis may still have a significant role to play in philosophy. So naturalists needn't disregard intuitions altogether.

1. Introduction

Conceptual analysis is currently undergoing a revival despite the enormous impact of naturalism in philosophy and a long history of proposed analyses being subjected to counterexample after counterexample. This renewed interest is due to a number of philosophers who have reinterpreted the role of conceptual analysis in philosophy, arguing that it is not only viable but necessary. These philosophers don't see themselves as merely restating old ideas. On the contrary, contemporary defenses of conceptual analysis are more sophisticated than earlier accounts, if only because they explicitly address some of the most pressing criticisms that led to conceptual analysis's original decline.

Recent advocates of conceptual analysis include George Bealer, David Chalmers, Frank Jackson, and David Lewis (among others).[2] Naturally, these different theorists develop their case for conceptual analysis in different ways,

---

[2] Bealer (1987, 1998), Chalmers (1993), Jackson (1994, 1998), Lewis (1994).

with varying ideas about its motivation and scope. For example, Bealer takes the primary interest of conceptual analysis to concern the study of specifically philosophical concepts. His overarching goal is to defend what he calls the authority and autonomy of philosophy—roughly, the idea that philosophy has a distinctive subject matter accessible in principle via a priori methods that trump the empirical methods of science. In contrast, philosophers like Jackson and Lewis take conceptual analysis to have far broader scope. Commenting on his views in the philosophy of mind, Lewis writes that "my reductionism about mind begins as part of an a priori reductionism about everything" (1994, p. 291). Similarly, Jackson takes conceptual analysis to be an integral component of a wide variety of intellectual activities. In particular, conceptual analysis is absolutely essential whenever the need arises to isolate the subject matter of a given theory or system of belief. If Lewis and Jackson are right, conceptual analysis isn't confined to such concepts as TRUTH and JUSTICE; it's relevant to just about any concept, including the likes of PROTON and MOLYBDENUM.[3]

Conceptual analysis has had a long and venerable history tracing back to the very origins of philosophy, but it took on a particularly heavy burden in the early 20th century when it was widely thought, following Carnap and others, that scientific concepts must be definable a priori and that it's philosophy's job to furnish the definitions. However, in the 1950s and 1960s, W.V.O. Quine and Hilary Putnam convinced many philosophers that this is a mistaken view. Quine and Putnam highlighted the limits of a priori inquiry, noting that science sometimes overturns even our most cherished beliefs. A famous example discussed in Putnam (1962) is the discovery that space is non-Euclidean, which breaks the apparent conceptual link between STRAIGHT LINE and THE SHORTEST DISTANCE BETWEEN TWO POINTS. Together with Saul Kripke (1972), Putnam also emphasized the fact that we can possess concepts in spite of being massively ignorant of, or mistaken about, the kinds our concepts pick out. For example, one can possess a concept like BUBONIC PLAGUE without knowing that the plague was spread by the transmission and proliferation of tiny micro-organisms and that the symptoms of the plague are causal effects of the activity of these micro-organisms in a host's body. One can even possess the concept in spite of wrongly taking the symptoms to be an infallible indicator of the disease's presence or while believing that the affliction is the result of some sort of supernatural causes.[4]

---

[3]   We'll use small caps to indicate concepts and quotation marks for mentioned words; however, we will often move back and forth between claims about concepts and words, since the difference will be largely irrelevant (see footnote 7).

[4]   Describing the historical ignorance and error surrounding the bubonic plague, Colin McEvedy writes (1988, p. 4):

> Strange as it may seem, in view of the frequency of the disease and the toll it exerted on the population, no one at the time had an inkling of its fundamental

What makes the contemporary revival of conceptual analysis particularly interesting is the fact that the new proponents of conceptual analysis explicitly address these sorts of worries urged by Quine, Putnam and others. Moreover, since a substantial portion of philosophy done in the wake of Quine's and Putnam's work has been premised on the rejection of analyticity and a priori methods, the work of contemporary conceptual analysts like Bealer and Jackson constitutes a radical critique of much recent philosophy, especially philosophy of mind.[5]

In this paper, we'll focus on Jackson's defense of conceptual analysis. We single it out not only because it is so ambitious but also because Jackson's treatment of the topic is perhaps the most explicit and detailed account available. In a series of important papers and particularly in his (1998) book, Jackson articulates a general framework for understanding conceptual analysis and presents several detailed case studies. If conceptual analysis comes to regain its position in philosophy, much of the credit will have to go to Jackson. Be this as it may, the burden of our argument in this paper is to show that Jackson's case does *not* succeed. We claim that Jackson's account remains susceptible to the sorts of worries raised by Quine and Putnam (though not in precisely the same way as earlier accounts) and that Jackson fails to undermine the familiar naturalistic and anti-a-prioristic views prevalent in contemporary philosophy of mind. Our response takes several stages, since Jackson himself relies on a number of distinct arguments. We begin, in section 2, by elucidating Jackson's primary argument, which assigns conceptual analysis a central role in what he calls "serious metaphysics". In section 3, we show that this argument is sufficiently flawed that Jackson will have to look elsewhere to find support for conceptual analysis. In fact, Jackson does look elsewhere, identifying a variety of supplementary arguments based on such phenomena as categorization, meaning change, communication, and understanding. In section 4 we present and criticize each of these arguments, showing that not one of them works. Jackson's radical critique of recent philosophy of mind is unsuccessful, as is his attempt to revive conceptual analysis. We conclude, in section 5, with a brief discussion of the role of intuitions in philosophy, given the foregoing critique of conceptual analysis. As we point out, philosophers who are opposed to conceptual analysis can't

---

nature, its ultimate cause or how it was spread. During the period of the Black Death people were inclined to attribute the disease to unfavorable astrological combinations or malignant atmospheres ("miasmas"), neither of which could be translated into a public-health program of any kind. More paranoid elaborations blamed the disease on deliberate contamination by witches, Moslems (an idea proposed by Christians), Christians (proposed by Moslems) and Jews (proposed by both groups).

[5]   We should note that while we adopt the broadly naturalistic framework that is at stake here, we are in no way committed to Quine's doctrines per se and especially not to his skepticism about semantics.

rely on intuitions in the way that Jackson encourages, but intuitions may still play an important role in their philosophy.

2. From Serious Metaphysics to Conceptual Analysis

Jackson's primary route to conceptual analysis is his argument that it is an essential element of so-called serious metaphysics. Jackson characterizes metaphysics in a familiar way, as the study of what sorts of things exist and what they are like, but as he sees it, it would be foolish of philosophers to be concerned with "any old shopping list of what there is and what it is like" (1998, p. 4). Serious metaphysics is a reflection of this sentiment. Its aim is a more comprehensive account of things in terms of a fundamental set of entities, one that restricts itself to "a limited number of ingredients" (1998, p. 5). For instance, physicalism is a doctrine in serious metaphysics, since it aims for a comprehensive account of the world in terms of just its physical structure.

Part of what goes into doing serious metaphysics is the conviction that higher-level phenomena can't be taken for granted. Jackson calls this the location problem. As he puts it, the location problem is the problem of showing how any given phenomenon can be reconciled with one's fundamental ontology. If it can't be, then one has to face up to being an eliminativist about the phenomenon.[6] There is, of course, a great deal of controversy surrounding the question of how lower-level phenomena are supposed to account for higher-level phenomena. Jackson's position on this issue is that, putting aside the eliminativist option, higher-level phenomena must be entailed by the basic phenomena of one's serious metaphysics. So if one were to have a completely fleshed out description of the world phrased in the proprietary basic terminology of one's serious metaphysics, it ought to entail any other true claims about the world. Likewise, to show that a particular claim about the world is acceptable, one has to show that it is entailed by the lower-level description.

It is at this point that conceptual analysis enters the picture. According to Jackson, conceptual analysis is required in order to show that a true claim about the world is entailed by some more fundamental description. He notes that the problem here is in some ways a special case of the more general problem of linking any two claims or theories that are framed in disparate vocabularies. To do this, you need a grip on the subject matter of the claims. Only when you know what the subject matter is will you be in a position to see how they relate to one another. "[M]etaphysicians will not get very far with questions like: Are there Ks? Are Ks nothing over and above Js? and, Is

---

[6]   This step in Jackson's argument is highly questionable—after all, one could always just expand the set of primitives. There is no a priori guarantee that the world is as orderly as some might like it to be. For related discussion see Stich and Laurence (1994).

the K way the world is fully determined by the J way the world is? in the absence of some conception of what counts as a K, and what counts as J" (1998, pp. 30-1). But as Jackson sees it, the only way to determine what counts as a K or a J is by appealing to intuitions about whether these terms apply in various possible cases, that is, by engaging in traditional conceptual analysis. Conceptual analysis reveals our implicit conception associated with a term and this, in turn, defines what the term is about.[7] The result is that conceptual analysis figures prominently in the practice of serious meta-physics. Without it, there'd be no way to evaluate some purported higher-level description of the world with respect to its most fundamental descrip-tion.

