

4

Moderate Intuitionism: A Metasemantic Account

Michael Johnson and Jennifer Nado

4.1. Introduction

Intuitions have for many years been considered indispensable to philosophical methodology. Recently, however, a growing body of empirical work has indicated that intuitions may be subject to various sorts of undesirable variation. These findings strongly suggest that philosophers have substantially overestimated the epistemological worth of intuition; it has even been suggested that intuition must be excluded from philosophical practice. Nonetheless, given that there's been no dramatic revolution in philosophical method, most philosophers seem to be hesitant to relegate intuition to the dustbin entirely. There is a strong temptation to say that intuitions must have at least *some* evidential weight—though they obviously occasionally go astray. Given their ubiquity in reasoning (philosophical and otherwise), it's difficult to accept the idea that our intuitions could be so unreliable that they'd have to be wholly abandoned as an evidential source. Therefore, in this chapter, we'll explore the potential for a 'moderate' account of intuition.

Moderate intuitionism (as we'll call it) recognizes that intuitions are generally reliable, but also frequently in error on certain classes of cases. On the methodological side, it allows that some revision of our philosophical practices might be in order, while stopping short of a complete rejection of an arguably central philosophical tool. Of course, moderate intuitionism can't be defended solely by an appeal to its pleasant consequences; what's needed is an account that explains why a moderate stance is appropriate. That is to say, there should be a theory that provides an explanation of the genuine yet somewhat fragile connection to the truth that moderate intuitionism aims to ascribe to intuitions. Our goal in this chapter is to outline such a theory.

A plausible place to start looking for the needed theory is in language. Our tendency to assume that intuition must be generally reliable seems linked to the fact

that it is impossible for our intuitions about *the meanings of words* to be substantially misleading. It's not at all plausible to suppose, for example, that 'dog' might (in the epistemic sense) turn out to refer to cats. This has led some philosophers to suspect that there exists some kind of deep connection between intuitions and meanings.

Unfortunately, however, the most popular existing metasemantic accounts don't seem to give us any kind of explanation for this apparent link—a fact that is particularly odd, given that philosophers tend to defend their preferred metasemantic account on the basis of intuitions. Thus, while intuitions regarding the reference of names in certain circumstances are standardly employed to defend Kripkean causal-historical accounts, such accounts don't themselves provide us with any particular reason to suppose that we would have intuitions in accord with them. Nothing internal to the account explains why we should have causal-historical intuitions rather than, say, descriptivist ones.

We propose to develop an outline for a metasemantic account which ties facts about meanings to dispositions to apply words when in possession of complete information. As we'll show, an account of this type both fits our metasemantic intuitions, and predicts a link between intuition and meaning that could underwrite the former's reliability. At the same time, the account also predicts that intuitions will fail under certain conditions. This fits pretty well with how the empirical evidence on intuition seems to be turning out. We take this general fit between our account and the evidence to provide support both for the metasemantic account itself (or some other account along the same lines), as well as for the moderate approach to intuition.

4.2. Moderate Intuitionism: Motivations

The nature of intuition itself is notoriously difficult to pin down. Rather than attempting to defend any particular analysis of intuition, we will just stipulate that by 'intuition' we mean to refer at the very least to spontaneous, not-obviously-inferential judgments such as those that frequently occur in response to thought experiments.¹ Our primary goal is to defend a moderate stance on the epistemological status of the sorts of things philosophers are inclined to call 'intuitions'; as far as we can tell, our argument is compatible with a fairly broad range of particular, considered views on the nature of intuition. As we'll discuss later, it is even compatible with the claim that there is no unified mental state-type underlying uses of the term 'intuition'.

The empirical arguments that have been recently marshaled against intuitions are well known, so we'll only briefly rehearse them here. In short, intuitions vary—both interpersonally as well as intrapersonally. Evidence suggests that intuitions vary across cultures (Weinberg et al. 2001; Machery et al. 2004), genders (Buckwalter and Stich

¹ We characterize intuitions as judgments here. Others may prefer inclinations to judge, or seemings that can lead to judgments. Nothing in what follows hangs on this issue.

2014), and socioeconomic groups (Haidt et al. 1993; Weinberg et al. 2001); and that they are sensitive to emotional states (Wheatley and Haidt 2005; Schnall et al. 2008), order of presentation (Swain et al. 2008), and more. Since these factors are plausibly irrelevant to the phenomena the intuitions are meant to describe—that is, e.g., the facts about knowledge don't vary as a function of the cultural background of the attributor—this variation suggests that intuition is not wholly tracking the truth. Thus, variation data has been taken to show that intuitions are unreliable, and that they are therefore unsuited for use in philosophical theorizing.

Of course, this last step is open to serious question. If the findings are robust, they show (barring relativism) that at least some intuitions are in error. But, it may be argued, this merely demonstrates that intuition is fallible—a fact that few philosophers would deny. As Ernest Sosa has noted, perception is also susceptible to various interfering factors that cause it to occasionally go awry, but perception is nonetheless a quintessentially respectable evidential source. “The upshot is that we have to be careful in how we use intuition, not that intuition is useless” (Sosa 2007, p. 105). Intuition isn't perfect, but that's no reason to think that it should never be employed in philosophical argumentation.

In fact, there are reasons to suspect that intuition can't be totally rejected as an evidential source without invoking radical skepticism, both with regard to philosophical inquiry as well as generally. As George Bealer (1992) and Joel Pust (2000) have noted, it seems that even arguments *against* intuition invoke intuitions—for example, intuitions that beliefs formed by unreliable processes are unjustified. Even worse for the anti-intuitionist, Timothy Williamson (2007) has argued persuasively that there is simply no clear distinction between 'philosophical' intuitions and everyday cases of concept application. If this is right, however, arguments against the reliability of intuition threaten to generalize to arguments against the reliability of concept application—a remarkably skeptical position.

More generally, it's patent that intuitions have at least some tie to the truth. The possibility that all or even most of our intuitions are mistaken does not warrant serious consideration. No one is worried about the possibility that murdering innocents for pleasure is in fact the *summum bonum*; nor is there any reason to believe that the term 'consciousness' in fact correctly applies to all objects within 100 meters of the Eiffel Tower. In a less silly vein, there appears to be little variation on intuitions regarding certain central cases of philosophical categories like knowledge. Almost all epistemological theories agree in dismissing random guessing as a source of knowledge—and empirical evidence from Weinberg et al. (2001) suggests that there is cross-cultural agreement on such cases as well.

For several reasons, then, the most attractive stance on intuition seems to be that it is significantly less epistemologically respectable than the most optimistic among us might have hoped, but still broadly reliable and capable of providing evidence for or against certain philosophical hypotheses. As noted in the Introduction, we will call this position 'moderate intuitionism.'

4.3. Moderate Intuitionism: In Search of Theoretical Explanation

Moderate intuitionism is attractive, but it needs explanatory support. Why *should* there be a link between intuition and truth at all? In the case of perception, we have a fairly thorough theoretical understanding of why beliefs based, e.g., on vision should be more or less reliable. We can tell a story about the causal mechanisms by which looking at dogs produces the belief that there are dogs. The story begins with light reflected from dogs hitting cones on the retina, and continues with edge detection, the segmentation of the scene into objects and their parts, and so on. The details of the story also explain how vision sometimes produces beliefs about dogs in situations where there are none. Given the fact that illusory dogs arise only rarely, the story predicts that visually-based dog beliefs will be generally reliable. Of course, the story is not complete—there are gaps in our understanding, e.g., at the level of consciousness. But, importantly, there's an account which explains more or less why—and how—the presence of dogs tends to lead to true dog beliefs.

