

1. Introduction.

Arguments ‘from reference’ seek to establish philosophically significant conclusions using theories of reference, e.g. conclusions about minds (arguments against physicalism), ontology (necessity of origin, natural kind essentialism), and science (anti-realism). In *Against Arguments From Reference*, Machery, Stich et al. argue that the variability of intuitions across and within cultural groups fatally undermine these arguments ‘from reference,’ since it is unclear which intuitions to privilege in our theorizing. Because these theories are grounded in intuitions about cases, to recover referential arguments we need some justificatory account for our use of intuitions in the face of this apparent variability—some story telling us why only certain intuitions are likely correct. In what follows I will argue that it may be harder than we realize to justify the use of intuitions both in elaborating theories of reference and deriving philosophical conclusions from reference. Instead we ought to *integrate* intuitions within philosophical theories: i.e., sometimes theories correct intuitions.

2. The Experimental Challenge

Kripke famously wrote that intuitive content is

...very heavy evidence. I really don’t know, in a way, what more conclusive evidence one can have about anything, ultimately speaking.¹

According to Machery, Stich et al., philosophers have taken Kripke at his word. They argue that the dominant method of motivating—or ruling out—a given theory of reference for a class of terms such as proper names has been to consider an actual or fictional case of reference and let our intuitions about the case guide our theorizing. They term this the ‘method of cases’, claiming that

it is clear from the arguments for and against specific theories of reference that the correct theory of reference for a term ... is commonly thought to be constrained by our intuitions about the reference for this term [...]

And again,

The correct theory of reference for a class of terms *T* is the theory which is best supported by the intuitions competent users of *T* have about the reference of members of *T* across actual and possible cases.²

Indeed, Kripke’s influential arguments to the effect that descriptivist theories do not adequately capture the reference relation proceed from a consideration of counterfactual cases in which the object of reference satisfies no particular descriptive property; yet intuitively it seems clear that we mean to refer to *that object* (and not some other object that may have, in the case being considered, the descriptive property in question).³

Several philosophers have defended the idea that the core evidence for a given theory of reference remains the intuitive conclusions drawn from reflection on salient cases, which then constrain theorizing in different ways: by providing counter-examples to existing theories, for example, or by setting desiderata any adequate theory must meet. These intuitions are “*a priori* justified,”⁴ or comprise a “basic evidential source,”⁵ so that having an intuition that *p* is *prima facie* evidence for the truth of *p*. More detailed proposals, such as Ernest Sosa’s, defines intuitions such that

¹ Kripke (1980), 42. Call this the ‘heavy evidence thesis’: the default evidential significance of intuitions is thus *very high*.

² Machery, Stich, et al., [forthcoming], 8-9. Max Deutsch (2009) disputes this account, arguing that philosophers of language do not, or should not, rely on intuitions at all. I agree, but discussion here will have to be put aside for the moment.

³ E.g. the Gödel/Schmidt case in *ibid.*, 83-84; under the descriptivist account, we would be talking about Schmidt, but Kripke comments “it seems to me we are not. We simply are not.”

⁴ Kornblith (1998), 129.

At t , it is intuitive to S that p iff (a) if at t S were merely to understand fully enough the proposition that p ... then S would believe that p ; (b) at t , S does understand the proposition that p ; and (c) the proposition p is abstract.⁶

Thus the *prima facie* truth of p follows from understanding the proposition in which it is expressed: intuitions are the *a priori* deliverances of conceptual grasp to be understood analogically to perception or introspection. In the case of reference, to ‘fully enough’ understand a proposition p containing a proper name would be to have a belief about the referent for that name, a belief that is *prima facie* true.

The evidential worth of intuitions here is particularly important, as theories of reference for a class of terms can be further deployed to establish substantive philosophical conclusions on a broad variety of topics. As Machery, Stich et al. note, theories of reference figure prominently in debates over eliminative materialism and physicalism in the philosophy of mind; over scientific realism; in ontology; and in ethics.⁷ Machery, Stich et al. analyze these arguments as essentially comprised of three parts:

In the first, philosophers ... adopt a substantive theory of the reference of a term t (or a class of terms T). In the second stage, they claim that the reference of t or of members of T has some specific properties. For instance ... that t refers or fails to refer. Finally, a philosophically significant conclusion is drawn.⁸

Because of the prevalence of such arguments and the significance of the conclusions they reach, it is an important task to demonstrate their validity. Such a task, I will argue, runs into two obvious problems. The first problem, that of intuition variability, is derived from Machery, Stich et al.; and the second is a related issue I’ll term the *handling* problem.