Though each step in Jackson's view raises questions, it is the commit-ment to conceptual analysis that really stands out. The business about one description entailing the other isn't nearly as objectionable, since, for Jack-son, entailment is merely a kind of necessitation. When he says that the physical description of the world entails all others, what this means is just that the physical way the world is somehow fixes or determines the other ways that the world is. We suspect that a lot of philosophers are prepared to accept some such claim. On the other hand, the suggestion that one can know these necessitations a priori is harder to swallow. It commits physicalists not only to the claim that all facts are necessitated by the physical facts, but also to the claim that all facts—chemical, biological, psychological, economic, political, and aesthetic—are *a priori deducible* from the physical facts. This is a startling claim, and one that physicalists are rightly wary of endorsing. Consider, for example, the claim that water is H_2O. We take it that many philosophers are prepared to grant that, given that water is H_2O, it is neces-sarily so. Yet on the face of it, this isn't the kind of claim that can be known a priori; on the contrary, it's a paradigmatically a posteriori claim.

Jackson is very much aware of the potential difficulties such cases pose for his view. Indeed, it is examples of just this sort that motivate one of the central components of Jackson's brand of conceptual analysis. According to Jackson, a posteriori necessities involve an a priori component and it is through this component that conceptual analysis continues to make its con-

---

[7]   Though Jackson often talks of concepts, his interest is really in natural language terms. "I use the word 'concept' partly in deference to the traditional terminology which talks of conceptual analysis, and partly to emphasize that though our subject is the elucidation of the various situations covered by bits of language according to one or another language user, or by the folk in general, it is divorced from considerations local to any particular language" (Jackson 1998, p. 33). This terminology is perfectly harmless, so we'll con-tinue to follow Jackson in moving back and forth between claims about concepts and claims about language. What matters to Jackson's defense of conceptual analysis isn't a view about what concepts are but the necessity of formulating analyses by testing one's intuitions against various possible cases.

tribution. In defending this claim, Jackson relies heavily upon what has come to be called 2-dimensional semantics. This takes some explaining.

Perhaps the best way to get oriented is to think about the semantics of indexicals and demonstrative such as "I", "you", "she", "here", "now", "this", "that", and so on. Statements involving these terms appear to be doubly dependent upon context. On the one hand, they are dependent on a contextual feature that is specific to the term in question (e.g., the reference of "I" depends upon who the speaker is, while the reference of "you" depends roughly on who the intended audience is). On the other hand, they are dependent on the possible circumstances under which they are evaluated. Consider the following sentence: "I am in the Rolling Stones". One sense in which the sentence is context-dependent is that its truth value depends on who is saying it. Were Mick Jagger to utter this sentence, it would be true; if either one of us were to utter it, it would be false. But there is another sense in which the sentence is context dependent. This has to do with whether it's evaluated with respect to the actual world or to some other possible world. As things actually are, the sentence is true when uttered by Jagger and false when uttered by either one of us, but things could have been different. Had things gone differently, perhaps one of us would have found our way into rock and roll fame.

Now we can define two types of semantic value that might be associated with sentences that contain indexicals and demonstratives. One allows for the evaluation of a sentence in different contexts, treating them as if they were actual. The other allows for the evaluation of a sentence in different contexts, taking the actual context (including who the speaker is) as fixed. The difference is easiest to see if we switch the example. Take the sentence "I am here". In one sense, this sentence has the interesting property that if a speaker utters it she is pretty much guaranteed to be saying something true. In another sense, however, the sentence says something that is clearly contingent—after all, the speaker might well have been somewhere else. Corresponding to these two readings are two different types of proposition, which Jackson calls the A-proposition and the C-proposition. Both types of proposition can be characterized in terms of the sets of worlds where they are true. The A-proposition corresponds to the reading where "I am here" is always true. This is the set of worlds, w, where the sentence is true, assuming that w gives the context of utterance. So in this case, it is the set of worlds where a speaker (whoever and wherever she may be) utters the sentence "I am here", and the speaker is in fact in that very location. The C-proposition, by contrast, is the set of worlds where the sentence is true given that the actual context is taken to be fixed (including who the speaker is and her location). Suppose, for example, that one of us (SL) utters the sentence while visiting the Louvre. Then the C-proposition is the set of worlds where SL is at the Louvre.

2-Dimensional semantics is essentially a generalization of this account to a wider class of terms including, importantly, "water" and other natural kind terms.[8] According to Jackson, there are two propositions associated with a sentence like "water is a valuable resource". The C-proposition fixes the referent of "water" relative to the stuff in the actual world that is *watery*. ("Watery" is the analogue of the general semantic features that pick out the referent of an indexical or demonstrative given a context of utterance. Jackson doesn't provide an official detailed specification of exactly what being watery amounts to, but he seems to have in mind the characteristic superficial properties of water: clear, liquid, drinkable, falls from the skies as rain, fills the rivers, lakes, oceans, etc. See, e.g., Jackson 1998, p. 38.) Given this, the C-proposition corresponds to the set of all worlds, w, where H_2O is a valuable resource. The A-proposition, on the other hand, corresponds to the set of all worlds, w, where the watery stuff in w (which may or may not be H_2O) is a valuable resource in w.

Jackson's 2-dimensional semantics plays a crucial role in supporting his overall claim that physicalism is committed to the a priori deducibility of all facts from the physical facts. Consider first the following simple argument, which is intended to show how to derive a priori a description of the water way that the world is from the physical way that the world is. The water fact that Jackson chooses is that water covers most of the earth (see 1998, p. 82):

(1) H_2O covers most of the Earth.   (premise)
(2) H_2O is the watery stuff of our acquaintance.   (empirical fact)
(3) Water is the watery stuff of our acquaintance.   (a priori)
_____

(4) Therefore, water covers most of the earth.

Notice how the argument works. It derives a fact about the distribution of water in the world from an a priori truth taken together with certain information about the relevant context. If Jackson is right, knowing that water is the watery stuff of our acquaintance is on a par with knowing that "I" refers to the speaker. Both can be known a priori. Then all you have to do is see who the speaker actually is or, analogously, what stuff is actually the watery stuff. "[W]e did not know that (1) entailed (4) until we learnt (2). But as soon as we learnt (2), we had the wherewithal to move a priori from (1) together with (2), to (4)" (1998, p. 82)[9]

We are now in a position to see how Jackson proposes to use the 2-dimensional framework to establish the more general claim that, in principle,

_____

8   Jackson seems to apply the account across the board, though for some terms (such as "square") he holds that their A-intension and C-intension will be the same.

9   For ease of exposition, we've renumbered Jackson's premises in this quotation and in the displayed argument.

one can derive a priori *all* facts from the physical facts. The inference in the simple argument just given relies on an a priori truth and information about the context, the context being just the way that things actually are. But if physicalism is true, then *there is a complete physical description of the way that things actually are*.[10] So physicalism ought to always have the resources to fill in the context. For Jackson, this means that all facts ought to be derivable a priori from the physical facts, given a 2-dimensional framework of sufficient breadth.[11]

3. 2-Dimensional Semantics and the Scope of A Priori Deduction

Jackson's primary argument for conceptual analysis—the argument from serious metaphysics—relies heavily on a 2-dimensional framework, and one can't help but wonder whether natural kinds really do have a 2-dimensional semantics. We aren't confident that they do. For present purposes, however, we'll put this question to the side.[12] Instead, we'll focus on Jackson's commitment to conceptual analysis within the 2-dimensional framework. We'll ask whether conceptual analysis delivers an a priori component for natural kind concepts, in particular, one that allows for an a priori derivation from the physical facts. We'll argue that the answer to this question is decidedly "no". The problem is (1) that people don't have a priori access to the A-intensions of natural kind concepts (sec. 3.1) and (2) even if they did, a physical specification of the context doesn't come close to doing what Jackson needs it to do (sec. 3.2). We discuss these points in order, using Jackson's stock example (the concept WATER) for illustration.

---

[10]   As Jackson says, "the actual context is givable in physical terms according to physicalism" (1998, p. 83).

