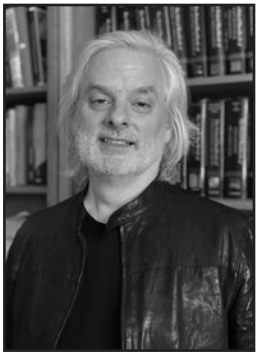
Inferentialism, Australian Style



David J. Chalmers
NEW YORK UNIVERSITY

Sanders lecture prepared for the ninety-fourth Pacific Division meeting of the American Philosophical Association in San Francisco, CA. The 2020 meeting was cancelled due to the coronavirus pandemic, and the lecture was delivered at the ninety-fifth Pacific Division meeting of the American Philosophical Association on April 10, 2021. The meeting was held virtually due to the coronavirus pandemic.

This lecture concerns inferentialism as an approach to the problem of naturalizing intentionality. My approach has a number of inspirations.[1]

My title is inspired by Jerry Fodor's 1984 article "Semantics, Wisconsin Style," which was a progress report both on work by others in naturalizing intentionality and on his own ideas. In the same way, this lecture is intended as a progress report both on recent exciting work by others in naturalizing intentionality, and on my own approach to the problem.

My approach is especially inspired by the late and much-missed Karen Neander, whose 2017 book *A Mark of the Mental* pioneered a two-tiered approach to intentionality that has been responsible for much of the recent progress in the field. I develop a two-tiered version of inferentialism about intentionality, partly inspired by the "Canberra plan" associated with Frank Jackson and others. The inferentialism that I favor is Australian style twice over, by combining Neander's two-tiered approach to intentionality with the Canberra plan.

I'll spell out all of this background in what follows, and then lay out my own preferred approach to grounding intentionality.

NATURALIZING INTENTIONALITY

Intentionality is aboutness. Our mental states seem to be about the world. When I believe that Biden is president of the US, my belief is

about Biden and about the US. When I see a dog, my experience is about the dog. Mental states that are directed at the world in this way are intentional states.

Intentional states typically have truth-conditions or at least satisfaction-conditions. My belief that Biden is president is true iff Biden is president. My desire that the pandemic is over is satisfied iff the pandemic is over. These truth- or satisfaction-conditions are *contents* of the intentional states. When I believe that it is raining, I have a belief whose content is *It is raining*. When I see a red square, I have an experience whose content is something like *Red square there*.

How can we explain intentionality? Or in the contemporary parlance, what are the *grounds* of intentionality? When I believe that it is raining, in virtue of what do I believe that it is raining? When I have a mental state with a given truth-conditional content, what makes it the case that I am in a state with that content?

The project of naturalizing intentionality aims to ground intentionality in natural terms—most often in physical terms. The project of naturalizing intentionality has been around for a long time, but it hit a peak in the 1980s, when it was one of the most active research programs in all of philosophy. Where theories of meaning in the philosophy of language had held center stage in the 1970s, an increasing conviction that meaning is grounded in mental content led to a focus on content in the philosophy of mind a decade later.

Approaches to naturalizing intentionality divided into a number of groups. *Causal* theories understand intentionality in terms of causal connections between the mind and the world, or in closely related terms characterized using information or counterfactual dependence. On one crude version of a causal approach, a mental symbol (a word in the "language of thought") such as CAT refers to cats if cats typically cause the symbol CAT to be tokened. Leaders here in the 1980s included Fred Dretske and Dennis Stampe, whose work Jerry Fodor chronicled in "Semantics, Wisconsin Style," as well as Fodor himself.

Teleological theories understand intentionality in terms of the *function* of cognitive systems, which is itself typically understood in terms of a system's evolutionary history. On one crude teleological approach, CAT refers to cats if CAT is produced by cats when the system is fulfilling its evolutionary proper function. Leaders here included Ruth Millikan, David Papineau, and Karen Neander.

Interpretivist approaches understand intentionality in terms of the best interpretation of a cognitive system, usually cashed out as whatever assignment of content serves best to rationalize a system's action. On a crude interpretivism, CAT refers to cats if this interpretation best rationalizes actions associated with the symbol. Interpretivist approaches have their roots in the work of Frank Ramsey in the 1920s (who was far ahead of his time) and W. V. Quine in the 1960s, with central figures in the 1970s and 1980s including Donald Davidson, Daniel Dennett, and David Lewis.

Inferential or *conceptual-role* approaches understand intentionality in terms of the role of mental representations in inference and in closely related processes. For example, a mental symbol AND picks out conjunction if it plays the core role of conjunction in inference (accepting P and accepting Q jointly lead to accepting P AND Q, accepting P AND Q leads to accepting P and leads to accepting Q). Inferential approaches have their roots in the work of Rudolf Carnap in the 1930s (who was also far ahead of his time) and Wilfrid Sellars in the 1950s. Key proponents from the 1970s through the 1990s include Ned Block, Robert Brandom, Hartry Field, Gilbert Harman, Paul Horwich, and Christopher Peacocke.

Phenomenal intentionality approaches understand intentionality in terms of consciousness. On some approaches, the intentionality of a perceptual state is grounded in the conscious experience of perceiving, and the intentionality of belief is grounded in the conscious experience of thought. This need not lead to a naturalizing of intentionality, but it still may tell us something about the grounds of intentional states. Although phenomenal intentionality has its roots in the work of phenomenologists such as Husserl, it was not as central in the 1980s discussion of intentionality. Interest was reignited around the turn of the century with key proponents including Terry Horgan and John Tienson, Brian Loar, David Pitt, and Charles Siewert.

The project of naturalizing intentionality was at the center of philosophy in the 1980s. In my graduate school years from 1989 to 1993, the program was still extremely active. Sometime in the mid-1990s, though, it suddenly ground to a halt. There was a good deal of work on the phenomenal intentionality program, but little of this work was devoted to naturalization. For two decades after 1995 or so, there was relatively little forward progress on naturalizing intentionality.

What happened? My diagnosis is that none of the main approaches provide a remotely plausible *global* theory of intentionality, assigning contents to all intentional mental states. Some theories, including

causal and teleological theories as well as phenomenal theories, are quite promising for perception and associated observational concepts, but seem hopeless for complex concepts such as democracy, mathematics, and education. Other theories, including interpretivism and inferentialism, are more promising for complex concepts but lead to rampant indeterminacies when viewed as global theories. For example, it is easy to permute contents of many representations while leaving the rational and inferential role of all representations preserved, so it is hard to see how rational and inferential role can determine content. Various attempts were made to extend these theories to global theories, but none were at all compelling. As a result, progress dried up for two decades.

TWO-TIERED THEORIES OF INTENTIONALITY

In the last few years, things have suddenly changed. There has been a renewed wave of activity in naturalizing intentionality, with a number of major books and important articles. What has happened to breathe life into the project?

The answer is pretty clear. We've seen a new wave of *two-tiered* theories of intentionality. Instead of applying a single global theory to all mental representations, these theories work in tiers. There is one theory of representation for simple representations, typically perceptual or nonconceptual representations. There's another theory for complex or conceptual representations.

The *locus classicus* here is Karen Neander's 2018 book *A Mark of the Mental*. This gives a teleological theory of perceptual representation, and instead of claiming a global theory of intentionality, limits the scope to just that. She says a theory of conceptual representation will wait for another day. Nicholas Shea's 2019 book *Representation in Cognitive Science* does something similar, offering a teleological theory of subdoxastic representation and stopping there, at least for now.

Robbie Williams's 2020 book *The Metaphysics of Representation* builds a second tier on top of Neander's foundation. He endorses Neander's theory of perceptual representation and builds an interpretivist theory on top of it. This theory avoids indeterminacy problems by assuming a first tier of source intentionality with fixed content. In this way a teleological and an interpretivist theory come together to yield a global theory of intentionality. The result is perhaps the most ambitious and comprehensive approach to naturalizing intentionality to date.