The first problem is that posed by the experimental philosophers: if intuitions are justified *a priori* and form reliable evidence, then how ought we handle apparent variability in intuitive response to cases? On the justificatory accounts typically given, such as the ones sketched out above, any subject with any intuitions is as good as any other. As we have seen, however, intuitions are meant to constraint theorizing. Different intuitions about the same case lead to competing desiderata on what constitutes, for example, a good theory of reference. It is unlikely that both intuitions can be accommodated within *one* theoretical framework. Consider the referential case, where we are asked the intuitive referent of a proper name. Should intuitions genuinely differ it is clear that one or the other is wrong, for we are after *semantic* reference, not speaker reference, ordinary use, or artefacts of pragmatic implicature.⁹ Thus we can state the variability problem plainly:

Variability Problem :

If S_1 has intuition p_1 about a case $xy\zeta$, and S_2 has intuition p_2 about the same case, and intuitions comprise *prima facie* evidence, then both S_1 and S_2 are justified in believing that p_1 and p_2 respectively and articulating the appropriate constraints on a theory T_n . A complete such theory T_n must give an account of $xy\zeta$ that accommodates intuitive response to the case. But p_1 and p_2 conflict. Therefore either S_1 or S_2 is not justified in their belief after all.

⁵ Goldman & Pust (1998), 179; Bealer calls intuitions “basic sources of evidence” in Bealer (2000). His defense is based on a notion of ‘modal reliabilism’ which, I think, is vulnerable to same sort of challenge presented in what follows.

⁶ Sosa (1998), 262.

⁷ Machery, Stich et al., [forthcoming], 3-8; the objection to physicalism has been pressed by e.g. Jackson and Chalmers, but is originally due to Kripke (1980).

⁸ Machery, Stich et al. [forthcoming], 2-3.

⁹ It may well be said that the ‘wrong’ intuition was not a real instance of an intuition at all; on this, see fn. 13. For now let us consider the possibility that semantic intuitions could genuinely differ, e.g. in an experimental design that controls for pragmatics, etc.

The upshot is that one or the other intuition must be wrong, but in the absence of considerations that speak for or against them it will be impossible to know which is, in fact, correct: both S_1 and S_2 can hold their ground and claim *their* intuitions are clear and distinct and the other's confused. What would such considerations look like? S_1 could judge that S_2 isn't a competent speaker—perhaps citing as evidence the intuition that p_2 . This argument cuts both ways, however. Within the philosophical community variability has not occurred; intuitions have been largely shared (or ingrained). Outside the community empirical evidence is mounting that intuitions of reference are *not* universally shared.¹⁰ The longer version of this paper describes in some detail this research. For now, take it for granted that we have evidence that intuitive responses vary but across and within cultural groups. I can discuss this in more detail in question period.

Apart from apparent demographic variation in intuitive responses to philosophical cases, experimental work has also uncovered significant biases that colour our responses to these cases—biases stemming from factors that seemingly have little to do with the cases themselves. To take an example from epistemology: emphasizing the stakes of an epistemic scenario changes intuitive answers about whether a subject 'knows' that p on the basis of given evidence.¹¹ Well, *do* stakes matter to evidence, or is the consideration of stakes a 'bias' to be eliminated? Jason Stanley says the former, arguing that quality of evidence crucially depends on practical import;¹² Keith DeRose, by contrast, suggests that the extension of 'know' is sensitive to context;¹³ Mark Phelan and Ram Neta argue that when and if properly asked, stakes actually *don't* matter in epistemic situations;¹⁴ Nat Hansen in turn suggests existing experimental design is too coarse-grained to provide a definite answer.¹⁵ This raises a host of interesting questions for analytic epistemologists—what are the right questions to ask? What factors are salient to epistemic judgments? Which should be eliminated in order to get an 'unbiased' response? What are the 'normal' cognitive conditions best-suited to the use of intuition?

Such biasing factors are ubiquitous: hand washing and over-all cleanliness affects judgments of moral cases;¹⁶ ascriptions of intentionality are seemingly consequentialist (the so-called 'Knobe effect');¹⁷ ordering affects evaluations of 'true temp' cases;¹⁸ small, seemingly inconsequential changes to the vignettes drastically changes responses to 'trolley' cases;¹⁹ and so on.²⁰ Thus apart from demographic variation in intuitions, we also find variation at the individual level due to framing, order and environmental effects. There is no indication that merely being aware of these effects is sufficient to inoculate oneself against them, or that we have only but started finding them. More problematically, we need to become clear on what *is* relevant to a case and what isn't. This is our second problem:

Handling Problem.

If S_1 has an intuition p_1 about case xyx in circumstance C , and S_1 has conflicting intuition p_2 about that same case in circumstance C^* , and intuitions form *prima facie* evidence, then S_1 must reject either C or C^* in order to be justified in believing that p_1 or p_2 .

¹⁰ Note however that empirical evidence of variability is not necessary to press the objection. One would have to deny the mere *possibility* of genuine intuitional variability to avoid having to give a justificatory account of some sort—at which point we fall afoul of the *no true Scotsman* fallacy: no wrong intuitions are genuine intuitions, so genuine intuitions are necessarily always correct.