[11]   Interestingly, Jackson claims to have a quicker and simpler argument for much the same conclusion (Jackson 1998, pp. 83-4). He starts by saying that physicalists have to agree that the simplest organisms—amoebae—are ones whose properties can be deduced a priori from their physical description. He then goes on to point out that, according to physicalists, human beings are different from amoebae only in their degree of complexity. The result is supposed to be that all of the facts about us (including, e.g., psychological facts) ought to be a priori deducible from the physical facts. The problem with this argument is much the same as with his main argument—physicalists needn't agree that the facts about amoebae *are* a priori deducible from the physical facts. Just because there is a microphysical description of an amoebae and its environment doesn't mean that we can deduce a priori claims about amoebae couched in terms of higher level descriptions.

[12]   One reason for suspicion is that whereas the semantic facts seem obvious for paradigmatic 2-dimensional items—indexicals and demonstratives—there is much residual disagreement with natural kind terms, where people's intuitions are disparate and malleable (see, for example, Crane 1991, Cummins 1998, Zemach 1976).

3.1. Epistemic Access to A-Intensions

Our first objection is that the A-intension of WATER can't be known a priori. A-intensions are effectively functions from worlds to referents. The A-intension of WATER is supposed to pick out the kind that it refers to in each world, w, assuming that w is actual. So to have an a priori grasp of this function, one has to be able to say what WATER refers to in each world w (considered as actual) on a priori grounds. Jackson often writes as if this were a relatively trivial problem, since he thinks that people can know a priori a suitable description that incorporates the concept's stereotype ("the watery stuff of our acquaintance"). Then the referent of the term in w is whatever uniquely satisfies the description. Unfortunately, Jackson's claim that the A-intension of WATER can be known a priori faces three problems.

(1) The first problem is that people can't have a priori access to a description that picks out the referent of WATER in each world, w, since people don't even have a priori access to a description that picks out the referent of WATER in the actual world. The source of the difficulty here is that nearly any element in a natural kind concept's stereotype is open to revision in light of empirical findings.[13] This point is emphasized by Putnam (1970) who argues that a concept's stereotype is invariably defeasible. His example involves the concept LEMON, whose stereotype presumably includes that lemons are yellow, tart, juicy, etc. Putnam notes two obstacles to working these characteristics into a definition of LEMON. One is that, for all anyone knows, they may be based on atypical or idiosyncratic samples. As he puts it, "[T]he normal members of the natural kind in question may not really be the ones we think are normal" (1970, p. 142). The other is that, for all anyone knows, the conditions of observation may affect the apparent characteristics of the kind so that these change over time. For example, the color of lemons could change from yellow to blue, given the introduction of some new gas in the atmosphere. All the same, "we would not say that lemons had ceased to exist" (1970, p. 142). These observations are extremely insightful. They make it highly implausible to suppose that there is an accessible stereotype which fixes the subject matter of such a concept a priori. For even if the observed lemons *are* typical and even if their stereotypical features *don't* change, these are themselves empirical facts.

Putnam's case applies with equal force to Jackson's example of the concept WATER. Recall that being watery is, roughly, a matter of being clear, liquid, drinkable, falling from the skies as rain, filling the rivers, lakes, oceans, etc. But *liquid* can't be part of the a priori specification of its A-

---

[13]   The point isn't so much that the stereotype is *revisable* but that it is revisable *in light of empirical findings*. We mention this to distance our claim from the more banal sense in which even paradigmatic a priori statements are revisable, given the potential that one has had a lapse of reasoning, made a mistaken calculation, etc.

intension, since this would rule out water in its gaseous and solid forms. Similarly, whether water is clear depends to a large extent on the amount and type of impurities in a given sample, and since what counts as an impurity is itself subject to empirical investigation, *clear* isn't a better candidate than *liquid*. Consider also the situation of someone who grows up with no exposure to lakes or oceans—maybe a nomad in the Sahara. If Jackson's analysis is right, such a person should be able to know a priori that water fills oceans and lakes or else she lacks the concept WATER altogether. We take it, however, that neither of these possibilities is at all plausible, so one has to cross off *fills the oceans and lakes* as well. In fact, we doubt that any of Jackson's purported "watery" characteristics are immune to revision. Jackson seems to be dimly aware of this objection. He qualifies his account by claiming that not all of the features that go into a stereotype need to be satisfied—just "enough" of them (1994, p. 171). Putting aside the fact that Jackson doesn't provide any indication of how much is enough—much less a *principled* account of how much is enough—it seems clear that this sort of move won't help, since the criticism that we are tracing back to Putnam is that it is consistent with the way that many of our concepts work that there may in fact be no limit to the potential revision of their stereotypes.[14]

Surprisingly, Jackson provides very little in the way of detailed specifications of the A-intensions of particular concepts—a curious fact in itself given the absolutely crucial roles he assigns to them in categorization, communication, understanding, science, and philosophy—but he does have some remarks that are suggestive when he turns to his two case studies, color and ethics. Consider what he says in connection with color concepts (1998, p. 87):

There is an important sense in which we know the live possibilities as far as colour is concerned. We know that objects have dispositions to look one or another colour, that they have dispositions to modify incident and transmitted light in ways that underlie their dispositions to look one or another colour, and that subjects have experiences as of things looking one or another colour. We also know that this list includes all the possibly relevant properties.

Interestingly, Jackson's list includes some patently empirical phenomena. It's hardly a priori that objects modify incident and transmitted light or that such modifications affect how they look. These are ideas that find their home in empirical inquiry, even if they are ones that are widely acknowledged among educated people. By claiming that they are part of an a priori analysis, Jack-

---

[14]   Cf. Block and Stalnaker (1999) on the concept of life and the proposal that it can be characterized a priori in terms of reproduction, locomotion, digestion, excretion, respiration, etc. After rejecting each of these characteristics one by one, they point out that all of them could fail collectively. "[N]othing in the concept of life rules out the possibility that there could be living beings that are immortal, and don't reproduce, that are tree-like (so don't locomote), get their energy by electromagnetic induction (so don't digest or excrete), and have no need for any substance in the air (so don't respire)" (Block & Stalnaker 1999, p.14)

son seems to be committed to the view that people who don't have these beliefs lack not only the concept COLOR but also particular color concepts. Yet why should we think this? Certainly, there are many people who aren't at all privy to even the most rudimentary scientific discoveries concerning color and light. Does Jackson really want to say that they don't have the concept RED? Likewise, does he really want to say that Aristotle didn't have the concept RED? And how can Jackson be so sure that his list is complete, especially given that it is based on empirical discoveries that emerged with developments in the theory of optics? Couldn't further developments reveal that our current understanding of optics is wrong in some significant way?

In short, the beliefs embodied by a concept's stereotype are often entirely on a par with the patently defeasible empirical beliefs in which it participates.[15] Again, what this means is that one can't know a priori a description that fixes the reference of a natural kind. As a result, one can't know a priori a natural kind term's A-intension.

(2) A second problem facing Jackson's claim that we can know A-intensions a priori is that we have no a priori guarantee that anything like the description theory is the correct theory of reference determination in the first place. If A-intensions are essentially functions from worlds (considered as actual) to extensions, then the particular A-intension associated with a concept is implicitly determined by the operative account of reference determination that fixes the referent of the concept in each world.[16] It seems that it is here that Jackson's stereotypes play their role. The problem is that while Jackson may take it for granted that a version of the description theory is the right account of reference determination, there are a number of competing theories—really, theoretical approaches—for explaining reference, as anyone who is familiar with the recent literature on mental content knows. In addition to description theories, there are causal-historical theories, descriptive-causal theories, nomic-causal theories, teleosemantic theories, and so on.[17] So it has to be asked which, if any, of these theories gives the correct account of reference determination. This is a complex issue and is currently a matter of lively debate. The point that bears emphasizing, however, is that the

---

[15] In principle, Jackson could try to square the empirical defeasibility of stereotypes with a commitment to conceptual analysis; he could argue that the potential defeasibility of a concept's stereotype is built into the concept and that our ability to recognize this is actually an example of conceptual analysis. Unfortunately, this rejoinder isn't effective. We explain why in detail in section 4.1.

[16] A-intensions are just semantic values associated with given sentences, but something must determine which semantic value is associated with which sentence.