Adam Pautz's 2021 article "Consciousness meets Lewisian Interpretation Theory: A Multistage Account of Intentionality" does something similar. He gives a phenomenal account of the intentionality of perception, and interpretivist accounts of the intentionality of thought and language. Pautz's approach assumes primitive intentionality at the perceptual level, so its naturalistic ambitions are not as great as on Williams's approach, but its scope is otherwise similar.

Angela Mendelovici's 2018 book *The Phenomenal Basis of Intentionality* can also be seen as offering a two-tiered account. She defines a class of *immediate* contents for mental states that are grounded in their phenomenology. These immediate contents are often quite thin, however. On one version of the story, they include only perceptual contents, while on another version, they include a thin layer of cognitive phenomenology but certainly not enough to ground ordinary concepts. Then she introduces a further class of *derived* contents, where the derived contents of a state are given by the way the state can be "cashed out" into immediate contents. This is effectively a sort of inferential or conceptual role theory, where full mental contents are captured by inferential connections to immediate contents. Like Pautz, Mendelovici does not offer a naturalistic reduction of the first tier of intentionality, but her approach is otherwise global in scope.

In all of these two-tiered account, a causal, teleological, or phenomenal account of the first tier is combined with an interpretivist or inferential account of the second tier. Restricting the first group of theories in this way bypasses their problems in accounting for complex contents. Restricting the second group allows them to appeal to a level of source intentionality that avoids indeterminacy and permutation problems.

My own view is that the most plausible and powerful sort of two-stage view involves an inferential approach to the second tier. Mendelovici, Pautz, and Williams do not classify their theories as inferential or conceptual-role theories, but on examination each of them involves a heavily inferential element. We've already seen that Mendelovici's account can be understood that way. In Williams's account (as I discuss in a critical notice of his book, "Interpretivism and Inferentialism"), what's maximized is typically the rationality of certain inferential connections, leading to fairly standard inferential accounts of logical connectives and numerous other expressions. In Pautz's account, a "use theory" of linguistic content is specified in terms of dispositions to make certain inferences involving sentences in public language. As a result, most contents of thought outside a limited perceptual circle are grounded inferentially.

As it happens, I've argued (somewhat briefly) for a two-tiered inferentialist account of intentionality in *Constructing the World* (excursuses 17 and 19) and in "Verbal Disputes." As a result, this recent progress in theories of intentionality suggests some convergence. No doubt any sense of the world converging toward my view of these things is an illusion of perspective. Still, in what follows I'll aim to spell out in more details the sort of two-tiered inferentialism that I think has the best chance of being true.

INFERENTIALISM

Inferentialism about meaning is the thesis that the meaning of a word is constituted by its role in inferences. For example, "and" gets its meaning from the standard inferential role for conjunction. Inferentialism about mental content is the thesis that the content of a mental state or a representation is constituted by its role in inference. Inferentialism is sometimes called "conceptual-role semantics" or "inferential-role semantics"—but those names are not so apt for a view about mental content these days, now that the term "semantics" is almost always restricted to the linguistic domain. These labels also tend to suggest a view on which meanings *are* inferential roles (semantics), rather than a view on which meanings are grounded in inferential role (metasemantics). Given that my main focus is on mental content and on metasemantics, I use the term "inferentialism" instead.

"Inference" is often understood narrowly to include just transitions between judgments or beliefs, but I'll understand it broadly to subsume many aspects of conceptual role. For my purposes, many different transitions between mental states can be understood as inferences, including transitions from perception to belief, from belief and desire to intention, from supposition to conclusion, and more.

The founding figure for inferentialism in the twentieth century is Rudolf Carnap in *The Logical Syntax of Language* (1934) and other works. Sometimes this role is ascribed to Sellars, but Sellars's first paper on the topic, "Inferentialism and Meaning" (1953), is largely a response to Carnap's inferentialist theses from two decades earlier.

Inferentialism about the mental usually presupposes a language of thought. The idea is that mental representations such as AND, CAT, and HESPERUS have their content in virtue of their inferential role.

There are any number of different versions of inferentialism. Holistic versus molecular inferentialism: Does the total inferential role of a representation matter, or just a privileged subset? (My approach is molecular.) Long-armed versus short-armed inferentialism: Do causal connections between mind and world count as relevant inferential roles, or just narrow connections between mind and mind? (My approach is short-armed.) Mental versus linguistic inferentialism: Does inferentialism offer an account of mental or linguistic content? (My approach can apply to both, but I am analyzing mental content in the first instance.) Local versus global inferentialism: Does inferentialism apply to all concepts and representations or just some? (My two-tiered approach is strictly speaking local, since inferentialism does not apply to concepts in the first tier, but the first tier is relatively small, so the approach is relatively global.) Semantic versus metasemantic inferentialism: Are meanings and contents identical to inferential roles, or grounded in those roles? (My approach is metasemantic.) Normative versus descriptive inferentialism: Are the relevant inferences those that should be performed, or those that are actually disposed to be performed? (I'll consider versions of each.)

A key distinction for my purposes concerns whether inferentialism is intended as an account of *truth-conditional* content. The large majority of inferentialist approaches to date are non-truth-conditional. Block, Field, and Harman explicitly separate conceptual role from truth-conditions as two different kinds of content, and do not try to account for one in terms of the other. Horwich is a minimalist about truth who is not interested in recovering truth-conditional semantic values. Brandom rejects truth-conditional content entirely, saying that mind and language have inferential but not representational content.

For my part, I am a pluralist about content, so I'm happy to allow that there are notions of content understood wholly in inferential terms. But in the current project, my own interest focuses almost wholly on truth-conditional content and related notions such as reference. The inferentialism I am interested in is in the same business as causal, teleological, and interpretivist theories of mental content: explaining truth-conditions and reference. That is, in the terms introduced above, I'm interested in inferentialism not as semantics but as metasemantics. So I will set aside the sorts of inferentialism in the previous paragraph. Among existing inferentialist works that I'm aware of, perhaps Carnap's "Meaning and Synonymy in Natural Languages" (1955), Peacocke's *A Study of Concepts* (1992), and my own remarks in *Constructing the World* (2012) come closest to trying to explain truth-conditions across many domains in inferential terms.

Carnap does not cast his 1955 account as an inferential account, but as I'll discuss, it clearly fits the mold. Peacocke's 1992 account is presented as an inferential or conceptual-role account of concept possession. He does not present it as an account of truth-conditional content, but combining his account of concept-possession with his "determination theories" of how conceptual roles determine semantic values, one can naturally understand it as an account of how subjects are in states with certain truth-conditional contents in virtue of being in states with certain conceptual roles. Both Carnap's and Peacocke's inferentialisms appear to be two-tiered: Carnap's for reasons I'll discuss, Peacocke's because it presupposes a layer of nonconceptual content (e.g., for perception) that is not grounded in inferential role.

Truth-conditional inferentialism has always faced a large problem: How can inferential roles cross the gap from mind to world? Inference is a mind-mind relation, while truth-conditions require a mind-world relation. This is a serious issue for a *global* inferentialism that tries to understand all content in inferential terms. However, it is much less of a problem for a non-global inferentialism. In a two-tiered inferentialism, a first tier of non-inferential content crosses the bridge from mind to world, so that the second inferential tier simply needs to build on this mind-world connection.

TWO-TIERED INFERENTIALISM

One way to motivate two-tiered inferentialism is through frameworks for conceptual analysis, such as the Australian-style "Canberra plan" associated with Frank Jackson and other figures. To see how this might work, we can start with a highly traditional sort of conceptual analysis, on which all concepts can be analyzed in terms of certain simple concepts. For example, Russell's descriptivist view yields an analysis on which all concepts are composed from a few simple concepts—concepts of sense-data, universals, and perhaps the self—each of which is itself grounded in acquaintance with its object.[2]

If Russell's descriptivism were true, a two-tiered view of intentionality would follow. On the first tier, we'd have the simple concepts deriving from acquaintance. On the second tier, we'd have complex concepts composed from these simple concepts. The project of naturalizing intentionality would then be reducible to the problem of naturalizing the content of simple concepts, and naturalizing their mode of combination.