Our understanding of the means by which vision produces true beliefs is thus reasonably good. In other cases, as with memory, we do not yet have as many details, but we can conceive of the general form an account would take. With intuition, however, the situation is otherwise. It's not at all obvious how to explain the link between, for example, knowledge and our intuitions about knowledge. Indeed, this ignorance has itself occasionally been fodder for the anti-intuitionist argument, as when Jonathan Weinberg notes that the “hopefulness” of intuition as an evidential source is undermined by the fact that “We just do not seem to know much about the underlying psychology of the propositional seemings that we term ‘intuitions’” (Weinberg 2007, p. 336). Tellingly, the most persuasive defenses of intuition—such as those of Sosa, Bealer, Pust, and Williamson mentioned above—are in a sense ‘negative’. They argue against anti-intuitionism, rather than in favor of intuitionism; they simply advert to the practical impossibility of abandoning intuition, or to the fact that such a move is unwarranted given the evidence at hand. Positive defenses of intuition, based on explanations of how intuitions actually produce true beliefs, are occasionally offered—however, in many cases, the explanations are ultimately unsatisfying.

When defenders of intuition do attempt to provide positive characterizations of the nature of the intuitive process, they frequently employ somewhat metaphorical explanations in terms of ‘perceiving’ or ‘grasping’ the truth of a proposition. For instance, Laurence BonJour claims that, when one has an intuition, one is “able to see or grasp or apprehend in a seemingly direct and unmediated way that the claim in question cannot fail to be true” (BonJour 1998, p. 101). Another popular term is ‘understanding’—Sosa’s view is that “we manifest a competence that enables us to get it right on a certain subject matter, by basing our beliefs on the sheer understanding of their contents” (Sosa 2007, p. 102). Standardly, an intimate link with ‘reason’ is supposed to be involved—terms such as ‘rational insight’ (BonJour) and ‘rational seeming’ (Bealer) have been offered to

describe the intuitive process. In all the above cases, however, the phraseology employed does not give us much in the way of explanation of the particular mechanisms by which such capacities supposedly operate. *How* does one come to grasp or understand the truth? What process underlies rational seeming, such that those seemings tend to be true?

The following is a common and potentially promising idea for unpacking the above explanations: intuitions tend to be true because the reliability of intuition is *guaranteed* by the existence of some sort of *constitutive relationship* between intuitions on the one hand and meanings or concepts on the other. Alvin Goldman, for instance, has argued that “it’s part of the nature of concepts . . . that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept” (Goldman 2007, p. 15). Similarly, Bealer writes that intuitions possess a “strong modal tie to truth, [which] is simply a consequence of determinately possessing the concepts involved” in the intuition (Bealer 1996, p. 2). And Frank Jackson claims that “the business of consulting intuitions about possible cases is simply part of the overall business of elucidating concepts” (Jackson 1998, p. 33). The idea, then, is that intuitions are intimately tied up with concepts in a way that guarantees that intuitions will reflect truths involving those concepts. The existence of a constitutive link of this sort, if adequately explained, could provide the theoretical support for moderate intuitionism that we’re seeking.

The basic idea of a constitutive tie between intuitions and concepts can, however, be expanded on in several ways. On Goldman’s view, for instance, “possessing a concept makes one disposed to have pro-intuitions toward correct applications and con-intuitions toward incorrect applications—correct, that is, relative to the contents of the concept as it exists in the subject’s head” (Goldman 2007, p. 15). In other words, intuition reveals psychological facts: your intuition that *x* is a case of knowledge is evidence that *x* falls under *your* knowledge concept, where ‘concept’ is used in a narrow psychological sense to indicate a particular mental representation which “is fixed by what’s in its owner’s head” (Goldman 2007, p. 13).

Although the truth of Goldman’s account would guarantee that intuitions are evidence regarding psychological facts, it would not automatically guarantee that intuitions are themselves generally true or generally lead to true beliefs.² This would be a separate, further claim; one which Goldman does not make. It is compatible with Goldman’s account to claim that an intuition of the form ‘the Gettier case is not a case of knowledge’ provides evidence for a certain psychological fact, while failing to provide evidence *that the Gettier case is not a case of knowledge*. Whether or not this failure arises depends on our account of the truth conditions of intuitions, and whether those truth conditions are determined by something other than the psychological concept that produces the intuition. So at the very least, Goldman would need to make supplementary claims in order to defend a move from ‘intuitions are evidence for facts about

² Whether or not it makes sense to call an intuition ‘true’ will depend on one’s particular view of intuitions. If intuitions are beliefs, then they can be true or false. If they are inclinations to believe, or intellectual seemings, or something similar, it may make more sense to merely say that they produce true beliefs without being themselves truth-evaluable.

one's personal concept of knowledge' to 'intuitions are evidence for facts about knowledge'. Because moderate intuitionism is a claim that intuitions are generally reliable *tout court*, rather than just reliable indicators of the nature of personal psychological concepts, we'll leave Goldman's account to the side.³

Bealer's version of the constitutivity approach, on the other hand, is non-psychological. Bealer has suggested that intuitions possess a 'strong modal tie' to the truth—necessarily, intuitions are true most of the time, or under normal circumstances. This strong modal tie is itself cashed out in terms of 'determinate concept possession', which is the sort of concept possession one has when one possesses a concept without 'misunderstanding or incomplete understanding' (Bealer 2000, p. 11); however, Bealer clearly means to use 'concept' to express something graspable by multiple persons, rather than something individual and purely psychological. Unfortunately, Bealer provides no explicit argument for the claim that humans in fact ever attain such determinate concept possession; instead, he tends to rest with the claim that there is no intrinsic barrier to achievement of determinate concept possession.

It's clear, however, that on Bealer's account, determinate concept possession is made possible at least in part by the fact that philosophical terms, unlike natural kind terms, are 'semantically stable'—that is, knowledge of their conditions of application does not require any contingent knowledge about the speaker's external environment. Conversely, "an expression is semantically unstable iff the external environment makes some contribution to its meaning" (Bealer 1996, p. 23). Thus, the constitutive tie between intuitions and concepts seems to involve a claim about the meanings of terms. If semantic stability is a feature that enables determinate concept possession, and thereby a strong modal tie to the truth, this suggests that our intuitions about philosophical terms and their applications are reliable in virtue of certain *metasemantic facts*. Though we need not be committed to Bealer's particular account of the constitutive link between intuition and truth, the idea that said link involves metasemantic facts is promising. Let's pursue it.

4.4. Metasemantics and Constitutivity

As mentioned above, one way that intuitions might have a constitutive tie to the truth is via their relation to metasemantics. *Metasemantics* is the theory of why expressions mean what they do (i.e. *in virtue of what* do they mean what they do?), rather than something else, or nothing at all. Now suppose, presumably contrary to fact, that the correct metasemantic view is one according to which the facts about meaning are fully determined by intuitions. If the connection were particularly direct, in a sense to be explained presently, then there would be a neat story to tell explaining the reliability of intuition.

³ Goldman offers his account as a way to defend the idea that intuitions can provide evidence for philosophical claims. As part of the account, he proposes that understanding concepts in the personal psychological sense is the primary goal of philosophical inquiry. We disagree with him on this point, but will not argue it here.