[17] See, for example, Putnam (1970, 1975), Boyd (1988), Devitt (1981), Devitt & Sterelny (1987), Dretske (1981, 1988), Fodor (1987, 1990), Millikan (1984, 1993), Papineau (1993).

nature of reference determination is itself a broadly empirical question.[18] At the very least, this is because developments in science have a bearing on the evaluation of theories of reference. For instance, one has to take into account the potential for empirical theories to overturn even what may once have looked like the most secure of analytic inferences. As we noted earlier, Putnam and others have argued that even a connection as tight as that between STRAIGHT LINE and THE SHORTEST DISTANCE BETWEEN TWO POINTS may be broken given the right theoretical developments. Such facts bear on the evaluation of which theory of reference is ultimately correct. This suggests that the choice of a theory of reference is not a wholly a priori matter.

In addition, a theory of reference determination has to answer to other broadly empirical concerns that are tied to the various explanatory goals associated with the individuation of mental states in terms of their content. Among other things, theories of reference need to mesh with what's known about misrepresentation, with realistic assessments of the level of indeterminacy in thought, with facts about learning and development, and (more generally) with current models of cognitive architecture and psychological processing.[19] This point is just the application of a familiar Putnamian point to the problem of reference. If reference is a real causal explanatory kind, then its nature is not simply a matter of stipulation, and our understanding of it is potentially hostage to empirical and theoretical developments concerning any number of related issues. Since the correct theory of reference determination is open to such empirical influence, the A-intension of a natural kind concept can't be known a priori.

(3) Finally, suppose that Jackson is wrong and that stereotypes and descriptions generally do not play the sort of role that Jackson imagines they do in determining the semantic values of kind terms. Suppose that some variant of the causal theory of reference is right, and that what determines the extension of a kind term in a given world is a matter of causal links, not satisfaction of descriptions. In that case it seems clear that we will *not* have a priori access to the A-intensions which are determined by these causal links, since we presumably do not have a priori access to the relevant causal relata.

3.2. The Specification of Contexts

Jackson's claim that the water way the world is can be deduced a priori from a physical description of the world already faces serious difficulties. Without a priori access to the A-intension of WATER he doesn't have the crucial link

---

[18] For discussion of the relevance of this point in a related context, see Laurence, Margolis, & Dawson (1999).

[19] Of course, not all theories of the determination of reference do make full contact with these issues, but we see that more as a reflection of theorizing about reference being in the early stages than as a statement about what theories of reference ought to aim to achieve.

that he needs between the two descriptions. Further problems arise when we turn to the role of context in his argument.

To see this, let's take another look at Jackson's illustration of how to derive a priori a water fact from a physical description of the world. We repeat the argument for convenience:

(1) H_2O covers most of the Earth.	(premise)
(2) H_2O is the watery stuff of our acquaintance.	(empirical fact)
(3) Water is the watery stuff of our acquaintance.	(a priori)

---

(4) Therefore, water covers most of the earth.

Premise (1) is supposed to be a partial physical description of the world, available on the assumption that physicalism is correct. Premise (2) supplements this with further information about the context, again, in physical terms (also available on the assumption that physicalism is correct). Finally, (3) presents information available a priori from the A-intension of WATER. As Jackson sees it, the inference is a priori, since the A-intension is itself a priori and only requires a context to be specified in order to support the conclusion about water's distribution.

Unfortunately, Jackson's formulation of this argument is very misleading. Notice that Jackson describes the argument as if it vindicates the a priori deducibility of water facts from *physical* facts. That is, the context that is specified in the first two premises is supposed to be in wholly physical terms. Moreover, what's meant by "physical terms" here isn't some loose interpretation that incorporates any purportedly natural phenomena. That would be question-begging, since Jackson is using physicalism as a case study in the workings of serious metaphysics, and serious metaphysics doesn't allow for the assumption that any phenomenon is admissible unless it is deducible a priori from one's most basic ontology. We take it, then, that Jackson himself is limited to the ontology of microphysics—particles, fields of forces, etc. But microphysics makes no mention whatsoever of H_2O, the Earth, and the host of other things that Jackson finds himself referring to in his first two premises.

Of course, there are moves which Jackson might make in order to address this problem. Both premise (1) and (2) are supposed to give partial physical descriptions of the world. Why not just use a complete description and bypass the questionable references altogether? After all, if physicalism is correct, there is a complete description of the world in microphysical terms. It's too long and unwieldy to be of much practical use, but the present concern is with what can be inferred a priori in principle, not in practice. So let's just replace the first two premises with a sentence that is stipulated to be a complete description of the world in microphysical terms, $S_{physical}$:

(1) $S_{physical}$
(2) Water is the watery stuff of our acquaintance.

---

(3) Therefore, water covers most of the earth.

But now the real problem with Jackson's argument should be apparent. How is one supposed to connect a truly microphysical description of the world with the knowledge that water is the watery stuff of our acquaintance? A microphysical description doesn't talk about things being watery. It has no use for the terms "liquid", "clear", "falls", "lake", "ocean", "acquaintance", etc. So someone could grasp the microphysical description and even know the A-intension of WATER and still not be able to assemble this information in a way that allows her to infer on a priori grounds that water covers most of the earth.

What Jackson needs is an a priori link between microphysical terms and the ordinary terms that express a natural kind's stereotype. We can envision two strategies he might try. The first is to claim that the troublesome terms can be given a priori definitions directly in microphysical terms. The problem with this strategy is simply that it is wildly implausible. There is no reason at all to think that any of these terms can be defined in this way. The second strategy is to hold out for a priori functional definitions and then assign each functional role a huge (perhaps open-ended) disjunctive set of microphysical conditions that instantiate it. Unfortunately, this strategy isn't much better since there continues to be a gap between the two descriptions. Knowing the microphysical description doesn't guarantee access to the higher-level functional description. Why couldn't you have a full grasp of the one and still be in no position to infer that the other applies as well? To be sure, the microphysical facts might necessitate that the functional role is fulfilled, but that's beside the point. For Jackson's argument to be vindicated, the necessitation would have to be knowable a priori. We see no reason to think that it is.

Another way of putting the point is that Jackson is overlooking that there are two vastly different modes of presentation at play, one that is couched in microphysical terms and one that is couched in terms that are common to natural kind stereotypes. The first is involved in the characterization of the context as it's given by the dictates of physicalism. The second is involved in the characterization of various possible contexts, as they are embodied in the A-intension of WATER. Jackson's difficulty is that there is no way of moving between these two modes of presentation on purely a priori grounds. What this means is that knowing the A-intension of WATER and knowing the context in a way that accords with physicalism isn't enough to establish an a priori deduction of the water facts. Since nothing in this objection turns on specific features of the concept WATER, the implication is that Jackson's entire program vis a vis serious metaphysics cannot get off the ground. He is

simply wrong to expect an a priori deduction of higher-level facts from a complete physical description of the world, no matter how detailed.

4. Jackson's Supplementary Arguments

We've just seen that Jackson's main argument for conceptual analysis is deeply flawed, as is his case for the a priori deducibility of higher-level facts from the microphysical facts. But Jackson has a variety of supplementary arguments to fall back on, at least in his defense of conceptual analysis. By our count, he has at least five. In this section, we'll argue that these don't work either.

4.1. Categorization and Conceptual Analysis

Jackson's first supplementary argument concerns the phenomenon of categorization. Categorization involves a psychological process in which an object, event, etc. is judged to fall under a concept or term. Jackson emphasizes that categorization isn't random or miraculous. From this he concludes that there must be some set of implicit[20] principles responsible for the judgements that we make in categorization, where these principles give the meaning of the concept in question. "[T]ypically we know something useful and non-grue-like, and are giving voice to this knowledge when we classify happenings as examples of grooming behavior, pain, rational inference, and so on" (1998, pp. 64-5). Conceptual analysis enters the picture as a way of articulating these principles. The very principles that guide our judgement about whether something is rational, for example, are the ones that are delivered by an a priori analysis of our concept of rationality.

The main problem with this argument is that categorization doesn't require analytic or a priori principles in order to operate. For example, it's important to the categorization of birds that they fly, lay eggs, produce songs, etc., but all the same it isn't analytic or a priori that birds have these features. Similarly, people often form judgements about other people's sex on the basis of hairstyle, clothing, etc. Yet it's certainly not analytic or a priori that, say, women have longer hair.

Now Jackson could respond by noting that categorization isn't a single psychological process. Perhaps it makes sense to distinguish between quick categorization (which is often based on minimal perceptual contact) and more considered categorization (which can appeal to any number of evidential sources). Then Jackson could claim that conceptual analysis is really only relevant to specifying the principles that guide considered categorization. Unfortunately, this distinction doesn't help him very much. That's because

---

[20] Jackson does not say what it is for an analysis to be implicit in one's categorization practice. We assume he means that the analysis is either represented subdoxastically or that it is entailed by things that are explicitly represented doxastically or subdoxastically.

either sort of categorization may be based on principles that are mistaken or otherwise flawed. This is the moral of the work by Quine and Putnam that we discussed above. As Quine and Putnam emphasized, even our most central beliefs—those that would enter into our most considered acts of categorization—are revisable in principle in light of unexpected empirical and theoretical developments. Moreover, if Quine is right about the holistic nature of confirmation, then our most considered acts of categorization would be ones that implicate large chunks of our total system of belief. Certainly, the a priori analysis of a single concept couldn't uncover principles that spill over like this.