These days, classical descriptivism is widely rejected. There are numerous reasons, but the most influential and powerful reason is that full-scale conceptual analysis almost never succeeds. Almost every conceptual analysis that has ever been put forward has been a failure, in that it is subject to counterexamples. The classic case is the concept of *knowledge*. Some analyzed this as *justified true belief*, but Gettier produced counterexamples. Some refined the analysis, but there were more counterexamples in turn. No one has ever produced a widely accepted counterexample-free analysis. What goes for *knowledge* goes for most ordinary concepts. So philosophers have largely become convinced that most ordinary concepts are not composed from simple concepts.

That said, a weaker form of conceptual analysis remains plausible. This is *casewise* conceptual analysis, where one describes a case and asks whether it is an instance of the concept. For example, Gettier exhibited certain cases of justified true belief and argued that they were not instances of knowledge. Others overwhelmingly agreed. One can use this process to refine improved if imperfect analyses of the concept, such as "justified true belief" plus a fourth condition. These analyses can be tested by cases and refined in turn. The same goes for most concepts. We can describe cases, and if the descriptions are fleshed out enough, we can usually make a judgment about whether the cases fall under the concept. Gettier-style counterexamples to full-scale conceptual analyses do nothing to undermine casewise conceptual analysis. In fact, they arguably rely on it, in that it's arguable that we classify these counterexamples using our conceptual competence alone.

Casewise conceptual analysis can be used to motivate an inferentialist analysis of intentionality. An analysis of meaning along these lines was proffered by Carnap himself in "Meaning and Synonymy in Natural Languages." Carnap said that to determine the *intension* of a term like "horse" as used by a subject, we should present that subject with cases (sometimes described in language, sometimes presented in the form of pictures or movies) and ask the subject whether the case is a horse or not. On Carnap's framework, the subjects dispositions to judge cases determine the intension of the word for the subject. Some possible cases will be counted as "horse," while others will be excluded. The intension of "horse" will be something like the class of possible cases that the speaker counts as horses.

As I discuss in "Revisability and Conceptual Change in 'Two Dogmas of Empiricism'," Carnap's framework has obvious limitations. One limitation arises from the fact that speakers can make mistakes about the

extensions of their terms, and the method requires heavy idealization to eliminate these. Another is that the method does not yield a global analysis of meaning or intentionality, since it presupposes a basic vocabulary used to describe the cases (or at least pictorial intentionality for pictures of cases, and so on). As a result, Carnap's method works best as a two-tiered account. If we presuppose the meaning of a first tier of expressions and use them to describe the cases, Carnap's method can be used to ground a second tier of meanings for other expressions and concepts. These meanings are in effect grounded in inferences from descriptions of cases (e.g., the Gettier case) to judgments involving the concept in question (e.g., John knows that p). A raft of dispositions to make inferences along these lines constitutes the meaning of a term such as *knowledge*.

SCRUTABILITY AND TWO-DIMENSIONAL SEMANTICS

For a two-tiered approach to reference and truth-conditions in this Carnapian mould to work in general, we need what I have called (in *Constructing the World*) a *scrutability* thesis. The rough idea behind a scrutability thesis is that given enough information about the world (or about a possible scenario), in a limited vocabulary, we are in a position to make inferences (under idealized reflection) leading to knowledge of what our expressions refer to and of whether our sentences are true.

More precisely, a sentence S is a priori scrutable from another sentence D for a speaker if the conditional 'If D, then S' is knowable a priori (under idealized rational reflection) for the speaker. In *Constructing the World*, I argued for an a priori scrutability thesis holding that all truths are a priori scrutable from a truth involving a limited class of concepts. The base truth I relied on most was PQTI: a conjunction of microphysical and phenomenal truths about the world along with indexical truths and a that's-all truth. The idea is roughly that given the information in PQTI, ideal a priori reflection alone should enable knowledge of all other truths.

While PQTI just describes the actual world, a generalized scrutability thesis applies to other possible scenarios. The rough idea is that there is a limited class of expressions C, such that for any sentence S that is epistemically possible (in the sense that ~S is not a priori), S is a priori scrutable from some epistemically possible sentence involving only expressions in C. PQTI-style specifications of scenarios are in effect generalized to describe arbitrary epistemically possible scenarios. The truth or falsity of all sentences are then scrutable from these descriptions.

Given the generalized scrutability thesis, we can define a Carnap-style intension for arbitrary concepts or expressions. On this picture, any concept can be applied inferentially to various scenarios described in the base vocabulary. We use basic concepts to describe scenarios, and we assign extensions to nonbasic concepts according to scrutability from those scenarios. For example, 'water is H_2O' is a priori scrutable from a description of an Earth-style scenario, while 'water is XYZ' is a priori scrutable from a description of a Twin Earth scenario. So the primary or epistemic intension of 'water is H_2O' (for a given speaker) maps the Earth scenario to "true" and the Twin Earth scenario to "false." We can also say that the primary intension of 'water' maps the Earth scenario to H_2O and the Twin Earth scenario to XYZ.

This semantic framework shares the basic structure of two-tiered inferentialism. The primary intension of a concept is characterized by its inferential role with respect to a first tier of basic concepts. In particular, it is characterized by how the concept applies to scenarios characterized using these basic concepts. We can start with sentences and thoughts. For a given sentence S (say 'water is H_2O'), the primary intension depends on whether S is a priori scrutable from D for various scenario descriptions D involving basic concepts. That's a matter of the a priori inferential role of S with respect to basic concepts.

We could in principle specify this inferential role as an infinitary set of entry inferences. For a sentence S, we can say $D_1 \vdash S$, $D_2 \vdash \sim S$, and so on for each scenario. D_1, D_2, etc. are restricted to concepts in the basic class, while *S* is not. We can use this method to also specify primary intensions for subsentential expressions, associating 'water' with a function from scenarios to extensions in those scenarios. (The details take some work, discussed in section 3.7 of "The Foundations of Two-Dimensional Semantics.") In effect, the intension of 'water' is defined in terms of its a priori inferential role. We can assign primary intensions in a similar way to mental items such as judgments and concepts.

The primary intension of S can be seen as a truth-condition for S. It yields a set of scenarios in which S is true and a set in which S is false. It's true that scenarios are characterized above in linguistic terms using expressions for basic concepts, so intensions across these scenarios do not yet cross the mind-world barrier. But if we have an independent account of the content of basic concepts, and in particular their reference or reference-conditions, then this will cross the mind-world barrier. This way, a description D specifies a possible world (at least an epistemically possible world) in which D is true, where possible worlds

are worldly entities constructed from properties and other entities. This way a primary intension yields a truth-condition across possible worlds.

This two-tiered approach also allows us to connect primary intensions to truth and reference. The primary intension of a sentence can be evaluated at the actual world (by taking a scenario description D that is true at the actual world), yielding the truth-value of the sentence. Something similar can be done for reference: the primary intension of a subsentential expression can be evaluated at the actual world, yielding the actual referent. This way, primary intensions serve as conditions of truth and reference.[3]

Primary intensions are coarse-grained contents. For example, all a priori sentences (including all mathematical theorems) have the same primary intensions. It is not hard to make them more fine-grained, though. One can define structured primary intensions by associating simple expressions with primary intensions and then associating sentences with structures of these intensions. One can also individuate primary intensions more finely by moving to an analytic scrutability thesis (discussed in chapter 8 of *Constructing the World* rather than an a priori scrutability thesis. This will involve finer-grained scenarios and will have the result that synthetic a priori truths (such as many mathematical and normative truths, arguably) will generally have different intensions. For simplicity I will stay with the coarse-grained account, but most of what I say will generalize.