The crudest version of such a metasemantic view might look something like this:

Crude Metasemantic Constitutivity (CMC): Term T applies to object O in the mouth of speaker S in virtue of the fact that S intuits that T applies to O .

If CMC were true, semantic application intuitions—that is, intuitions that a term T applies to object O —would be a sort of self-fulfilling prophecy. Having the intuitions makes the intuitions true, because the intuitions themselves *ground* the meaning facts that the intuitions are about.

This model could accommodate more than just intuitions about the application of terms, under a reasonable assumption.

Assumption: For each T in S 's vocabulary, then if there is a property P that T expresses, P is such that:

- (i) An object⁴ O is P iff T applies to O .
- (ii) S intuits that O is P iff S intuits that T applies to O .⁵

Let's take an example. If I intuit that case C is an instance of knowledge, then by (ii) I intuit that 'knowledge' applies to C . So, by CMC, 'knowledge' does apply to C . By (i) if 'knowledge' applies to C , then C is knowledge, and my (non-semantic) intuition that C is knowledge is true. Thus, though CMC is a metasemantic claim, its truth would underwrite reliability for intuitions that are not explicitly linguistic.

What this suggests is that a close tie between intuitions and meanings, through a metasemantic theory that connects them, as given in CMC, would be a promising and powerful tool for connecting intuitions with the truth of their contents. There are, however, at least three reasons to think that nothing quite *this* crude is going on. First, as we've pointed out, intuitions vary under circumstances that don't plausibly alter the truth of the facts at issue (e.g. the order in which cases are presented to subjects to elicit their intuitions). CMC, however, doesn't admit *any* separation between intuition and truth; according to CMC as stated, there's no possibility of error in intuition (granting the Assumption). At the very least, a less crude account would need to require that speakers were attentive, or appropriately trained, or something similar before their intuitions 'count' as determining meaning.

Second, and more seriously, there's reason to think that the crude model given in CMC is altogether too internalist. Our hesitation to commit to a fully internalist conception of meaning is not merely based on intuition, but on general theoretical grounds. The worry is well illustrated by a passage from Jerry Fodor:

... words can't have their meanings *just* because their users undertake to pursue some or other linguistic policies; or, indeed, because of any purely *mental* phenomenon, anything that happens purely 'in your head.' Your undertaking to call John 'John' doesn't, all by itself, make 'John'

⁴ Or event, or what have you.

⁵ An example: For $T =$ 'water', there is a property $P =$ *being water*, such that when you intuit something is water, you intuit that 'water' applies to it, and when 'water' does in fact apply to it, it is water.

a name of John. How could it? For ‘John’ to be John’s name, there must be some sort of *real relation* between the name and its bearer; and intentions don’t, per se, establish real relations. This is because, of course, intentions are (merely) intentional; you can intend that there be a certain relation between ‘John’ and John and yet there may be no such relation. A fortiori, you can intend that there be a semantical relation . . . and yet there may be no such relation. (1990, p. 98)

If, like us, you accept Fodor’s claim, then application intuitions can’t be self-fulfilling in the way described by CMC, because something purely mental (*viz.*, my intuition that *T* applies to *O*) cannot in itself bring about a real relation (*T*’s actually applying to *O*).

Finally, though somewhat paradoxically, there are well-known intuitive reasons for thinking that intuitions just don’t bear *that* close a tie to meanings. That is, several thought experiments and related actual cases (Kripke’s ‘Peano’ and ‘Gödel/Schmidt’ cases, Putnam’s Twin Earth) elicit externalist intuitions, and have been used to support externalist metasemantic theories wherein the purely mental has little say in the application-conditions of our terms. CMC is, of course, inconsistent with this externalist approach. So much, then, for CMC—we’ll have to look for a less direct link between intuition and meaning. Perhaps one can be found in the externalist theories just mentioned.

4.5. Externalist Metasemantics and Intuition

Consider the following prompt:

You’re of course familiar with the name ‘Gödel’. Likely, you believe that Gödel proved that arithmetic is incomplete. Suppose what I tell you now is true. There were two men, X and Y. X discovered a proof that arithmetic is incomplete—indeed, the same proof that frequently has the name ‘Gödel’ attached to it in articles and textbooks today. X, however, had the name ‘Schmidt’ but not ‘Gödel’ written on his birth certificate. Another man, our Y, killed X in secret and stole his proof before X had a chance to publish it. Y then published the proof under the name he (Y) had on his birth certificate, which happened to be ‘Gödel’. This is why the proof now frequently has the name ‘Gödel’ attached to it in articles and textbooks today.

Again, supposing all this to be true, what of your original belief? *Did Gödel prove that arithmetic is incomplete?*

A fair number of influential philosophers are inclined to answer the question negatively,⁶ and historically this has been taken as evidence that *the* intuition is a negative one: Gödel did not prove the incompleteness theorem, if the situation is as stated in the prompt. Further, intuitions on this case, on Hilary Putnam’s Twin Earth case (Putnam 1975), and so on, have historically been taken to provide strong evidence for certain metasemantic theories and against others.

⁶ Beginning with Saul Kripke (1972), who formulated it.

Let's give a quick example of how such arguments go. Most individuals who have heard of Gödel likely only have heard that he proved the incompleteness of arithmetic.⁷ On descriptivist accounts, a name refers to whatever satisfies the mentally represented description that the speaker associates with that name; thus, on a descriptivist account, 'the individual who proved the incompleteness of arithmetic' is one of the few candidate descriptions that can determine the meaning of 'Gödel'. However, this predicts that in the scenario described in the prompt, Gödel did in fact prove that arithmetic is incomplete—for it is a tautology that the individual who proved the incompleteness of arithmetic proved the incompleteness of arithmetic. The prediction of descriptivism is thus at odds with the intuition, and this is interpreted as evidence that descriptivism is false.⁸

Note that the intuitions on such cases aren't taken merely as crudely suggestive, but as fine-grained tools for getting at the metasemantic reality. For example, Gareth Evans (1973) introduces several examples that are intended to show that Saul Kripke's suggested model can't be right, even though it comports with our intuitions most of the time. One such case involves the reference of 'Madagascar'. On Kripke's account, reference is determined by an initial baptism, and maintained by causal links between speakers through which reference is transmitted. However, the landmass that was initially baptized 'Madagascar' is in fact a portion of the African mainland, and thus Kripke's account inaccurately predicts that 'Madagascar' refers to that region. The attention paid to such counterexamples suggests that getting the intuitions right 'most of the time' isn't enough; you have to get *all* the intuitions right, or at least all the ones that seem fairly systematic. But why should any of this be so? Why think that metasemantic intuitions of this sort are reliable guides to metasemantic facts?

Normally in science, we use theories to generate predictions (with the aid of certain background assumptions, etc.), and we design experiments to test those predictions; the results of those experiments can then potentially confirm or disconfirm the theory. But consider the application of this idea to externalist metasemantic theories. Kripke's theory of reference—the view that name *N* refers to the object *O* that stands at the end of a certain historical chain beginning with a baptism event and successive *N*-inheritance from speaker to speaker—does not at all predict that anyone will have intuitions that accord with it. There's nothing contradictory about Kripke looking at Evans' 'Madagascar' case and saying "No, 'Madagascar' refers to the portion of

⁷ As a reviewer points out, many will only have heard that he is a logician: of course, this is even more problematic for a descriptivist.