We suspect that, at this point, Jackson might respond by trying to run his argument at a higher level. That is, he could concede that Quine and Putnam establish that certain specific proposals in conceptual analysis are wrong but then claim that the very fact that Quine and Putnam can show this justifies conceptual analysis in the long run. After all, people's reaction to Quine's and Putnam's considerations aren't random or miraculous. So there must be some implicit principles guiding their reaction; it's *these* principles that are to be made explicit by doing conceptual analysis.

The first thing to note about this proposal is that Jackson's claim would no longer be the claim that A-intensions are analytic and a priori. The claim would rather be that some set of principles more general than the A-intensions associated with particular concepts are analytic and a priori—perhaps, for example, the principles governing our understanding of the nature of substances in general.[21] The hope is that conceptual analysis might then have a role to play in articulating these principles. This isn't much more than a hope, however, since Quine and Putnam's critique of the a priori isn't confined to any particular level in the conceptual system—certainly not the level of concepts like WATER and LEMON. Their point is that empirical developments can impinge upon all sorts of claims or inferences that might initially seem unassailable. We don't see any reason why this doesn't include the principles governing a concept like SUBSTANCE or other similar higher-level concepts.

What's more, it's important to recognize that even if there is at some distant point an unrevisable set of principles guiding people's categorization, this would do little to vindicate the vital role that Jackson wants to assign to conceptual analysis. As he repeatedly emphasizes, he thinks that conceptual analysis is a pervasive feature of philosophy and science and that it is absolutely essential for coordinating different theories (broadly speaking), since it's only through conceptual analysis that one can isolate a theory's subject matter. But under the present proposal, the significance of conceptual analysis

---

[21] We don't know how abstract such principles may be. Our example (that they may concern such things as the nature of substances in general) is only a suggestion.

is drastically minimized. For any given natural kind, there's almost nothing you can know about it a priori. At best, you can hold out for the possibility that there is some extremely general principle, or set of principles, that guides people's judgments about the kind in response to novel empirical information. However, saying what these principles are would require knowing how human categorization responds to any potential scientific development. And since we can't know in advance how science will go, we can't know in advance what these underlying principles are. This puts Jackson in the awkward position of maintaining that conceptual analysis is actually hostage to empirical inquiry *in the long run*. Not only can it tell us nothing at the outset of investigation, but it may have to wait until all of science is complete before it can have anything to say at all.[22]

4.2. Causal Theories of Reference and Conceptual Analysis

Jackson's next supplementary argument is meant to turn the tables on opponents of conceptual analysis. Jackson is aware that many people think that there is something fundamentally wrong with a general commitment to conceptual analysis given the widespread occurrence of a posteriori necessities. Jackson may address some of this worry by appealing to a 2-dimensional semantics, but one may still wonder whether he is simply missing an important insight of Putnam, Kripke, and others, who have helped to motivate the need for a causal theory of reference.

---

[22]   One response that is available to Jackson—and one that he seems to have sympathy with—is to claim that the dependence on empirical information needn't compromise a prioricity because any empirical claim can always be treated conditionally. The idea is that what's important is only what we would say if a world of such-and-such sort were actual, not whether the actual world really is such-and-such. This "conditionalizing strategy" seems to allow for all scientific theorizing to be a priori, amazingly enough. Science is a matter of making theoretical moves in response to possible sets of evidence. The suggestion is that, from the point of view of constructing and evaluating theories, it doesn't matter which set of evidence is actual. All of science can be done in the absence of that information.

   There are a number of difficulties worth noting here. First, and most importantly, the conditionalizing strategy is question-begging in the present context, since it presupposes that we are already able to assign meanings to terms in order to represent how things are in each possible world. However, it is not at all obvious that we would be able to do this, or if we could, that we could determine what was in the extensions of these terms a priori. Consider, for example, the following possible world (which may or may not be actual). There are things that appear to be fruits, that are smallish, yellow, tart, etc. *Are they lemons?* Note that there will be many worlds like this. In some, the small, yellow, tart fruits are a small atypical subset of a larger group of fruits of the same kind, in others they are perfectly typical. Which are lemons? It may depend on how things are here in the actual world. Moreover, the same point may well apply to "fruit", "yellow", "tart" and so on. Another problem facing Jackson's conditionalizing strategy is that the nature of justification (especially the justification of non-demonstrative principles) may also depend on experience; which principles work well may very well be a function of the way the world actually is, and which worlds are local.

Jackson's response to this concern is to claim that the Kripke/Putnam tradition is itself implicitly committed to conceptual analysis. Far from showing that conceptual analysis is misguided, this tradition is actually supposed to exemplify the benefits of conceptual analysis. This is because, as Jackson sees it, causal theories of reference are motivated by thought experiments, in particular, Twin Earth style thought experiments (Jackson 1998, pp. 38-9):

> Putnam built his impressive case concerning the reference of theoretical terms out of intuitions about how to describe possible cases. He told stories about, for famous example, Twin Earth, and invited us to agree with him that what counted as water on Twin Earth was not the stuff on Twin Earth with the famous superficial properties of water—being a clear potable liquid and all that; for short, being watery—but rather the stuff that on Earth made up (most of) the watery samples that we were acquainted with when the term 'water' was introduced. We agreed with Putnam. But we were not under external instruction from some linguistic dictator to agree with him. Our agreement was endogenous. It, therefore, reflected our folk theory of water. Putnam's theory is built precisely on folk intuitions.

In other words, given the role of intuitions in establishing causal theories, it only makes sense to view them as exonerating conceptual analysis, whatever their proponents may say. Since intuitions about possible cases just are the stock and trade of conceptual analysis, the upshot is supposed to be that the causal-historical theory with its reliance on Twin Earth intuitions is best viewed as a case study in conceptual analysis, rather than as a prima facie counterexample.[23]

Jackson's focus on the causal-historical theory is bit misleading, since the original sentiment that he wants to address isn't tied to the causal-historical theory per se. Really, it turns on a variety of theoretical approaches to reference that stand in opposition to Jackson's more traditional description theory; among others, this includes nomic-causal theories and teleosemantic theories. Taking this broader set of theories into account, it becomes clear that Jackson's claim that "causal theories" are solely or primarily based on the Twin Earth thought experiment is simply wrong. The different theories in this broad family have a variety of motivations that have little, if anything, to do with Twin Earth. For example, one of the main motivations for teleosemantic theories is that its proponents want to account for misrepresentation and feel that the only sensible way to do this is by holding that representations have functions, thereby enabling them to malfunction.[24] On the other hand,

---

23 See also Jackson (1994) pp. 164-5.

24 E.g., Ruth Millikan writes that "[beliefs and intentions] display the characteristic mark that all things defined by proper-function categories display. It makes sense to speak of their being *defective*. Nothing that can be said to be defective is *what* it is merely by virtue of what it is actually like or what it actually does or would do if. What is defective is, just, that which is *not* what it *should* be or can*not* do what it *should* do, hence is something defined by its 'shoulds' and 'coulds' and 'woulds.' And I can see no way of unpacking the 'should' in this sort of context by reference to present structure or disposition. How could any 'should's or 'supposed to's be applied to the inner arrangements of the newly

one of the main motivations for certain nomic-causal theories is the feeling that teleosemantic theories can't account for misrepresentation short of leaving the mind far more indeterminate than it really is.[25] We don't propose to take sides on these issues here but do want to emphasize that the controversies they raise are at the very center of a rather lively debate concerning the nature of mental representation. Jackson's excessive interest in Twin Earth has blinded him to such matters.[26]

The debate over misrepresentation is indicative of the way that explanatory concerns, rather than bare intuitions and thought experiments, drive a good part of the theory of representation. We'd also suggest that it's the explanatory concerns that should take precedence. Despite the enormous amount of attention that Twin Earth has drawn, it offers a fairly weak argument for causal theories, even for believers in the philosophical significance of intuitions. At the very least, this is because the standard intuition associated with Twin Earth isn't widely shared. People often have exactly the opposite intuition that they are supposed to have. Who's to say that they have "the wrong intuition"?