Under certain plausible assumptions, primary intensions are a sort of narrow content, depending only on features internal to the organism. As a result, Oscar on Earth and Twin Oscar on Twin Earth will have the same primary intensions associated with their terms "water." To a first approximation one might think of this as an intension picking out the dominant clear drinkable liquid in a given scenario, though this would need much refinement. Of course, Oscar's and Twin Oscar's terms may have different referents, H_2O and XYZ, respectively. The primary intension of "water" picks out H_2O in an Earth scenario and XYZ in a Twin Earth scenario. So reference is a form of wide or environment-dependent content. There will also be different wide truth-conditions for sentences on Earth and Twin Earth as a result.

This yields the familiar two-dimensional semantic framework for thinking about content and truth-conditions, with a narrow primary intension and a wide secondary intension. For present purposes, however, I am focused largely on the primary intension. It is arguable that in combination with facts about the thinker's environment, primary intensions themselves

ground reference and wide content. This grounding claim requires further argument, but it's an attractive claim for an inferentialist to make, as it allows that if short-armed inferential roles ground primary intensions (narrow content), then in conjunction with facts about the environment, they will also ground reference and secondary intensions (wide content).

I won't try to defend the two-dimensional framework or the scrutability theses that underlie it here. I'll simply assume the framework, and see what follows for the project of naturalizing intentionality.

NORMATIVE INFERENTIALISM

The two-dimensional semantic framework I've presented is not a metasemantic framework. It characterizes certain contents (primary intensions) in terms of certain a priori inferential roles, but it does not make claims about how the contents are grounded, and in particular whether the contents are grounded in the roles or vice versa. So the framework does not have immediate consequences for naturalizing intentionality. Still, it suggests a possible approach to naturalization. Since the framework characterizes primary intensions in terms of certain a priori inferential roles, we can postulate that intensions are grounded in these roles. To naturalize intentionality, we then need only naturalize these roles.

The inferential roles that define primary intensions are normative rather than descriptive roles. The key claim is that conditionals such as "If D then S" are knowable a priori. In inferential terms, this amounts to saying roughly: if the subject were to suppose D, they ideally *should* conclude S via a priori reasoning. In related epistemic terms, one could also say that there is an a priori warrant for the subject to make this inference, whether or not they have the capacity to use it. This contrasts with Carnap's descriptive inferentialism where the defining roles concern what the subject *would* infer from case descriptions like D.

Can we use all this to develop a metasemantic framework where primary intensions are grounded in normative inferential roles, and where these roles are used to naturalize intentionality? It's worth trying, but there are a number of obstacles.

First, some will argue that the inferential roles used to define primary intensions are so heavily idealized that they are not genuine normative roles at all. No human subject can entertain full world-descriptions like

D or make inferences from them. Given that ought implies can, it follows that no subject ought to make these inferences, so that there are no inferential norms here.

Now, I think there are idealized oughts to which ought-implies-can does not apply. Or to put a small twist on the same thought, there are idealized oughts that imply only a heavily idealized "can." These are the oughts that matter for my purposes. But if someone denies that there is any "ought" here, not much rests on the word "normative." What matters is that the facts about idealized apriority are well-defined (as I argue they are in *Constructing the World*). If so, there remains the possibility of grounding content in a priori inferential roles, whether or not these roles are genuinely normative.

Furthermore, these idealized inferential roles may well be grounded in far less idealized roles. For a start, we need not always invoke inferential roles with respect to scenarios described in terms of basic concepts. For a given concept C, we can instead invoke less constrained inferential roles: how thoughts involving C are a priori scrutable from non-C-involving thoughts (instead of from thoughts involving only basic concepts). We can avoid circularity by supposing a partial ordering on concepts, with basic concepts at the bottom. We always specify a concept C's inferential role with respect to concepts that are below C in the ordering, but which need not themselves be basic concepts. Given plausible assumptions, this less idealized inferential structure will suffice to determine a primary intension for C.

We can also invoke perceptual or introspective presentation of scenarios instead of linguistic descriptions. As Carnap suggested, we can sometimes use a picture or a movie (or perhaps a virtual reality environment) to specify a case, which subjects can then use to form an image-based state of imagining that a certain scenario obtains, and making inferences from there. Or we can start with states that mix imagery with non-imagistic judgment and consider inferences from there.

These inferential roles involve a less demanding basis for specifying scenarios, but they still require specifying a full scenario in each case, or at least spelling out enough detail to make it conclusive whether the case is a case of C. Still, these roles may be grounded in turn in simpler inferential roles that involve simpler specifications of cases with sometimes defeasible inferences. For example, it is arguable that the concept *knowledge* satisfies a hierarchy of inferential roles like these: (1) ~true(p) ⊢~K(p); (2) ~believe(p) ⊢~K(p); (3) justified(p) & true(p) &

believe(p) $\vdash^{*}\sim$K(p), (4) grounded-in-falsehood(p) $\vdash\sim$K(p), (5) lucky-true-belief(p) $\vdash^{*}\sim$K(p), and so on. Here "$\vdash$" represents a nondefeasible inference and "$\vdash^{*}$" represents a defeasible inference. It's arguable that even the defeasible inferences here are part of the concept of knowledge: they're inferences for which one has a conceptual warrant, albeit a defeasible one. The idea is that one proceeds through the hierarchy until the case is nondefeasibly classified. If no rule nondefeasibly classifies the case, then the last defeasible classification carries the day if there is one; else the case is indeterminate.

In this way, a hierarchy of relatively simple inferential roles can define a primary intension for a speaker. Of course, the roles may need to be complex to handle some cases. But this way the primary intension will often be defined across an enormous proportion of cases without needing heavily idealized roles, with idealized complex roles required only for very hard cases. One also has the option of ignoring inferential roles beyond a certain level of idealization (perhaps the level where "ought implies can" threatens to exclude these roles), thereby defining a sort of approximate primary intension that will be useful for many purposes. These approximate primary intensions may not perfectly capture the concept's application conditions: they will yield a finite analysis that like other finite analyses will be subject to counterexamples in hard cases. But this way, the need for idealized roles will at least be restricted to a few hard cases rather than being required across the board.

By far the most important objection to normative inferentialism is the plausible thesis that normative inferential roles are themselves grounded in mental content. If normative roles are grounded in content, then content is not grounded in normative roles. For example, the apriority of "If D, then S" plausibly depends on the meaning or content of S. It's true that we can sometimes classify a sentence as a priori without focusing on its content. In the two-dimensional semantic framework, apriority is used in this way to explicate semantic values (since our grip on apriority may initially be firmer than our grip on primary intensions), but this is not to say that apriority grounds content. On the face of it, the apriority of a sentence is most naturally understood in terms of the apriority of the proposition it expresses. Something similar applies to the apriority of judgments and inferences. More generally, the inferences one ought to make from a judgment plausibly depend on its content and are grounded in that content. If the epistemic status of a judgment or of inferences from it are grounded in its propositional content, then they cannot ground those propositional contents.

A proponent might stipulatively define primary intensions in terms of normative inferential roles in such a way that the intensions are grounded in the roles by definition. But this does not escape the problem. We still have reason to think that the normative role of a sentence or judgment is grounded in its content. It's just that the grounding content will be something distinct from its primary intension. If primary intensions are grounded in roles that are grounded in some other sort of content, we'll still need to naturalize content of the latter sort, and the current framework is no closer to showing us how to do this.