⁸ Even defenders of descriptivism have accepted the evidential force of the intuition. For example, various philosophers have attempted to outline a version of descriptivism that makes predictions in accordance with these and other thought experiments in the literature (e.g. Katz 1994). Such a version might say that the description associated with 'Gödel' that determines the latter's referent is 'the individual upon whose birth certificate appears the name "Gödel"'. In this case, Gödel would not have proven the incompleteness of arithmetic in the scenario described, since the individual upon whose birth certificate appears the name 'Gödel' in the scenario did not prove the incompleteness of arithmetic in the scenario. It's worth noting that such a view is clearly designed to comport with the intuitions.

the mainland baptized as such. That's what my theory says.⁹ *What's intuition got to do with it?*" So the fact that our metasemantic intuitions do or don't accord with Kripkean theory seems irrelevant to whether or not we have confirmation of it. The argument from intuition to Kripkeanism seems to be missing a crucial step—one which links the theory with the intuitions.

It is, in fact, not obvious how such a link might be provided in the absence of some kind of constitutivity hypothesis. The intuition in question involves a claim about the reference relation; there's been a temptation, therefore, to claim that the intuition reflects the subject's theory—we might even say her *concept*—of reference. Such a theory of reference might be either explicit, or tacit. The explicit option, of course, is unhelpful; if one's explicitly-held metasemantic theory is causing the intuition about which individual 'Gödel' refers to in the prompt, then the 'intuition' is just a statement of the consequence of one's philosophical theory. The other option involves the idea that individuals possess something like a tacit 'folk metasemantics' that allows them, when given a non-semantic description of a situation, to produce a judgment about what the semantic facts are. But the problem remains: why think that such judgments provide evidence about the metasemantic facts themselves?

Consider a parallel. People have natural intuitions about how non-animate objects will move when subjected to various forces. The tacit principles associated with these intuitions are known collectively as 'folk physics', and they appear to roughly resemble the principles of the (false, discredited) medieval impetus theory. Obviously, no physicist is of the mind that physics should strive to capture either the content of folk physics or the intuitions it generates. Physicists are concerned with scientific theories about how actual objects subject to actual forces behave, not psychological theories regarding how people *think* they will behave. Folk physics is irrelevant.

Much the same could arguably be said for semantics. Attributing meaning or reference to natural language expressions serves a scientific, predictive, explanatory purpose. The assigned contents have to meet certain constraints of which most individuals are not aware: they need to be objective enough to be communicable; context-independent enough to be compositional; fine-grained enough to capture certain logical relations, etc. So, if tacit folk metasemantics drives judgments on the Gödel case, this seems of little moment—why does folk metasemantics have anything more to do with metasemantics than folk physics has to do with physics? Metasemantics is an empirical theory about why expressions mean what they do, rather than something else or nothing at all; it is not a psychological theory about why people *think* things mean what they do.

In general, there's no obvious reason to think that 'folk theories' are a good guide to true theory. So we still have no reason to suppose that, for externalist metasemantic

⁹ Well, this would be somewhat unlikely, given that Kripke claims not to be offering a theory. For our purposes, we'll adopt the convenient fiction that Kripke's remarks on the metasemantics of names and natural kind terms constitute a theory that he endorses.

accounts, our intuitions about reference are anything less than independent of the facts about reference. As the case of folk physics shows intuitions can, quite generally, ‘float free’ from the facts. Perhaps evolutionary considerations suggest that the intuitions must line up to some extent with the facts, but again, as the case of folk physics shows, this requirement need not be very substantial.

If the folk metaseantics approach is to provide the link between intuition and theory, then one needs to claim that metaseantics is in some kind of epistemically privileged position. This is, in fact, *prima facie* tempting—for, while physics involves external phenomena, metaseantics is in a sense about *us*. The use of language is a human activity, and so it might seem that humans would have special insight into its workings. One might try to make a parallel with reliance on grammatical intuitions in linguistics;¹⁰ in linguistics, after all, it’s common to assume that there is an intimate tie between grammatical intuition and grammatical truth, one which simply arises from the nature of language (in other words, a constitutive tie). So the analogy with grammar might lead us to propose some sort of constitutive link—not directly between intuitions and meanings, but between folk metaseantics (which generates intuitions) and meaning.

However, with regard to the grammar analogy, a couple of points are in order. First, a close tie between represented grammar rules and grammar facts need not preclude a divide between grammar intuitions and grammar facts: multiple center embeddings (e.g. “A theory that a philosopher that a grant agency awarded money endorses is true”) seem unacceptable, but are in accord with English grammar (on standard accounts).¹¹ More importantly, though, it’s relatively clear why grammaticality should be determined by internal rules and representations—because its function is to explain, among other things, the learnability of languages. The fact that an infinite set of expressions is learnable from a finite, highly impoverished set of stimuli is explained by a severe, innate restriction on the class of potential grammars. If the principles restricting possible grammars were not internal to the mind, they couldn’t do the explanatory work required of them. These principles (with their learned parameters set) determine what is grammatical (but of course not what is parsable). Thus grammaticality is internal (if only partly accessible through intuition).

On the other hand, it’s far less plausible to suppose that what semantics is supposed to explain is a purely internal affair. Representation is a relation between us and the world. And as Fodor urged in the last section, such relations can’t hold “*just* because their users undertake to pursue some or other linguistic policies; or, indeed, because of any purely *mental* phenomenon, anything that happens purely ‘in your head’” (1990,

¹⁰ Cf. Stich 1996, pp. 41–2.

¹¹ The reason things work out like this in our linguistic theories is that grammaticality is a feature that’s supposed to explain the learnability of languages, and the evidence suggests that our language parsers (the generators of our intuitions of acceptability) only employ fallible heuristics for detecting grammaticality. (It’s easier and faster to parse when you allow yourself to get certain systematic classes of cases wrong.)

p. 98). Merely representing a folk semantics does not bring it to pass that meanings are determined by that folk semantics.

If this is right, then there's reason to doubt the hypothesis of a constitutive link between folk metasemantics and metasemantics itself. Worse, though, even the existence of such a link wouldn't really support standard causal metasemantic accounts—because, as we'll discuss presently, such accounts don't even capture the metasemantic intuitions that were supposed to motivate their acceptance.

4.6. From Causal to Dispositional Metasemantics

Supposing we did represent a tacit metasemantic theory of a broadly Kripkean sort, one would expect to find that our intuitions about reference track the predictions of Kripkean theory rather closely. But, in fact, they depart from such predictions in very significant ways. For example, it seems intuitive that descriptive terms or languages are possible. Suppose that English and Cantonese differ in the following respect: in English, name *N* refers to object *O* iff *N* is acquired through an anaphoric chain that terminates in an ostensive or descriptive baptism of *O* with *N*. In Cantonese, *N* refers to *O* iff *O* is the unique object that satisfies a description *D* that speakers associate with *N*. *It strikes us that this is possible*. Now, suppose that our intuitions are right and it is in fact possible. Then, the fact that name-referent pairs $\langle N, O \rangle$ in English stand in an anaphoric-baptism relation would be an interesting generalization about English, but not an explanation of why *N* bears the reference relation to *O*. Pairs $\langle N, O \rangle$ in other languages might stand in the selfsame relation, but nevertheless not be referentially related.