Ironically, Putnam's main contribution to the development of causal theories isn't Twin Earth but rather his observations concerning purported a priori analyses and how they don't mesh with either the history of science or with our ordinary experience. As we have already seen, he emphasizes the way that we are able to coherently overturn deeply held beliefs (ones that otherwise might be identified as embodying analytic a priori truths) and that purported a priori analyses are often subject to empirical correction. To take another example, nothing seems more inviolable than the belief that if *a* is older than *b* today, then *a* will be older than *b* next year too; but Relativity Theory implies that this isn't so. Only with hindsight are we able to see that what prevented people from acknowledging the revisable character of the claim was the absence of an appropriate empirical theory. Putnam takes examples like

---

arrived, randomly created double? How could anything in him be defective? Beliefs, on the other hand, are *essentially* things that can be true or false, correct or defective" (1984, p. 94).

25 E.g., Jerry Fodor writes, "It might be argued that there is a real indeterminacy about whether frogs snap at flies or at little black dots. But, surely, if there are any matters of fact about content, it's one of them that frogs don't snap at flies under the description *fly or bee-bee*. Yet, as far as I can see, it's equally OK with Darwin which way you describe the intentional objects of fly snaps, so long as it's reliable … that all the local flies-or-bee-bees are flies… So evolutionary teleology *cannot tell these frogs apart*" (1990a, p. 73; see also his 1990b).

26 But even if there are these other motivations and justifications for theories of content, couldn't the causal theory be motivated and defended solely on the grounds of Twin-Earth-style thought experiments? And if so, doesn't this mean that the theory of content *can* be justified on a priori grounds? As we note in the text, we think Twin Earth only provides fairly weak support for causal theories of content. The question of whether such justification would be a priori is more interesting (see section 5 below).

this as a sign that we have to be cautious. There's always the possibility that an apparent a priori analysis is really just a reflection of a lack of theoretical imagination. So an apparently unrevisable belief can't be taken at face value. The implication for the theory of reference is that it has to allow for such unexpected developments. One of the advantages of causal/anti-descriptivist theories is that they do.

Another way of capturing the appeal of anti-descriptivist theories is to put the matter in terms of how much ignorance and error is consistent with the possession of a concept. Theorists who are opposed to descriptivism tend to emphasize that the potential for ignorance and error is considerable. Take, for example, a concept like MEASLES. People's beliefs about measles vary. Some know only about the spots associated with measles, while others know that it is caused by a virus (but couldn't say which one), while still others may mistakenly think that it is a bacterial condition or perhaps even caused by spiritual forces, such as God's will. Despite these differences, all of these people can still possess the very same concept. This is what allows them to disagree about the nature of measles and to change their minds as they come to learn more about the disease. Perhaps the most attractive feature of causal theories of reference is that they readily accommodate all of this variation.[27] Twin Earth needn't come into it.

4.3. Meaning Change and Conceptual Analysis

Jackson's third supplementary argument is based on the fact that it is at least *possible* to change the meanings of one's concepts. Make enough changes to your network of beliefs and eventually the embedded concepts will come to embody a different subject matter. Of course, not any change in belief will lead to a meaning change, but the point that Jackson wants to emphasize is that meaning change can, and does, occur. This being the case, we need an explanation of how meaning change is possible. As Jackson sees it, the explanation can only be that certain beliefs are essential to a concept and that a change in one or more of these amounts to conceptual change. Conceptual analysis comes into the picture as the means of identifying the essential beliefs. Since there must be essential beliefs, conceptual analysis must be possible (1998, p. 38):

---

[27] Diehard advocates of conceptual analysis could respond to the measles case by digging in their heels and maintaining that natural kind concepts are subject to a priori analyses, despite the apparent significance of ignorance and error. What we've been describing as cases of people disagreeing or changing their mind *about a disease* could be described, instead, as disagreements *about which of several competing disease concepts* is most useful given the way the world is. The main problem with this strategy, however, is precisely that it doesn't allow for substantive disagreements about the world. And yet, it does seem as if we disagree with one another, and with our former selves, not about the utility of certain concepts, but about how things are. Giving up on this conviction may offer a consistent position, but it's not a particularly attractive one.

The role of intuitions about possible cases so distinctive of conceptual analysis is precisely to make explicit our implicit folk theory and, in particular, to make explicit which properties are really central… . For surely it *is* possible to change the subject, and how else could one do it other than by abandoning what is most central to defining one's subject? Would a better way of changing the subject be to abandon what is *less* central?

This argument makes an elementary mistake, so we will be brief. The mistake is to assume that there is only one model of meaning change. But in fact, nearly any theory of content brings with it the possibility of meaning change. There's nothing special about Jackson's model at all. It's just the model that comes readily to mind given a description-theoretic account of content like Jackson's. However, given a different theory of content, meaning change would have a different explanation. For example, consider how things work with nomic-causal theories. On these accounts, content is constituted by the existence of a nomic dependency between a mental representation and the property it expresses. Then meaning change occurs when the representation comes to stand in a different dependency relation; in contrast, content is preserved despite any change in belief so long as the same dependency relation is preserved. Meaning change is a real and fascinating phenomenon, but since alternative theories of content that have no truck with conceptual analysis are equally compatible with this phenomenon, it does nothing to support conceptual analysis.

4.4. Communication and Conceptual Analysis

Jackson's fourth supplementary argument is closely related to his concerns about meaning change and categorization. Jackson argues that our ability to communicate useful information depends on the fact that the words we use, and the concepts they encode, have analytic entailments. The whole point of someone's describing a situation using certain words is that these words convey something worth knowing about the situation. For Jackson, this seems to imply that our concepts have information built into them in the form of analytic entailments, ones that can only be elucidated by doing conceptual analysis. "For only then can we explain the manifestly useful information we give about what the world is like to each other and to our later selves, through diary entries and notes on fridges…" (1998, p. 65).

We find this argument especially puzzling. What exactly is the link between being able to communicate something using a word and its having such-and-such analytic entailments? The only model that suggests itself is that in grasping the meaning of the word, one inevitably grasps its analytic entailments and that these, in turn, provide the information that makes using the word so useful. The problem is that much of the information we communicate isn't, in fact, analytic; indeed, much of what we communicate is obviously *non*-analytic. London is in England. It's not analytic that London is in

England, but everyone knows that it is. So if you are told that Bill will be in London this summer, you'll come to be in possession of the useful information that Bill will be in England this summer. Likewise, suppose you are told that the stove is hot. It doesn't follow analytically that it may burn your hand, but everyone knows that hot things burn, particularly hot stoves. So you come to possess the useful information that you shouldn't touch the stove. What these examples show is that words convey the useful information that is associated with them *whether or not that information derives from their meaning*. Moreover, even assuming that there are analyticities for conceptual analysis to uncover, it's unlikely that grasping these could be needed for communication anyway. After all, conceptual analysis isn't supposed to be easy—witness the huge post-Gettier literature on "knows". So it's unlikely that people are invariably in a position to grasp a word's analytic entailments just by grasping its meaning. For these reasons, the useful exchange of information in communication can't depend on the possibility of conceptual analysis.

4.5. *Understanding and Conceptual Analysis*

Jackson's final supplementary argument is more elusive than the others, yet, in the end, it may be the most illuminating. The backdrop to the argument is a general account of understanding a sentence. Jackson notes that a proposition effects a partition in the set of possible worlds; it divides the space of possible worlds into those in which it is true and those in which it is false. Since understanding a sentence involves grasping the proposition that it expresses and knowing its truth conditions, Jackson concludes that understanding a sentence gives one access to this partitioning. "[Our folk theory] ties together understanding, truth, and information about possibilities; and the obvious way to articulate this folk theory is to identify, or at least essentially connect, understanding a sentence with knowing the conditions under which it is true; that is, knowing the possible worlds in which it is true and the possible worlds in which it is false" (1998, p. 71). A problem arises, however, with a posteriori necessities, such as "water is H_2O". If the foregoing account is right, how could someone understand this sentence without seeing that it is true in every possible world? But if someone could see that, then there wouldn't be any doubt about its necessity; merely understanding the sentence should generate an appreciation of its modal status. In other words, on what Jackson takes to be an uncontroversial model of understanding a sentence, it's a serious puzzle how there could be any a posteriori necessities. As Jackson remarks (1998, p. 72):

[I]t seems that understanding a necessarily true sentence should, at least in principle, be enough to reveal its necessary status. For understanding it would require knowing the condi-

tions under which it is true, and how could you know them—really know them—and yet fail to notice that they hold universally?