Now, the relationship between normativity and content is not entirely obvious. Normativists such as Brandom think that normative statuses of mental states can be grounded without presupposing content. As an analogy, someone might claim that the property of *belief* is defined in terms of content: to have a belief is to believe a content and is grounded in that content. All the same, many philosophers think one can understand the property of being a belief in content-independent terms. For example, perhaps to have a belief is for one's cognitive system to have a sentence in the language of thought that plays a certain specified functional role—or that is found in the "belief box," as it is sometimes put. It is not obviously out of the question that to have an a priori belief might be equivalent to a functional kind in a similar way. Likewise, it is not out of the question that to make an a priori inference could be a matter of having two states that stand in an appropriate psychological relation that can be specified without mentioning content. If so, a priori inferential role need not be grounded in content.

Still, I am not sure I have a good grip on the normativist view and I will not pursue it here. If one goes this way, the naturalization project will then turn on questions about how the normative inferential roles are grounded if they are not grounded in content. Either they are primitive, in which case there is no naturalization, or they are grounded in something other than content. One natural candidate is that they are grounded in dispositional inferential roles. That will bring us back to the issues we consider in the following section. In that way the naturalization projects for normative and non-normative inferentialism could end up converging.

If normative inferential roles are grounded in content, then content is not grounded in normative inferential roles. However, it remains possible that content (and thereby normative inferential roles) are grounded in *dispositional* inferential roles. I'll pursue this idea in the next section.

DISPOSITIONAL INFERENTIALISM

What grounds the primary intensions and the normative inferential roles of our concepts and judgments? It is possible that these things are primitive, but I'd be surprised. It is also possible that both things are grounded in something other than inferential roles.

For example, if there were a sufficiently rich cognitive phenomenology associated with a judgment, one could argue that this phenomenology grounds its normative inferential role. As before, phenomenal intentionality may not provide a clear path to naturalizing intentionality—there would still be questions about whether the phenomenology itself involves content, and about what grounds the phenomenology—but it may at least provide some insight into the grounds of inference. Still, I am doubtful that most of our judgments have cognitive phenomenology rich enough to ground their entire normative inferential role and their entire content. If they do not, we need another route to grounding primary intensions and normative inferential roles.

The most obvious route to grounding these things is an appeal to dispositional inferential roles. One simple idea is an adaptation of Carnap's: the primary intension of S for a subject is true at a scenario described by D iff the subject is disposed to judge that if D, then S (or to infer S from the supposition that D).

As before, an obvious problem for this simple idea is that these inferential roles are too idealized. Ordinary speakers cannot entertain huge descriptions like D, so it's hard to see how they can have dispositions about what they'd judge if they entertained D. Perhaps one could appeal to special counterfactuals about what speakers would do (or what they could do) if their capacities were greatly expanded. As Kripke (1981) has noted, it is not obvious that these counterfactuals are well-defined. I'm somewhat more optimistic than Kripke that these counterfactuals can give us some purchase, but it moves us a long way from standard inferential roles. So I'll set aside the simple idea here.

Instead, a dispositional inferentialism needs to ground primary intensions in less idealized inferential dispositions. We've seen in the previous section that a primary intension can be determined in principle by far less idealized inferential roles. Given these less idealized roles, it is more tenable to hold that some subjects have the relevant dispositions.

Even where less idealized dispositions are concerned, there are still substantive questions about whether a subject could possess the

concept without having the relevant dispositions. I'll return to this issue later. But for present purposes, I'll make my task here easier in two ways. First, I'll restrict my attention to conditions for a subject *mastering* a given concept, thereby setting aside cases where a subject possesses a concept merely in virtue of deferring to a linguistic community that uses the concept. Second, I'll be happy if we can establish a *sufficient* condition for mastering the relevant concept, even if this is not a necessary condition. A merely sufficient condition for content would still be a large step on the road to naturalizing intentionality, especially if the sufficient condition is inclusive enough to capture at least some human beings, or creatures not too far beyond our own capacities.

How might a subject's actual inferential dispositions involving a concept determine its content? There's a standard inferentialist proposal. A concept has that content that makes the inferences involving it (or a special subset of those inferences) as truth-preserving as possible, or as rational as possible. I'll adopt a version of that approach here.

A first decision point here is whether the determination of content by inferences is holistic or molecular: that is, whether all inferences involving a concept are relevant to determining its content, or just a special subclass. My own approach is molecular. A holistic approach threatens to give the inference from *bachelor* to *untidy* as much weight in constituting content as the inferences from *bachelor* to *unmarried*. But the former inference seems more or less irrelevant to the primary intension of *bachelor*. Primary intensions are contents that are tied especially to the a priori inferential role of a concept, not to its inferential role more generally. Intuitively, the relevant inferences are the *a priori* inferences. As before there's case for individuating primary intensions more finely and using something akin to analytic inferences—but I'll focus on a priori inferences here.

Quineans often reject molecular approaches on the grounds that they require something like an analytic/synthetic distinction or an a priori/a posteriori distinction. I'm happy to endorse these distinctions, and I've argued that Quine's arguments against them fail, so this objection doesn't worry me. There does remain a substantive issue about whether I can appeal to the normative concept of apriority (or the semantic concept of analyticity) at this point for the purposes of a naturalistic reduction. If apriority is grounded in content, as I largely conceded in the previous section, a circle threatens. For the purposes of naturalization, I will eventually need to cash out the relevant class of inferences in some other way.

For now, I'll call the relevant class of inferences the a priori inferences, to at least roughly fix reference to the relevant class. I'll start by assuming that the members of this class correspond to conclusive (or indefeasible) a priori inferences, those inferences that can in principle be justified with certainty. Conclusive apriority plays a special role in the two-dimensional framework. Conclusive a priori sentences and judgments have a primary intension that is true at all scenarios, and conclusive a priori inferences are truth-preserving at all scenarios.

How might dispositions to make a priori inferences determine a primary intension? There's a natural story here. For a given concept C, we collect all a priori inferences that the subject is disposed to make between judgments involving only C and basic concepts. These will include both entry inferences (from judgments not involving C to judgments involving C) and exit inferences (vice versa), but they may also include inferences with C on both sides (which may be helpful in some cases). We then assign C a primary intension (that is, an extension at each scenario specifiable using basic concepts) such that all these inferences are truth-preserving at all scenarios. If more than one primary intension has this property, then the content of C will be indeterminate between these primary intensions.[4]

We can extend this picture to allow entry and exit inferences that go beyond basic concepts. As in the previous section, we need only suppose a partial ordering on all concepts, with basic concepts at the bottom. Then our key content-constituting inference can include a priori entry and exit inferences where the non-C concepts can include any concepts that fall below C in the partial ordering. We can then assign primary intensions to all concepts just above the bottom level and so on recursively. We can also include perception and image-based imagination to play a role in our specification of scenarios. This allows a more expansive class of inferences to serve as the basis for primary intensions, thereby increasing the determinacy of these intensions.

One objection to this way of proceeding is that there may be relatively few conclusive a priori entry and exit inferences involving most ordinary concepts that are accessible to ordinary thinkers. We have already seen that there are counterexamples to more or less every proposed analysis of terms such as *knowledge*, at least when the analysis is short enough to be graspable by ordinary thinkers. You might worry that the same applies to inferences: when an inference is simple enough that ordinary thinkers are disposed to make it, there will usually be counterexamples so that the inference is not conclusive.

The matter is not cut and dried. One can argue that despite the unanalyzability of knowledge, the inferences from knowledge to truth and belief remain conclusive. Likewise, unanalyzability is consistent with there being conclusive a priori inferences from many briefly described cases (e.g., a Gettier case) to a classification as knowledge or not. Still, many classificatory inferences will be defeasible by new information about the case. If this sort of defeasibility is the rule rather than the exception, then conclusive (indefeasible) inferences by nonideal reasoners may not have enough structure to determine primary intensions.

An alternative is to allow a role for defeasible (or nonconclusive) a priori inferences: roughly, those that are justified a priori even though they are not justifiable with certainty. We saw in the last section that inferences such as that from *justified true belief* to *knowledge* may have this status. We've also seen that in principle, a collection of defeasible and indefeasible inferences can determine a primary intension that is not determined by the indefeasible inferences alone. And it is arguable that ordinary subjects can have dispositions to make many more defeasible a priori inferences than they make conclusive a priori inferences.