There are in fact plausible real-world examples of descriptive terms: for instance, the empty expression 'phlogiston'. Phlogiston was a substance that the alchemist and physician J. J. Becher implicated in the combustion and rusting of materials. Oxygen is crucial for both of these processes, and was arguably the substance with which Becher was confronted when introducing the term. Yet 'phlogiston' is universally taken not to apply to oxygen, but rather to have an empty extension—and even, according to Kripke, a necessarily empty extension. We seem to take the descriptions associated by Becher with 'phlogiston' to be constitutive of its nature: since oxygen is not released in combustion or in oxidation, phlogiston is not oxygen. It's apparently more important to us whether the substance Becher postulated had the properties he postulated for it than whether there is a single substance involved in the processes he identified. Our intuitions here run counter to the predictions of causal accounts.

Another well-known anti-Kripkean intuition: the original baptismal event can be irrelevant to the current meaning of an expression (even in cases where there's no intentional reference shift, as when I call my dog 'Aristotle'), as can the descriptions under which objects or kinds are baptized. In Evans' 'Madagascar' case, although a portion of the mainland of Africa is the baptized location that stands at the end of our 'Madagascar'-chain, and although at no point did anyone involved in the chain

intentionally shift the reference to Madagascar, still it is the island and not a portion of the mainland that is properly so-called now. Since where we'll be taken if we ask to go to 'Madagascar' is more important to us than what was originally called 'Madagascar', 'Madagascar' means the former, and not the latter.

Similarly, 'jade' was introduced presumably with about the same understanding with which 'water' and 'gold' were introduced: as a putative substance kind. Jadeite and nephrite have similar, though not identical, appearance, hardness, beauty, and economic value. When it was discovered in the nineteenth century that the two types of jade were separate minerals, it was open to us to reject 'jade' as a classifier or to use it to apply to just one of the minerals. However, it was more important to us to have a term that applied to minerals of a certain appearance and value than it was to cut solely at the joints of nature. Compare fool's gold, which we had no interest in classing with the element Au.

Can we capture this group of intuitions, as well as the externalist intuitions mentioned previously, with a single metasemantic account—while retaining the idea of a constitutive link? One possibility: an account that runs not through a single tacit, broadly Kripkean metasemantic theory, but instead through a variety of less unified represented rules. Let's look at an example. Keith Donnellan (1993) argues that the "force"¹² of our intuitions in Twin Earth cases is strong evidence that we are following a hard-to-consciously-access, internal-to-the-mind "semantic rule" that leads us to infer from the premise that something shares the underlying nature of the paradigm cases of water to the conclusion that it is water (pp. 157–8). Donnellan then urges something like the following constitutive connection between semantic rules and metasemantic facts:

Donnellan's Metasemantic Constitutivity (DMC): (For any T, R, O, S) if term T bears relation R to object O , then T applies to O in the mouth of speaker S *in virtue of the fact that* S follows an internal semantic rule to infer from the claim that T bears R to O to the claim that T applies to O .

Donnellan concludes that "there may be a sense in which what is 'in our heads' determines the extension of a term such as 'water'" (Donnellan 1993, p. 158). The idea here is that semantic rules (for example, the rule to infer from the claim that the liquid in lakes and rivers around here was baptized 'water' to the claim that 'water' applies to the aforementioned liquid) ground the metasemantic facts. Thus our intuitions in Twin Earth cases are reliable, because the semantic rule we follow that nothing lacking the underlying nature of the paradigm cases of water is water makes it true that no such thing is water. The rule at once *generates* the intuition and *grounds* the fact that the

¹² Curiously, Stephen Stich (1996, p. 47) argues from the fact that for many prompts, we *don't* have firm reference intuitions, to the conclusion that our internally represented metasemantic theories underdetermine many cases. The analogous conclusion here would be that although the force of Twin Earth intuitions might establish that we represent and follow a semantic rule for 'water', the lack of force in many other, similar cases shows that not every expression is governed by a semantic rule (or that some are governed by indeterminate rules).

intuition is about. This position solves the problem of how an ‘externalist’ metasemantic theory can be connected with intuition: the ‘externalist’ theory itself is grounded by an internal representation of it.

But notice, too, that DMC could at least in principle account for the anti-Kripkean intuitions just discussed. Descriptive terms and languages are possible according to DMC, because it’s possible for expressions to be governed by descriptive semantic rules. So if ‘phlogiston’ has a descriptive semantic rule governing it, we could explain why there isn’t any phlogiston; and if ‘jade’ is governed by a semantic rule, like “‘jade’ refers to the economically interesting kind encompassing such-and-so paradigm cases’, we could explain jade’s current disjunctive status; and similarly if ‘Madagascar’ is governed by a semantic rule like “‘Madagascar’ refers to the place you get taken when you ask to go to “Madagascar”’. So we have in DMC a constitutive account with the flexibility to account for any metasemantic intuition one happens to have, by postulating a represented rule that underlies that intuition.

However, we want to argue against this particular version of constitutivity for two reasons. First, as we’ll argue in section 4.7, it’s implausible that any of the terms just discussed are or were actually governed by the semantic rules just hypothesized. But more fundamentally, as we have already emphasized more than once, purely internal things like semantic rules don’t plausibly engender, by themselves, representation relations. What’s needed is something that at once has the externalist virtues of the anaphoric-baptismal account and the internalist, constitutivity-granting virtues of the semantic rules account.

Here’s our attempt at embodying these virtues, through a somewhat less direct approach to constitutivity. The proposed account,¹³ which we’ll call a ‘dispositional’ theory of reference, is as follows:

Dispositional Constitutivity (DC): A linguistic expression *E* means some object, property, kind, relation, etc., *X*, in the mouth of speaker *S*, in virtue of the fact that *S* would be disposed to apply *E* to *X* if *S* had all the relevant information.

‘Relevant information’ consists of the facts *F* that would, were *S* to be apprised of *F*, influence *S*’s dispositions to apply *E*.¹⁴ The inclusion of the ‘relevant information’ clause creates a gap between one’s current disposition to apply a term and that which

¹³ We would like to emphasize that this is *not* an attempt at a naturalization of intentionality: the account contains clearly intentional terms like ‘apply’ and ‘relevant information’ (defined in terms of ‘apprising’ *S*).

¹⁴ This may even include semantic facts; for instance, one’s disposition to apply the term ‘dog’ may be influenced by semantic facts about what the term ‘dog’ means in the mouths of experts. However, when determining that an expression *E* means *X* in the mouth of speaker *S*, one fact that should be excluded from the relevant information is the very fact being determined by that information, namely, ‘expression *E* means *X* in the mouth of speaker *S*’. We think this exclusion is motivated, since the fact in question is not ‘independently grounded’. When a set of facts grounds another fact, the latter fact cannot be in the grounding set; things do not ground themselves. Thus, the fact that expression *E* means *X* in the mouth of speaker *S* cannot be part of the set of facts, knowledge of which determines the dispositions that ground the meaning of *E* in the mouth of *S*.

determines the meaning of the term. Thus, the fact that you are disposed to apply ‘cow’ while viewing a horse on a dark night is not determinative of the meaning of ‘cow’; for, if you were apprised of certain facts, you would no longer be so disposed. Looking ahead, ‘relevant information’ often involves the sort of information provided in the course of a thought experiment—e.g., the information that the watery stuff on Twin Earth is not chemically identical to the watery stuff on Earth.