Jackson's solution to this problem is to say that understanding a sentence only requires grasping its A-proposition—grasping its C-proposition isn't a requirement. Notice how this helps. The A-proposition of "water is H_2O" is the set of worlds, w, where the watery stuff in w is H_2O. This doesn't include every possible world, since there are many worlds where the local watery stuff has some other chemical composition, for example, on Twin Earth (as it is standardly characterized) the watery stuff is XYZ. This being the case, a grasp of the A-proposition doesn't by itself reveal that "water is H_2O" is necessary. The connection with conceptual analysis, of course, is that conceptual analysis delivers the A-intensions of a sentence's component words. So any account of understanding must rely heavily on conceptual analysis.

The first thing to say about this argument is that Jackson writes as if his model of understanding were inevitable. The best way to see that it isn't is to work through a simple alternative, one that is not committed to conceptual analysis. For these purposes, let's suppose that there is a language of thought and that understanding a sentence involves the occurrence of something like a translation between the two systems of representation—the natural language and the language of thought. To a first approximation, understanding a sentence on this model is a matter of entertaining the corresponding internal representation, i.e., one that has the same content as the target sentence.[28] In particular, to understand the sentence "water is H_2O" is to entertain an internal representation with the content *that water is H_2O*. Jackson's problem about a posteriori necessities never arises. (The problem is probably spurious anyway, since presumably one can grasp a sentence that expresses even an *a priori* necessity without appreciating that it is necessary. This certainly seems to be the case for some sentences concerning logical or mathematical truths. People can clearly understand a sentence expressing a mathematical proposition such as Goldbach's conjecture—or it's negation—despite being unable to determine whether it is true or false).

Not only are there alternatives to Jackson's model of understanding, but it's doubtful that his model is actually defensible. He claims that understanding a sentence requires knowing the worlds in which it is true. But what exactly does this mean? One possibility is that understanding a sentence involves explicitly representing each of these worlds. But that won't work, since there will almost always be an infinite number of them. The alternative is that understanding a sentence involves being able to tell, given a world, whether or not the sentence is true in that world. But this won't work either,

---

[28]   For purposes of exposition, we are overlooking indexicals, demonstratives, and other complications. We should also note that the internal representation would have to have a functional role appropriate to understanding.

since our ability to tell whether a given sentence is true in a world depends on how that world is presented. It's not as if we can tell that a sentence is true in a world regardless of the mode of presentation under which the world is considered; even the truth of a very simple sentence may go unappreciated given the wrong mode of presentation.

Interestingly, the problems Jackson faces here may shed new light on Jackson's most famous argument, viz., the so called Knowledge Argument against physicalism (Jackson 1982). The Knowledge Argument is based on a thought experiment. We are to imagine that there is a brilliant future scientist, Mary, who spends her entire life in a colorless environment. Despite this, she becomes the world's leading expert on the neurophysiology of vision, learning all there is to know about the neural mechanisms for color perception. By hypothesis, she knows all the relevant physical facts (broadly construed to include chemical, biological, and functional facts). Suppose she now leaves her room and has her first color experience; she sees red for the very first time. Jackson claims—or at least he used to claim—that when this happens she comes to learn a new fact. But since she already knew all of the physical facts, this new fact must be non-physical. So physicalism is false.

Now one common reaction to the Knowledge Argument is to note that even if Mary does learn something new, what she learns may just be an old fact under a new mode of presentation; prior to leaving her room, Mary knew all of the physical facts but only qua physical facts. Like many others, we take the mode-of-presentation objection to be a powerful objection to the Knowledge Argument. Surprisingly, however, Jackson has never acknowledged its force (see Jackson 1986). Why not? Our suggestion is that his resistance stems from the same considerations that are at play in his model of understanding a sentence. Take two descriptions of Mary's experience of red, one under a physical mode of presentation and the other under an experiential mode of presentation. If understanding either one of these entails being able to tell for any world whether it is true in that world, then it simply doesn't matter how things are couched. You ought to be able to see that the two descriptions are true in exactly the same worlds. So if physicalism were true, how could Mary, who knows all of the physical facts, come to learn something new? She couldn't. In this way Jackson's unfortunate views about understanding preclude him from appreciating the full impact of the mode of presentation objection.

Another way to see the point is to recast the Knowledge Argument in terms of Jackson's current 2-dimesional framework. As Jackson might put the argument these days, if physicalism is true and sensations are physically grounded, Mary should be able to deduce the C-proposition for any sensation statement. This is because her knowledge of the physical facts gives her all she needs to know about the context. Thus, she shouldn't learn anything

really new on seeing red for the first time. But as we argued earlier (see section 3.2), it's not enough to know the context. You have to know it under the right mode of presentation. In particular, a presentation of the context in solely physical terms doesn't allow for the a priori deduction of higher-level facts, even if the physical facts necessitate the higher-level facts. Jackson's failure to fully appreciate the significance of different modes of presentation would appear to be an important and constant theme in his writing.

5. Conclusion: The Role of Intuition in Philosophy

Jackson claims that conceptual analysis plays an essential role in all rational inquiry. As he sees it, conceptual analysis is required to account for categorization, communication, meaning change, and linguistic understanding, as well as to fix the subject matter of our concepts. Jackson even goes so far as to claim that conceptual analysis is presupposed by philosophers who strive to avoid it by adopting causal theories of reference. There's little doubt that Jackson's sweeping program, if correct, would call into question a vast amount of work in contemporary philosophy, work that is built around the anti-a-prioristic conclusions of Quine, Putnam, and others. We've shown, however, that Jackson's radical critique of contemporary philosophy does not succeed and that conceptual analysis does not play the central role that Jackson claims for it. Conceptual analysis is not required to fix the subject matter of our concepts or to account for categorization, communication, meaning change, or linguistic understanding. Nor do causal theories and other philosophical theses adopted by philosophers who are opposed to conceptual analysis depend on conceptual analysis in the way Jackson alleges.

The distinctive features of conceptual analysis is a reliance on people's intuitions about the application of concepts in actual and hypothetical circumstances. Many philosophers opposed to conceptual analysis come down very hard on intuitions. For example, Robert Cummins writes that "philosophical intuition is epistemologically useless" (1998, p. 125). We suspect that this sentiment expresses a widely held view in contemporary philosophy, especially the philosophy of mind; these days, one often hears the claim that if a view conflicts with intuition then so much the worse for intuition.[29]

---

[29]   See, for example, the discussion of Davidson's Swampman example in a special forum in *Mind & Language*, volume 11.1 (March 1996). Swampman is a physical duplicate of a normal human being but happens to be created by a freak accident. Intuitively, it seems that Swampman has beliefs and desires just as his duplicate does, but this conflicts with various theoretical accounts of the nature of such states, accounts that take an organism's history to be crucial to whether it can have contentful mental states. Daniel Dennett says that the example is "not worth discussing" (1996, p. 77). David Papineau, summarizing his own and Ruth Millikan's earlier responses to the example says that "we both gave the same response to the intuition that such a being would, contrary to our theories, have contentful beliefs and desires. Namely, that since we were offering a posteriori theories of representation, rather than conceptual analyses of an everyday notion, we were pre-

Though we think this sort of reaction is a useful antidote to the uncritical appeals to intuition that many philosophers are happy to make, we do not believe that intuitions are philosophically worthless—not at all. At the same time, trying to say what legitimate role intuitions have in philosophy is a large and difficult project. We'll end our discussion with just a few programmatic remarks.[30] What we want to suggest is that there are at least two ways that intuitions may continue to be relevant to philosophical inquiry.

The first is perhaps the easiest to see. What makes intuitions philosophically useful is that they reflect one's beliefs about a kind. To the extent that intuitions do this, they can be helpful in orienting an investigation into the kind's nature, especially in the early stages when there is little else to work with. This is the kernel of truth to Jackson's idea that the central inferences in a network of belief define the subject matter of the embedded concepts. If you want to know what a given concept (e.g., the concept of a star) is about, a sensible place to start is with what you already believe about the relevant domain (e.g., the domain of celestial bodies).[31] On the other hand, there is no reason to say of any of the inferential links supported by intuition that they are analytic or conceptually unrevisable. Such labels have no explanatory justification, nor are they necessary for capturing the way that a subject matter is fixed. What determines a subject matter is simply the fact that the concepts in question have content. Any theory of content will do; the question is, and always has been, which is the right theory.

This brings us to our second role for intuitions. Intuitions can provide evidence about the content of a concept. They can do this even if they aren't constitutive of conceptual content, since intuitions may nonetheless be correlated with the conditions that are.