How might we exploit defeasible a priori inferences in the current framework? We can give conclusive a priori inferences the same role as before: collect them up and use them to determine at least a class of primary intensions that make these inferences truth-preserving at all scenarios. If there are relatively few such inferences, then the resulting class may be quite large. At this stage, we can invoke defeasible a priori inferences, selecting those primary intensions that makes the defeasible inferences truth-preserving at as many scenarios as possible. This requires a measure on the space of scenarios. We will also need to respect defeasibility structure, as in the last section: where one defeasible inference can trump another, the former should be given more weight. (For example, if $A \vdash C$ is trumped by $A\&B \vdash \sim C$, one should try to make the latter but not the former truth-preserving at A&B-scenarios.) Working out the details would take considerable work. But this way of proceeding offers perhaps the most hope of using actual dispositions to determine something like a primary intension.

One could object that most ordinary subjects have only limited dispositions to make even defeasible a priori inferences. But this matter is now very far from cut and dried. Ordinary subjects are disposed to make any number of inferences involving concepts, and it is plausible that a reasonable number of these inferences have at least defeasible a priori status. Furthermore, recall that my thesis is restricted to cases of

concept mastery, and my aim is only a sufficient condition for having a primary intension in these cases. If there are subjects (young children, for example) who possess a concept but nevertheless do not have the requisite a priori inferential dispositions to determine its primary intensions, these will often be cases of concept possession without concept mastery. Even if there are cases of concept mastery (perhaps in unreflective subjects) without the requisite dispositions, which is far from clear, then as long as there are also cases where the requisite dispositions are present (perhaps in reflective subjects), this is no objection to a substantive sufficiency claim. For that purpose, we need only augment our account with a condition (spelled out in noncircular terms) that is satisfied by members of the second class but not the first. The account will then say that among subjects who satisfy this condition, inferential dispositions determine primary intensions.

Perhaps the most important objection to dispositional inferentialism is that it is unclear how to delineate the meaning-constitutive inferences in a noncircular way. So far I have said these are the a priori inferences. But as we saw in the previous section, the apriority of an inference is a normative status that is plausibly grounded at least partly in the contents of the judgments involved. If so, a dispositional inferentialism that appeals to apriority does not offer a grounding of content, let alone a naturalization of content.

One general worry is that being an inference at all is always grounded partly in accepting certain judgments, which is grounded in accepting their contents, so there are no facts about the inferential role of a judgment or a concept that are not grounded in contents. In response, we can at least try to isolate certain structural features of inferences that are present over inferences with many different contents and that do not turn on any specific contents. If so, then while the notion of inference may presuppose the notion of content, we can count certain transitions as inferences without this being grounded in any specific content of the judgments involved. A transition may be an inference from a judgment that p to a judgment that q in virtue of the contents involved, but it may be an inference simply in virtue of structural features that do not involve any contents.

In the same way, we can look for a structural characterization of our class of a priori inferences in a way that does not presuppose specific contents for the judgments involve. One can glean various proposals for doing this in the literature.

First, Peacocke characterizes them as inferences that are primitively compelling, in that the subject finds them compelling in a way that is not in virtue of finding other premises compelling and is not answerable to anything else. For our purposes compellingness should be understood as a psychological rather than a normative notion, in order to avoid the circularity issue. What matters for my purposes is that an inference is not psychologically based in accepting other claims (and in particular is not based in perception or in empirical judgment), whether or not it should be so based.

Second, there is a Quinean characterization of the a priori as unrevisability, so a priori inferences are inferences that the subject will never reject, or will never reject in certain ways. This characterization is more apt for the conclusive a priori than for the defeasible a priori.

Third, there is a phenomenological characterization: a priori inferences have a distinctive phenomenology, and the relevant class of inferences can be characterized as those with that phenomenology.

All of these characterizations may be imperfect markers of a priori inferences. Perhaps one can find certain empirical inferences primitively compelling, and may be disposed never to revise them, and they may have the phenomenology of apriority. Still, it is not out of the question that these psychological correlates of apriority might be usable at least to a first approximation in building an inferential theory of content.

I think that finding psychological markers of a priori inference is itself an open research program, and I won't try to definitively settle the issue here. For my purposes, I'll simply call the relevant class of inferences the C-inferences, where C can stand for "constitutive," "compelling," or "conceptual." I'll assume that the being a C-inference is not grounded in the specific contents of the judgments involved.

On one version of this account, C-inferences correspond at least roughly to *conclusive* a priori inferences. On this account, we assign primary intensions so that C-inferences are always truth-preserving across all scenarios. This method works best if we have an understanding of C-inferences so that they are always conclusive a priori. If C-inferences are not always conclusive a priori inferences, the intensions defined using C-inferences will differ from standard primary intensions: in particular, they will not have the standard property that a judgment is conclusively a priori iff it has a necessary primary intension. That said, these intensions might still be near-relatives of the standard version

that are still useful contents for many purposes and that are more straightforwardly definable in naturalistic terms.

On another version, C-inferences include both conclusive and defeasible a priori inferences. On one account in this vicinity, being a C-inference may come in degrees. Pure C-inferences correspond to the conclusive a priori. Impure C-inferences correspond to the defeasible a priori. We assign primary intensions as before so that pure C-inferences are truth-preserving across all scenarios if possible, or as near as possible, while impure C-inferences are truth-preserving as widely as they can be. If C comes in degrees, we can weight inferences in this process according to their degree. Again, the closer C-inferences are to a priori inferences, the closer the resulting intensions will be to primary intensions. In this way there's at least some prospect for grounding something like primary intensions noncircularly in inferential dispositions.

OBJECTIONS

How does the current framework deal with standard objections to dispositional inferentialism?

Plus and quus (Kripke's Wittgenstein). Kripke famously argues that dispositions do not fix the content of a word like "plus," because the dispositions are finite and can involve errors. This is a problem for a global dispositional inferentialism, but not for the two-tiered variety. Addition can straightforwardly be defined in terms of concepts such as *integer* and *successor*. So as long as these concepts are included in the first tier, a priori inferences involving them will fix the content of "plus" in the second tier. More generally, we can expect the first tier to fix reference to properties and relations that involve "rails to infinity," as the successor relation does. As a result, inferential roles with respect to these concepts can define more complex rails to infinity, and there is no longer an obstacle to finite dispositions fixing infinitary contents.

How do dispositions ground normativity? This is a hard question for any dispositional account. How and why does being disposed to make certain inferences make those inferences (or any other inferences) inferences that one *should* make? I don't have a full answer to this. But I like the standard inferentialist ideas that when certain inferences are constitutive of a given content, those inferences are necessarily truth-preserving in virtue of this constitutive role, and this suffices for them to be justified. This leaves residual questions such as how and why inferences constitute content in the first place, and why truth-

preservation in virtue of constitutive role constitutes justification. But perhaps there is more to say here, and if not, all explanations (especially explanations of normativity and intentionality) have to come to an end somewhere.

The circularity problem (Cummins, Fodor, Luzon). Intuitively, concepts are disposed to play the inferential roles that they do in virtue of their content. If so, it is hard to see how they can have their content in virtue of their inferential roles. These two "in virtue" claims together lead to a circle: content is grounded in inferential roles which are grounded in content. In a familiar move, I deny that where the constitutive C-inferences are concerned, inferential roles are grounded in content. Rather, content is grounded in inferential roles, which are grounded in more basic psychological facts. This may require rejecting an intuitive claim about inferences, but the intuition is far from non-negotiable. Note the difference with the case of normative inferentialism. It is hard to see how an inferential norm involving a concept could fail to be grounded in its content (this is a matter of what it is to be an inferential norm), whereas it's much less hard to see how an inferential disposition involving a concept could fail to be caused or grounded by its content (this is merely a folk-psychological intuition about mental causation).