The basic motivation behind the dispositional proposal is this. Fodor argued that nothing purely mental could *by itself* establish semantic relations. But we need not consider the purely mental *by itself*: “linguistic policies don’t make semantic relations; but maybe they make causal relations, and maybe causal relations make semantic relations” (1990, p. 99). Maybe we do represent tacit metasemantic theories, or semantic rules, or whatever. Those things, *in themselves*, can’t determine what means what. But they might dispose us to behave in certain ways: accepting the semantic rule “‘water’ applies to the liquid in lakes and rivers around here’ might dispose us to withhold ‘water’ applications from substances we know to not be the liquid in lakes and rivers around here. Such a disposition isn’t purely mental: it’s a real relation between our application behavior and the extra-mental events of the extra-mental world. And according to the account on offer, this disposition (subject to the relevant information clause) establishes a semantic relation between ‘water’ and H_2O .

What we’ve just given is only the barest possible sketch of a metasemantic account. But we’re not particularly interested in defending an account of reference; instead, we’re interested in outlining a method for elucidating a constitutive account of the evidential status of intuition, without the need for an internalist metasemantics. What we will try to show here is that something *roughly* like the account on offer is true, and preferable to either a straight-up anaphoric-baptismal account, or to a semantic rule account like DMC. This is a good thing, because if it’s true, we’ll have an explanation that can underwrite moderate intuitionism. The principal idea, to be elaborated upon, is that our judgments (intuitions) regarding the application of terms given certain prompts are reliable but fallible indicators of our dispositions to apply terms in the circumstances described by those prompts. The latter dispositions are constitutively linked to the metasemantic facts¹⁵—to which the terms correctly apply—and thus our intuitions are reliable but fallible indicators of that to which our terms correctly apply.

4.7. Some Relevant Details

It’s important that we establish that the dispositional account just offered, or something roughly similar, is true, or roughly true. If it isn’t, then it can’t explain intuition’s

¹⁵ Only when, of course, the prompts specify enough relevant information—that is, when our dispositions when given the specified information are not substantially different from what our dispositions would be when given all relevant information. This will of course be a matter of degree, as will the strength of the constitutive link.

connection with the truth. And our goal here is to explain intuition's (fallible) connection with the truth. That being our goal, however, we can't spend too long on a defense of our theory, or we'll never get to what we want to do with it. So we propose to just review how it handles the problem cases for the anaphoric-baptismal account, and in so doing illustrate how we intend it to work.

According to the dispositional theory, the reason why descriptive names, descriptive kind terms, or other descriptive expressions *E* are possible is that it is possible that a speaker *S* be disposed to apply *E* to objects only when *S* takes it that they satisfy a certain description *D*. *S* may apply *E* to objects that don't satisfy *D*, under the mistaken assumption that they do; however, were *S* given information that distinguishes the objects that don't satisfy *D* from the ones that do (that is, were *S* not to be mistaken about them), she would no longer be disposed to apply *E* to them. Thus, *E* applies, in *S*'s mouth, to all and only objects satisfying *D*. On this account, it is entirely possible that the relevant description and its integral role in the meaning of *E* was never *represented as a semantic rule* by *S*. *S* might well accept a non-descriptivist tacit semantic rule like "phlogiston" applies to the most natural substance in the vicinity of paradigm examples *X*, *Y*, and *Z*. And maybe oxygen was just such a substance. But for us, none of that is relevant: if *S* is truly disposed to reject the application of 'phlogiston' to anything upon learning that its release from an object does not cause oxidation (say), and no other relevant information could change her mind, then 'phlogiston' is empty.

Consider an alternate set of dispositions. It might be that though *S* now rejects the application of 'phlogiston' to anything, she would apply 'phlogiston' to oxygen, were she to learn more of the history of phlogiston theory, or of the behavior of oxygen, or whatever. In such a case, the individual would be now lacking relevant information, in the sense of lacking information that would influence her dispositions to apply 'phlogiston': thus her current lack of a disposition to apply 'phlogiston' to oxygen would be only a fallible indicator of her disposition to apply 'phlogiston' to oxygen, if she had all relevant information.

As indicated by the examples earlier, our intuitions often seem to track, not causal factors, but simply distinctions that we find important to make—and what makes such distinctions important can vary from case to case. 'Jade' provides a paradigm example. 'Jade' is problematic for causal accounts; it's presumably introduced as a natural kind designator, but it isn't a natural kind. The baptism is defective, yet 'jade' is non-empty—it means *either jadeite or nephrite*. Donnellan's semantic rule account doesn't clearly provide a good account for jade, either—for it's implausible that jade is disjunctive because of some explicitly represented 'escape clause' in its semantic rule that called for disjunctive contents when unified ones weren't to be had.¹⁶ Instead, the explanation simply seems to be that the economic and cultural importance of jade trumps our joint-carving interests in this particular case.

¹⁶ *Pace* Bealer (2002), who makes just such a claim.

Again contra Donnellan, it's highly implausible that this fact about our interests was ever embodied in a tacit semantic rule. But the dispositional account gets the case right; 'jade' refers to *jadeite* or *nephrite* simply because, even knowing these to be distinct minerals, and even under circumstances where we can distinguish them, we are still disposed to call both 'jade'. Our dispositions reflect certain highly contingent social and economic facts. Similarly, although tea is closer to chemically pure water than the stuff in the Hudson River, the latter but not the former is water. Again, this is just how we choose to call things, given our fairly contingent and idiosyncratic interests, and we are not moved by what was baptized what, or what is more 'kindy', or whatever.

Frequently, how others are disposed to apply terms is relevant to how we are disposed to apply them—we defer to experts (more generally, *to others*). The details of this deference are quite messy. Geoff Nunberg once¹⁷ gave the following example: 'carp' means something different to each community—as you go from pond to pond, different fish are called 'carp' and no one is motivated by considerations of what other speakers a town over call 'carp' to change their practice. The case of the color term 'puce' is somewhat different. Many of us would be likely to reconsider our application of that term upon encountering disagreement from another English speaker. However, Americans are apparently not as motivated by what French speakers have to say on the matter, even though the term was originally borrowed from the French—in the United States, 'puce' is a purplish brown, while in France it is a dark reddish brown. 'Liberty' seems different still: a Frenchman's use of 'liberté' may well influence how I decide to apply 'liberty'. In the limiting case, I defer to no one: no information about other speakers' usage can sway me. These are the terms of my private language. Wittgenstein thought this to be impossible, but we don't: in such a case, my expressions are true of that to which I'm disposed to apply them (when I'm not in error about *what the things are* that I'm applying the expressions to). Our account derives semantic values determined by "deference to experts" as a special case. Each of us decides which experts are relevant to the application of our terms, and our terms apply to whatever we would apply them to, when given the relevant facts about expert applications.

As we've noted, the standard causal-historical picture is apparently limited in scope: it applies sometimes (to 'water') but not always (to 'phlogiston' or 'jade') and not necessarily ('water' could have meant *watery stuff*). Why does it apply to some natural kind terms and not to all such terms, or to other kinds of terms? Again, one possible story is Donnellan's DMC: that with each term we associate a semantic rule that grounds the meaning determining facts for that term. So if the semantic rule instructs us to apply 'water' to the local stuff baptized 'water', then that's what *is* water. But this view is unsatisfying: it's implausible that 'phlogiston' is empty because of being governed by a 'descriptive' semantic rule—why should 'phlogiston' have been assigned a descriptive rule, while 'water' received a causal rule? But again, the dispositional account can handle this

¹⁷ At a talk attended by one of the authors.

case, for it's much more plausible that our dispositions were influenced in various subtle ways by contingent facts regarding the history of science, or what have you.