Wittgenstein famously championed the slogan "meaning is use". While many philosophers have followed his lead, not every theory of content is naturally construed as a use-based theory. On the other hand, all plausible accounts of content see use (broadly construed) as having important links to the nature of content. Given that categorization is clearly a central component of use, it turns out that our dispositions to categorize are non-accidentally correlated with conceptual content (to varying degrees) on virtually all theories of content. This is true even for those theories that deny that there are any content-constitutive inferences. Consider, for example, nomic-causal theories, which take content to be a matter of a representation standing in a particular type of causal relation to its referent. It bears emphasizing that this relation must be mediated and that typically the mediation involves cognitive

---

pared to reject the common-sense intuition that Swampman has contentful states" (1996, p. 130).

[30] For further discussion of these issues, see Margolis & Laurence (2003).

[31] See, e.g., Kornblith (1998).

resources—the beliefs, theories, perceptions, etc. that implicate the concept. Nomic-causal theories allow for a good amount of variation here. They don't care which particular beliefs, perceptions, etc. are in place, so long as some such set is able to support the crucial mind-world relation. But surely among the cognitive mechanisms that support this mind-world relation are the categorization dispositions that we all have. So even in this case categorization correlates with content (again, to some interesting extent).

This gets us half way to where we need to go; it establishes the link between categorization and content. The other half, concerning the link between intuition and categorization is relatively straightforward. It seems clear that intuition reflects our categorization dispositions, since forming an intuition is a matter of determining how we would categorize things in a given situation. Thus intuitions correlate with categorization dispositions and categorization dispositions correlate with content. *The result is that intuitions are broadly correlated with content on virtually any theory of content.* For this reason, they can be used as evidence for content, that is, evidence that a concept truly applies when intuition says it does.

We hasten to add that intuition is only a fallible guide to content—fallible for a number of reasons. One is that intuitions may not perfectly track categorization processes. What we actually do say about what we would say (i.e., in some hypothetical situation) isn't always accurate. Moreover, intuitions are apparently subject to variation from a great variety of cognitive and emotional influences, including, for example, the many different sorts of reasoning biases studied by Tversky, Kahneman, Nisbett, and others. Even more seriously, however, categorization doesn't always track content. At the very least, they come apart in the many different sorts of cases that Putnam has called our attention to and that have been discussed throughout this paper—cases where categorization is based on incomplete or erroneous information, especially information that may be corrected by empirical inquiry. Naturally, the degree to which intuitions are good evidence depends directly on the degree to which they are reliable. They aren't absolutely reliable, but they aren't absolutely unreliable either. They fall somewhere in between, providing a corresponding degree of evidence regarding content. Of course, it should go without saying that intuition isn't the only source of evidence here, nor is it especially privileged. Broad explanatory considerations are paramount to the study of content and continue to shape theories in the naturalistic tradition.

Conceptual analysis is an appealing theory to many philosophers, and it is important to try to understand the nature of this appeal. This second role for intuitions that we have identified may help to clarify the situation. We suspect that part of the temptation to embrace conceptual analysis derives from something like a folk semantic theory, which is implicated in the sys-

tematization of people's intuitions. People have a (relatively) pretheoretic conception of meaning and what determines it. Built into this conception is the idea that the central uses of a concept are constitutive of meaning and consequently that certain inferential links in the conceptual system are special (and a reflection of meaning) where others are not. This idea manifests itself both in the incredible historical dominance of the Classical or Definitional view of concepts, and in the broadly Wittgensteinian[32] idea that meaning is use.

Our suggestion is that the practice of conceptual analysis is rationalized by the folk theory. If the folk theory is taken seriously, then it makes perfect sense to think about one's categorization dispositions (as reflected in intuition) and use these to identify the central uses of a concept. On the other hand, folk theories are not always true, and in this case, as we have seen, there are good reasons to call it into question. All the same, the very intuitions that feed into conceptual analysis can continue to be relevant to philosophical inquiry. If we strip away the illicit contribution of folk semantics, the correlation between intuitions and content allows intuitions to have a certain amount of (defeasible) evidential value regarding conceptual content. This, we think, captures much of the pretheoretic appeal of conceptual analysis, while at the same time doing full justice to the arguments of Quine and Putnam. So while intuitions do have a role to play in philosophy, they don't have the final word or even the most important word.

---

[32]   We make no claim to historical accuracy concerning Wittgenstein's own views, which seem more hostile to positive theory than is compatible with the view alluded to here.

References

Bealer, G. (1987) The Philosophical Limits of Scientific Essentialism. In Tomberlin, J. (ed.) *Philosophical Perspectives*, 1, 289-365. Ridgeview: Atascadero.

Bealer, G. (1998) Intuition and the Autonomy of Philosophy. In DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc, pp. 201-39.

Block, N. & Stalnaker, R. (1999) Conceptual Analysis and the Explanatory Gap. *Philosophical Review,* 108.1, 1-46.

Boyd, R. (1988) How to be a Moral Realist. In G. Sayre-McCord, (ed.) *Essays on Moral Realism*. Ithaca: Cornell University Press.

Chalmers, D. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.

Crane, T. (1991) All the Difference in the World. *The Philosophical Quarterly,* 41, 1-26.

Cummins, R. (1998) Reflection on Reflective Equilibrium. In DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc, pp. 113-27.

Dennett, D. (1996) Cow-sharks, Magnets, and Swampman. *Mind & Language,* 11.1, 76-7.

Devitt, M. (1981) *Designation*. New York: Columbia University Press.

Devitt, M., & Sterelny, K. (1987) *Language & Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.

Dretske, F. (1981) *Knowledge & the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. (1988) *Explaining Behavior*. Cambridge, MA: MIT Press.

Field, H. (1996) The A Prioricity of Logic. *Proceedings of the Aristotelian Society*, 96, 359-379.

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1990) A *Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Jackson, F. (1982) Epiphenomenal Qualia. *Philosophical Quarterly*, 32, 127-36.

Jackson, F. (1986) What Mary Didn't Know. *Journal of Philosophy*, 83, 291-5.

Jackson, F. (1994) Armchair Metaphysics. In F. Jackson (1998) *Mind, Method, and Conditionals: Selected Essays*. London: Routledge, pp. 154-76.

Jackson, F. (1998) *From Ethics to Metaphysics,* Oxford: Oxford University Press.

Kornblith, H. (1998) The Role of Intuition in Philosophical Inquiry: An Account with No Unnatural Ingredients. In DePaul, M. and Ramsey, W. (eds.) *Rethinking Intuition*. Oxford: Rowman & Littlefield Publishers, Inc, pp. 129-41.

Kripke, S. (1972) *Naming and Necessity*, Cambridge, Massachusetts: Harvard University Press.

Laurence, S., Margolis, E. & Dawson, A. (1999) Moral Realism and Twin Earth. *Facta Philosophica*, 1.1, 135-65.

Lewis, D. (1994) Reduction of Mind. In D. Lewis (1999) *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, pp. 291-324.

Margolis, E. & Laurence, S. (2003) Should We Trust Our Intuitions?: Deflationary Accounts of the Analytic Data. *Proceedings of the Aristotelian Society*, 103.3, pp. 299-323.

McEvedy, C. (1988) The Bubonic Plague. In T. Brock (ed.) (1990) *Microorganisms: From Smallpox to Lyme Disease*. New York: W.H. Freeman and Company, pp. 3-12.

Millikan, R. (1984) *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Millikan, R. (1993) *White Queen Psychology; and Other Essays for Alice*. Cambridge, MA: MIT Press.

Papineau, D. (1993) *Philosophical Naturalism*. Oxford: Blackwell.

Papineau, D. (1996) Doubtful Intuitions. *Mind & Language,* 11.1, 130-2.

Putnam, H. (1962) The Analytic and the Synthetic. In H. Feigl & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science, Volume III*. Minneapolis: University of Minnesota Press.

Putnam, H. (1970) Is Semantics Possible? In H. Kiefer & M. Munitz (Eds.), *Languages, Belief and Metaphysics* (pp. 50-63). New York: State University of New York Press.

Putnam, H. (1975) The Meaning of Meaning. In K. Gunderson (Ed.), *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press.

Stich, S. & Laurence, S. (1994) Intentionality and Naturalism. In P. French, T. Uehling and H. Wettstein (eds.) *Midwest Studies in Philosophy,* v. 19: *Naturalism,* pp. 159-82. Notre Dame, Indiana: Notre Dame.

Zemach, E. (1976) Putnam's Theory on the Reference of Substance Terms. *Journal of Philosophy*, 73.5, 116-27.