Deference (Burge). In cases of semantic deference to a linguistic community, subjects possess concepts such as *arthritis* without being disposed to make the inferences that are constitutive of its content. In response, I've restricted my claims to cases of concept mastery (which is not present in these cases) rather than mere concept possession. One could also extend the current framework to cases of deference by arguing that these always involve deferential dispositions: for example, dispositions to make an a priori inference from "x has arthritis" to something like "x has the disease called 'arthritis' in this community" and vice versa. That will ground a primary intension for the speaker's use of 'arthritis' that picks out its extension in a scenario partly in virtue of how language is used in that scenario.

Understanding-assent links (Williamson). Williamson argues against these links by arguing that speakers can possess concepts such as conjunction without being disposed to make the inferences often regarded as constitutive of the concept. Part of Williamson's case rests on social externalism. Here I would invoke the discussion of deference above. If the idea is that these cases involve concept mastery without the relevant dispositions, one can reasonably question whether these cases involve genuine mastery of the concept of conjunction, as opposed to some other concept. There are other cases that rest less on

social externalism: notably, nonstandard logicians who arguably have mastered a logical concept while rejecting standard inferences. I'm inclined to say that even where these logicians reject these inferences in their theories, then if they have genuinely mastered the concepts, they will still be disposed to make these inferences at the level of thought. In addition, I note as before that my main thesis involves a sufficient condition for concept mastery, so a counterexample would really require a case where the dispositions are present but the concept is not mastered.

Defective concepts (Prior). What about concepts such as *tonk*, whose constitutive inferential roles are A ⊢ A *tonk* B and A *tonk* B ⊢ B? I don't think there's any such concept, or at least I don't think anyone can possess it. No one can have a concept with these a priori inferential roles, since if they could then an inference from A to B could be a priori, which it can't be. On my view, the constitutive inferential roles for a concept are constrained to be a priori and therefore cannot be explosive in this way. If so, there is no danger of anyone having a concept with these roles to cause trouble.

Variability. Users of names such as "London" and perhaps natural-kind terms such as "water" may use them with very different inferential roles, as brought by Kripke's case of Pierre who has words "London" and "Londres" that he does not connect. Do these all have different contents? Yes, Pierre has two concepts that are not a priori connected, so they have different primary intensions. In general, although I don't hold the extreme holist position that every inference affects the content of the concepts involved (only a priori inferences do), I think it is common for different users of a name (and of other words) to use it with different primary intensions. This is still compatible with communication as long as the intensions are not too different and share a referent.

Presupposing intentionality. Intentionality is presupposed at various places in this account: in the very notion of an inference, as well as in notions such as judgment and reference. I concede this point. Even if everything above works, it is not yet remotely a full naturalization of intentionality. It would ground the content of a concept in terms of factors that don't presuppose that very content (pending the issue below), but a lot more work would be required to turn this into a full naturalization of intentionality. And given my attitude to the first tier, such a naturalization may be impossible in any case.

In my view, the most serious objection to dispositional inferentialism is the problem of delineating the content-constitutive inferences in

noncircular terms, as discussed at the end of the previous section. One aspect of this problem is finding a descriptive counterpart for the normative notion of a priori inference or the semantic notion of analytic inference. Another is that we need a class of inferences constrained enough to only let in a limited class of inferences, but broad enough that subjects who have mastered a concept are actually disposed to make the relevant class of inferences at least some of the time. Requiring only a sufficient condition helps with the second problem, but if the condition is almost never satisfied in actual subjects, much more work will be required until we have an account that applies to to real thinkers. For these reasons, I am far from entirely confident that the dispositional inferentialist can succeed. But I think it is a project worth pursuing.

GROUNDING BASIC CONCEPTS

To naturalize intentionality in a two-tiered inferentialist framework, we need to naturalize inferential roles and naturalize the content of the basic concepts. I've talked about the roles already. Now I'll focus on the basic concepts. These aren't my main focus in this paper, in part because I'm unsure what to say about them and in part because I don't have anything new to say here beyond what I said in *Constructing the World*, but I'll put what I see as the main options on the table.

First, which basic concepts are needed? In *Constructing the World*, my initial scrutability base PQTI includes microphysical, phenomenal, indexical, and totality concepts. That's not a final first-tier base, for a number of reasons. First, some of the concepts can be analyzed further and may be scrutable from a smaller base. Second, one may need to expand the base to accommodate generalized scrutability and analytic (as opposed to a priori) scrutability. Third, some concepts in even a minimal scrutability base (conjunction, for example) may be grounded in inferential roles.

In chapter 8 of *Constructing the World*, I consider the nature of a fully adequate base in more detail. Microphysical concepts are plausibly analyzable by the Ramsey method, perhaps ultimately in terms of mathematical, nomic, and observational concepts. The relevant observational concepts arguably come down to spatial, temporal, and secondary quality concepts and/or phenomenal concepts. Totality concepts can arguably be analyzed in terms of fundamentality. Even phenomenal concepts can arguably be analyzed in terms of awareness of certain primitive qualities such as Edenic color qualities and spatial

properties. These qualities may also be needed to characterize other epistemically possible scenarios where they are instantiated.

There are many options here depending on one's philosophical views, but leading candidates for basic concepts in a minimal scrutability base may include the following.

Logical: e.g., *not*, *there exists*
Mathematical: e.g., *set*, *number*, *successor*
Phenomenal: e.g., *conscious*
Spatiotemporal: e.g., *space*, *time*
Secondary qualities: e.g., *color*
Nomic: e.g., *law*
Fundamentality: e.g., *in virtue of*
Normative/evaluative: e.g., *ought*, *good*
Indexical: e.g., *I*, *now*, *this*

Regarding normative/evaluative expressions: it is arguable that normative truths are always a priori scrutable from non-normative truths, in which case normative concepts are not needed in an a priori scrutability base. However, it is plausible that in this case the relevant a priori entailments are synthetic a priori, in which case normative concepts may still be conceptually basic. For the purpose of isolating conceptually basic concepts, we may want to focus on an analytic scrutability base, which may be somewhat more expansive than an a priori scrutability base given that analyticity entails apriority but not vice versa. In particular, if various mathematical and normative truths are synthetic a priori, it may be that some such truths are needed in an analytic scrutability base, with a corresponding role for mathematical and normative concepts. This base will be especially relevant if we want to individuate concepts by their analytic inferential role rather than by their a priori inferential role, yielding a somewhat more fine-grained analysis of meaning and content.

What grounds the contents of these concepts in a minimal scrutability base? In *Constructing the World*, I argued that the basic concepts can be divided into *acquaintance* concepts such as perceptual concepts, phenomenal concepts, indexical concepts, and perhaps spatiotemporal concepts, and *structural* concepts, such as logical and mathematical concepts, nomic and causal concepts, fundamentality, and perhaps evaluative and normative concepts.

What grounds the content of acquaintance concepts? For many of them, the obvious first move is to ground them in terms of *perceptual* contents.

The content of one's RED concept may be grounded in experience of redness. This is an especially natural view for an inferentialist, since the corresponding transitions from perception to thought (taking an experience of a certain shade at face value, for example) may be regarded as a priori justified and as constitutive of perceptual concepts. Color concepts and spatiotemporal concepts can plausibly be grounded this way. It is arguable that this is itself a sort of inferential grounding: for example, a perceptual concept RED_{31} may be grounded in a quasi-inferential connection to experiences of red_{31}, inheriting its content from those experiences by the constraint that face-value transitions from perception to thought should be content-preserving. Possibly something similar could be done for phenomenal and indexical concepts too.