Our interests dispose us to go one way or another in various cases, and *we've never been wrong following our dispositions* when we had all the relevant information. It just couldn't turn out that even though we know the whole story about jade, and we're content to call both jadeite and nephrite 'jade', nevertheless, 'jade' means in our mouths, right now, *just jadeite*. This strongly suggests a close connection between our dispositions to apply terms (under full information) and what those terms mean. What's more, the disposition story isn't perniciously internalist, like the story about semantic rules: a disposition to apply a term to a thing is a real relation between the term and the thing, just as a disposition to sleep during lectures is a real relation between a student who has it and the lectures he attends. So, we propose, the dispositional account is a good place to look for a theory of reference that can support moderate intuitionism.

4.8. The Dispositional Theory and Moderate Intuitionism

As we mentioned earlier, we think that a broadly dispositional metasemantics can provide the needed link between intuitions and truth. It does this via a claim that our application intuitions are generally a guide to what terms apply to which things. But how exactly are our intuitions a guide to what terms apply to which things? Let's look at a particular example. Suppose an English speaker from 1750, George, is given the following prompt:

H₂O/XYZ: You're of course familiar with the word 'water'. Likely, you believe that water is the tasteless, odorless, drinkable clear liquid that flows in lakes and streams. Suppose what I tell you now is true. All tasteless, odorless, clear, etc. liquid on planet Earth is composed of particular amounts of specific kinds of stuff. I could, as it were, write down a 'recipe' for this tasteless, odorless, etc. stuff, and I could also write recipes for everything else. The 'ingredients' in these recipes would be the most basic things in the universe, of which there are about six score.

Now I know that all this sounds strange, but imagine it to be true. Indeed, stranger still, imagine that far away there is another planet much like Earth, with verdant forests, hills, valleys, mountains, lakes and streams. And in the lakes and streams on this other planet flows a tasteless, odorless, drinkable clear liquid. However, this stuff, this liquid, has a very different recipe from the tasteless, colorless, clear, etc. liquid around here.

Again, supposing all I've told you is true, does your word 'water' truly apply to the tasteless, odorless, clear, etc. liquid on this other planet?

Let's suppose George says "No, that wouldn't be water." What we want to suggest is that George's response is good evidence that if he were given all relevant information, he would apply 'water' to H₂O but not XYZ, and thus 'water' for him now means H₂O but not XYZ.

The way this works is that the prompt asks George to suppose that the watery stuff on Earth has a certain underlying nature. We don't know if this information is relevant or not in advance, but if it is, then he's been provided with relevant information. Then the prompt stipulates that there is a sample of stuff that has a different underlying nature. We don't rely on George to tell apart H_2O and XYZ, we just tell him we have an instance of XYZ. So George is mentally 'simulating' a case where he is apprised of (maybe) all information relevant to 'water' application, where he can tell apart two samples, and is asked to further simulate how he would apply 'water' in these circumstances. To the extent that George's simulations of how he would respond in certain circumstances mirror how he would in fact respond, George's application intuitions in response to the prompt are evidence for what his terms mean.

Evidence, we say, but no certain criterion. First, there's no guarantee that the underlying nature of the substances involved is the only piece of information relevant to George's 'water' applications. We (and George as well) can only guess what information might be relevant from our own intuitions, from past experience, and from what we think we know about what's important. It might be that future science cares only for teleological (as opposed to compositional) kinds, and that H_2O and XYZ have one and the same purpose.¹⁸ Apprised of this information, George might lean toward applying 'water' to XYZ as well. As already mentioned, dispositions are responsive to a myriad of highly contingent social, economic, and cultural factors which may not be represented in the information provided by the thought experiment prompt. Second, people are imperfect at predicting their own behavior. I might view myself as cool under pressure, and predict that I would be disposed to risk my life to save a child from a burning building; my real dispositions under such circumstances might be rather different. Similar considerations apply to assessment of one's linguistic dispositions.

Our metasemantic account predicts that linguistic application intuitions—intuitions about when a term applies to an object—will be generally reliable to the extent that a prompt eliciting them specifies enough relevant information. Yet the account also predicts that intuitions are fallible in many cases, such as those described above. Of course, what we ultimately want is an account on which intuitions, including *non-linguistic* intuitions, are similarly reliable-yet-fallible. Fortunately, we've already discussed how to make this step. All that is required is one particular assumption, which we will repeat here:

Assumption: For each T in S 's vocabulary, if there is a property P that T expresses, P is such that:

- (i) O is P iff T applies to O .
- (ii) S intuits that O is P iff S intuits that T applies to O .

¹⁸ By 'teleological kind' we mean a kind defined by its purpose, such as a heart. By 'compositional kind' we mean a kind defined by its material composition, such as gold.

Given this assumption, one's intuition that 'knowledge' applies to case *C* will be just as reliable as one's intuition that case *C* is a case of knowledge.

What does this mean for philosophical intuitions? It means that the intuitions of English speakers about, e.g., consciousness will be exactly as reliable as their intuitions regarding their disposition to apply the term 'consciousness' under possession of all relevant information (and *mutatis mutandis* for French speakers, etc.). And such intuitions can err for exactly the reasons discussed above. For example, as mentioned above, a thought experiment may not always successfully stipulate all information that is relevant to determining an individual's dispositions to apply a term. The course of future neuroscience and psychology could quite plausibly affect our dispositions to apply the term 'consciousness'; insofar as thought experiments don't specify such facts about the future scientific world picture, they leave room for error.

4.9. Clarificatory Issues

Before we conclude, there are a few necessary points of clarification. First: the dispositional account is not directly constitutive, in the following sense. On our account, intuitions are not constitutively related to meaning facts—dispositions are. But there is still a very intimate (but non-constitutive) link between intuitions and truth, for the simple reason that there is an intimate (but non-constitutive) link between intuitions and the dispositions that do underwrite facts about meaning.

Our emphasis on dispositional facts as the ground for metasemantic facts means that we are neither under pressure to recognize any sort of special conceptual competence underlying intuition, nor any sort of a priori insight into meanings. This distinguishes our account from certain superficially similar accounts, such as that offered in Chalmers and Jackson (2001). David Chalmers and Frank Jackson propose that we have a priori access to certain facts about meaning, in the form of 'application conditionals'. One such conditional, for example, might be $G \rightarrow \sim K$, where *G* is a description of a Gettier scenario and *K* is a claim that a certain mental state of that scenario's protagonist falls under the extension of knowledge. More generally, they claim, we have a priori access to conditionals of the form $E \rightarrow T$, where *T* is a statement characterizing the extension of a given term, and *E* is sufficiently detailed information about a given possible world—thus, we know a priori that if some situation *E* is actual, then the extension of term *T* is such and so.

However, on our account, no such a priori access is implied. The facts about dispositions that ground reference are empirical facts; they just happen to be facts to which we have particularly direct epistemic access, in the same sense that, e.g., we have fairly direct access to features of our own personality. But our knowledge of such facts is a posteriori, and is not in any way a special sort of 'conceptual knowledge' with a special epistemic status. Further, unlike Chalmers and Jackson, we don't recognize any special second 'dimension' of meaning (*A*-intensions/primary intensions). Finally, we're not anywhere near as optimistic as Chalmers and Jackson about the prospects for

conceptual analysis via thought experiments—because, as mentioned earlier, thought experiments are generally not guaranteed to specify all relevant information. Relevant information can include deeply varied and idiosyncratic facts; in many cases, we simply won't be able to construct a useful thought experiment without knowing a whole lot more truths about *the actual world*.