This quasi-inferential grounding of perceptual concepts simply shifts the issue to grounding the relevant contents of perceptual experience, which are now in the first tier of our inferentialist framework. There are various options here. If one is aiming for a naturalistic reduction, one might pursue a causal or teleological analysis, such as a Neander-style analysis at least of perceptual concepts and spatiotemporal concepts and possibly phenomenal concepts.

For my part, I am doubtful about any fully naturalistic reduction here. On my view, the basic concepts include phenomenal concepts and Edenic concepts (Edenic space, time, color, and so on) that are hard to naturalize. As I discuss in *Constructing the World* and elsewhere, these concepts may be grounded in a sort of Russellian acquaintance relation with their referents: phenomenal properties and Edenic properties. One can also make a case that basic indexical concepts rest on acquaintance with the self and with one's experiences.[5]

Can acquaintance itself be naturalized? I am somewhat skeptical, for reasons tied to the difficulty of naturalizing consciousness. There is certainly a reasonable naturalist program of trying to ground acquaintance in causal, teleological, or inferential terms. But if this does not work, as I suspect, then acquaintance may serve as a basic sort of intentionality.

What about structural concepts? Here there are two main options. First, one could argue that they are also grounded in acquaintance. Perhaps we are acquainted with causation and value in experience, or we have a sort of intellectual acquaintance with them and with logical concepts, nomic concepts, and fundamentality. I take this view seriously.

Second, these concepts have distinctive inferential roles, and one could argue that their content is grounded in these roles. This is most

straightforward for the famous case of conjunction. But one can also find characteristic inferential roles for the other structural concepts, and then identify those concepts as representations that play these roles. Unlike second-tier concepts, these inferential roles won't involve a specific class of other concepts. Instead they are structural inferential roles. This approach will be subject to some of the standard problems for inferentialism, but perhaps the C-inference approach used for a dispositional approach to second-tier concepts could also be deployed here.

All this would take much further work to analyze properly. For now, my own view is that first-tier concepts may be grounded in a combination of acquaintance and inferential role. If we grant that second-tier concepts are grounded in basic concepts along with inferential role, then it follows that all concepts are grounded in acquaintance and inferential role. Those who want to pursue full naturalization may wish to further ground acquaintance in natural terms. For my part, a partial naturalization that takes acquaintance as primitive is naturalization enough.

NOTES

1. Based on the Sanders Lecture at the APA Pacific Division, April 10, 2021. Thanks to audiences at the Sanders Lecture, the NYU seminar on Mind and Language, and the London philosophy of mind group, as well as Ned Block, Tim Button, Anandi Hattiangadi, and Adriana Renero.

2. In what follows I sloppily sometimes use "concept" for abstract objects (contents) and sometimes for concrete objects (mental representations). I hope that context disambiguates!

3. Things work most easily if we assume the thesis of super-rigid scrutability from *Constructing the World*, where there is a scrutability base consisting of super-rigid and indexical concepts alone. This helps to avoid Twin Earth cases and Frege cases for basic concepts that complicate the analysis.

4. At this point, many inferentialist frameworks assign contents in a way that requires only that the inferences are truth-preserving. Intuitively, the core inferences should be stronger than that, being analytically or a priori truth-preserving. The 2D framework allows us to invoke truth-preservation at all scenarios, handling this stronger requirement perfectly. It's harder to see how other semantic frameworks will handle the stronger requirement.

5. See Pautz's article for reasons for rejecting reductionism for perceptual content and for embracing a form of acquaintance.

REFERENCES

Block, Ned. "Advertisement for a Semantics for Psychology." *Midwest Studies in Philosophy* 10, no. 1 (1986): 615–78.

Brandom, Robert B. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994.

———. *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press, 2000.

Burge, T. "Individualism and the Mental." *Midwest Studies in Philosophy* 4, no. 1 (1979): 73–122.

Carnap, R. *The Logical Syntax of Language*. Routledge, 1937.

———. "Meaning and Synonymy in Natural Languages." *Philosophical Studies* 6 (1955): 33–47.

Chalmers, D. J. "Epistemic Two-Dimensional Semantics." *Philosophical Studies* 118 (2004): 153–226.

———. "Revisability and Conceptual Change in 'Two Dogmas of Empiricism'." *Journal of Philosophy* 108, no. 8 (2011): 387–415.

———. *Constructing the World*. Oxford: Oxford University Press, 2012.

Cummins, Robert. "Conceptual Role Semantics and the Explanatory Role of Content." *Philosophical Studies* 65, no. 1-2 (1992): 103–27.

Davidson, Donald. "Radical Interpretation." *Dialectica* 27, no. 1 (1973): 314–28.

Dennett, Daniel C. *The Intentional Stance*. Cambridge, MA: The MIT Press, 1987.

Dretske, Fred I. *Knowledge and the Flow of Information*. Cambridge, MA: The MIT Press, 1981.

Field, Hartry H. "Logic, Meaning, and Conceptual Role." *Journal of Philosophy* 74, no. 7 (1977): 378–409.

Fodor, J. A. "Semantics, Wisconsin Style." *Synthese* 59 (1984): 231–50.

———. "Having Concepts: A Brief Refutation of the Twentieth Century." *Mind and Language* 19 (2004): 29–47.

Harman, Gilbert. "Conceptual Role Semantics." *Notre Dame Journal of Formal Logic* 28 (April 1982): 242–56.

———. "(Nonsolipsistic) Conceptual Role Semantics." In *New Directions in Semantics*, edited by Ernest LePore, 55–81. London: Academic Press, 1987.

Horgan, Terence, and John Tienson. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by David J. Chalmers. Oxford: Oxford University Press, 2002.

Horwich, P. *Meaning*. Oxford: Oxford University Press, 1998.

Jackson, F. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press, 1998.

Kripke, S. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press, 1982.

Lewis, David. "Radical Interpretation." *Synthese* 27 (July-August 1974): 331–44.

Loar, B. "Phenomenal Intentionality as the Basis of Mental Content." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by Martin Hahn and B. Ramberg Cambridge, MA: The MIT Press, 2003.

Luzon, Bar. "The Euthyphro Challenge in Metasemantics." Manuscript.

Mendelovici, Angela. *The Phenomenal Basis of Intentionality*. Oxford: Oxford University Press, 2018.

Millikan, Ruth Garrett. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: The MIT Press, 1984.

Neander, Karen. *A Mark of the Mental: A Defence of Informational Teleosemantics*. Cambridge, MA: The MIT Press, 2017.

Papineau, D. "Representation and Explanation." *Philosophy of Science* 51, no. 4 (1984): 550–72.

Pautz, A. "Consciousness Meets Lewisian Interpretation Theory: A Multistage Account of Intentionality." In *Oxford Studies in Philosophy of Mind*, edited by Uriah Kriegel. Oxford: Oxford University Press, 2021.

Peacocke, Christopher. *A Study of Concepts*. Cambridge, MA: The MIT Press, 1992.

Pitt, D. "The Phenomenology of Cognition, or, What Is It Like to Think that P?" *Philosophy and Phenomenological Research* 69 (2004): 1–36.

Prior, A. N. "The Runabout Inference-ticket." *Analysis* 21, no. 2 (1960): 38-39.

Quine, W. V. *Word and Object*. Cambridge, MA: The MIT Press, 1960.

Sellars, Wilfrid. "Inference and Meaning." *Mind* 62, no. 247 (1953): 313–38.

Shea, N. *Representation in Cognitive Science*. Oxford: Oxford University Press, 2018.

Siewert, C. *The Significance of Consciousness*. Cambridge, MA: The MIT Press, 1998.

Stampe, Dennis W. "Towards a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2, no. 1 (1977): 42–63.

Williams, J. R. G. *The Metaphysics of Representation*. Oxford: Oxford University Press, 2020.

Williamson, Timothy. "Understanding and Inference." *Aristotelian Society Supplementary Volume* 77, no. 1 (2003): 249–93.