It's worth mentioning that our account is compatible with all sorts of views on the nature of intuition. But more importantly, it's compatible with several more or less deflationary views on intuition. Our account doesn't imply that intuition is the special provenance of philosophy, for example—we agree with Williamson (2007) that there is no principled distinction between 'philosophical' intuitions and everyday judgments involving concept application. After all, our account provides the same explanation for our philosophical intuitions as it does for our intuitions about how to apply the term 'footstool'.

Our account is also compatible with the claim (defended by Nado, Forthcoming) that intuition is fairly heterogeneous, in the sense that the actual psychological processes underlying, e.g., moral intuitions are likely quite different from those underlying, e.g., mathematical intuitions. The reliability of intuition isn't found in some specific psychological process that intuitions have in common—instead, it simply reflects the fact that immediate reactions, e.g., to thought experiments, however such reactions are produced, tend to provide some indication of the dispositions that underwrite meaning. This is so, plausibly, because those psychological processes (whatever they are) are likely to be partially determinative of the relevant dispositions—the complex psychological processes governing folk psychology, for instance, are likely partially determinative of how we would be disposed to apply the term 'belief' when in possession of full information. Where that link is weaker, intuition will be increasingly fallible, as in the case of folk physics.

With the clarificatory notes in order, we're now ready to summarize the argument as we see it. Current empirical data on intuition provides us with good reason to adopt a moderate stance on intuition, but it's not immediately clear what explains the truth of that moderate stance. A claim that there is a constitutive link between intuition and meaning would support moderate intuitionism; this link could plausibly come from metasemantic theory. Unfortunately, popular externalist theories don't motivate the needed link, and internalist theories are theoretically undesirable. Locating the constitutive tie in dispositions to apply terms under full information solves the dilemma. Even better, it fits remarkably well with the original variation data. Subjects in different cultures are likely to have different dispositions to apply terms, due to the influence of various contingent social and cultural factors on the subjects' interests and beliefs about the world. Some of these differences are quite likely to evaporate under conditions of increased information. Some may not. To the extent that the differences do disappear, we can assign error to the original intuitions of one or the other group. To the extent that the differences do not disappear, we must recognize a difference in meaning.

With regard to the intrapersonal variation data, the situation is somewhat more complex. Take as an example a study by Schnall et al. (2008) which showed that moral judgments become harsher when subjects are exposed to disgusting surroundings. Thus, intuitions appear to vary according to emotional state. The inclination is to assign error to the emotionally heightened subjects, and our account can potentially explain why this is so. The variability of intuition is in and of itself relevant information—plausibly, when provided with the information that their own disposition to apply moral judgments is being influenced by their heightened emotional state, subjects would have the disposition to defer to their own moral judgments under more neutral conditions. This is exactly analogous to ordinary cases of deference to experts. This is of course empirical speculation, but what's important is that it demonstrates the general method by which the dispositional account might explain away certain variant intuitions as not reflective of meaning-generating dispositions, and as therefore in error.

The empirical data suggest moderate intuitionism. The dispositional account, if true, would provide an explanation for the link between intuition and truth that moderate intuitionism requires. Given the plausibility of moderate intuitionism, we take this to provide abductive support for the truth of the dispositional account. But in addition, the dispositional account has plausibility in and of itself; and since it predicts moderate intuitionism, we take this as reason to endorse moderate intuitionism. As we see it, the two positions form a virtuous circle—each provides evidential support for the other. And we think the overall picture, at least in broad outline, is likely to be pretty close to true.

References

- Bealer, G. (1992). "The Incoherence of Empiricism," *Proceedings of the Aristotelian Society, Supplementary Volume* 66, pp. 99–138.
- . (1996). "On the Possibility of Philosophical Knowledge," *Philosophical Perspectives* 10, pp. 1–34.
- . (2000). "A Theory of the A Priori," *Pacific Philosophical Quarterly* 81, pp. 1–30.
- . (2002). "Modal Epistemology and the Rationalist Renaissance," in T. Gendler and J. Hawthorne (eds), *Conceivability and Possibility*. Oxford: Oxford University Press, pp. 71–126.
- BonJour, L. (1998). *In Defense of Pure Reason: A Rationalist Account of A Priori Justification*. Cambridge and New York: Cambridge University Press.
- Buckwalter, W., and Stich, S. (2014). "Gender and Philosophical Intuition," in J. Knobe and S. Nichols (eds), *Experimental Philosophy*, vol. 2. Oxford: Oxford University Press, pp. 307–46.
- Chalmers, D., and Jackson, F. (2001). "Conceptual Analysis and Reductive Explanation," *Philosophical Review* 110, pp. 315–60.
- Donnellan, K. (1993). "There Is a Word for that Kind of Thing: An Investigation of Two Thought Experiments," *Philosophical Perspectives* 7, pp. 155–71.
- Evans, G. (1973). "The Causal Theory of Names," *Proceedings of the Aristotelian Society, Supplementary Volume* 47, pp. 187–208.

- Fodor, J. A. (ed.) (1990). "A Theory of Content II," in *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.
- Goldman, A. (2007). "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status," *Grazer Philosophische Studien* 74, pp. 1–26.
- Haidt, J., Koller, S., and Dias, M. (1993). "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?," *Journal of Personality and Social Psychology* 65, pp. 613–28.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.
- Katz, J. (1994). "Names Without Bearers," *Philosophical Review* 103, 1, pp. 1–39.
- Kripke, S. (1972). "Naming and Necessity," in D. Davidson and G. Harman (eds), *Semantics of Natural Language*. Dordrecht: D. Reidel, pp. 253–355.
- Machery, E., Mallon, R., Nichols, S., and Stich, S. (2004). "Semantics, Cross-cultural Style," *Cognition* 92: B1–B12.
- Nado, J. (Forthcoming). "Why Intuition?," *Philosophy and Phenomenological Research*.
- Pust, J. (2000). "Against Explanationist Skepticism Regarding Philosophical Intuitions," *Philosophical Studies* 106, pp. 227–58.
- Schnall, S., Haidt, J., Clore, G. L., and Jordan, A. H. (2008). "Disgust as Embodied Moral Judgment," *Personality and Social Psychology Bulletin* 34, pp. 1096–109.
- Sosa, E. (2007). "Experimental Philosophy and Philosophical Intuition," *Philosophical Studies* 132, pp. 99–107.
- Stich, S. (1996). *Deconstructing the Mind*. Oxford: Oxford University Press.
- Swain, S., Alexander, J., and Weinberg, J. (2008). "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp," *Philosophy and Phenomenological Research* 76, pp. 138–55.
- Weinberg, J. (2007). "How to Challenge Intuitions Empirically Without Risking Skepticism," *Midwest Studies in Philosophy* 31, pp. 318–43.
- , Nichols, S., and Stich, S. (2001). "Normativity and Epistemic Intuitions," *Philosophical Topics* 29, pp. 429–60.
- Wheatley, T., and Haidt, J. (2005). "Hypnotically Induced Disgust Makes Moral Judgments More Severe," *Psychological Science* 16, pp. 780–4.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.