¹¹ Stanley (2007).

¹² Ibid.

¹³ DeRose, K., (1992).

¹⁴ Neta, R., Phelan, M. [2008].

¹⁵ Hansen, N. [unpublished].

¹⁶ Schnall, Benton et al., (2008).

¹⁷ Knobe, J., (2003); Knobe, J., [forthcoming].

¹⁸ Swain, Alexander, Weinberg (2008).

¹⁹ Petrinovich, O'Neill (1996).

²⁰ I owe many of these textual references to an unpublished talk by Stephen Stich, [given in Buffalo in October of 2009 entitled "~~Experimental Philosophy and the Bankruptcy of the Great Tradition.~~"]

The issue, of course, is that rejection of C or C^* will require an explicit rationale explaining why the factors present are biasing judgment and why one particular circumstance is *the* correct presentation of the case. This is not a sceptical problem, meant to undermine confidence in any given intuition; rather the issue will be in determining what factors should count in our theory-building. If intuitions are both variable and susceptible to bias, why should we trust them or rest satisfied with how something ‘seems to us’? Because common accounts of intuitions rest upon their intuitive or direct plausibility, friends of intuitions must offer some response to the variability and handling problems. Crucially this response must justify *intuitions* as such. The elaboration of theoretical criterion for the admissibility of intuition must not overshadow the purported evidential worth of the very thing it seeks to justify: then the reasons for which we should *want* intuitions of a certain type may just as well support the theory directly. But we don’t want to *explain why* we have these intuitions, we want to justify their use in constraining theories. Call this the *reasons pitfall*.

3. Responses to the Challenge

According to Hintikka, the current vogue of intuitions in philosophy has its roots in a certain reading of Chomsky. Philosopher’s use of intuitions amounts to equating conceptual grasp with something like Chomskyan grammatical competence, so that possessing a concept is sufficient to begin the process of its analysis.²¹ Chomsky himself has a biological story about grammar, so he’s off the hook *for intuitions of grammaticality*:

It seems now reasonably well-established that there is a special component of the human brain that is specifically dedicated to language [...] This aspect of biological endowment appears to be close to uniform across the species.²²

It is obvious that the postulated innateness of *grammar* does not deliver to us a justification for the use of conceptual intuitions. Yet could a *similar* story be as a way of securing the deliverances of our intuitions? Timothy Williamson has argued to this effect, stating that the ‘armchair’ intuitive method involves “a general cognitive ability to handle counterfactual conditionals.”²³ While a plausible account, as an error-theory it misses the critic’s point. Williamson himself admits that under his account, evaluation of a counterfactual involves “background knowledge” consisting of a “general sense of how things go.”²⁴ This has the result of devaluing ‘pure’ intuition as ‘material for generalizations’²⁵ and instead results in a highly fallible methodology.²⁶ Indeed, if anything Williamson’s frank acknowledgement of these limits bolsters support for Hintikka’s methodological *cri de coeur* about intuitions:

[intuitions] are not premises for philosophical arguments; they are raw material to be critically weighed, corrected, and integrated into a coherent view. Their presuppositions have to be uncovered and their tacit limitations recognized before such integration is possible.²⁷

²¹ Wilson (2006) has argued persuasively that this presupposes a *dual-natured* view of concepts, where we grasp the same concepts exemplified in the external world, as both “manifest the same content in some primitive fashion.” Wilson’s main criticism is that this “substantially [exaggerates] human linguistic capacity,” as concepts in fact do not tidily carve up the world as such a view predicts. *Ibid.*, 90-91, 93. We will return to this in section 4.

²² Chomsky (1997), 13-14.

²³ Williamson (2005), 16-17. He is concerned with analytic epistemology, but his account can be generalized to cover intuitions of reference readily.

²⁴ *Ibid.*, 17.

²⁵ I owe the formulation to Hintikka (1999).

²⁶ Writes Williamson, “a reasonable speculation is that their [the cognitive capacities we use for counterfactual evaluation] applications in philosophy have roughly the same degree of reliability as their applications elsewhere.” Williamson (2005), 26.

²⁷ Hintikka (1999), 138-139. There is little in Williamson’s paper to suggest he would disagree with Hintikka’s assessment here. In fact, Hintikka proposes a ‘general,’ fallible faculty not unlike Williamson’s; and I have yet to discover a better proposal in the literature.

If this were the state of play, then Machery, Stich et al. would have little to complain about. But as they correctly note, the “dominant way of independently motivating” theories of reference involves pure intuition about cases.²⁸ Admitting that the cognitive processes underlying counterfactual evaluation are fallible products of reflection on background knowledge saves the methodology to the exact extent that it weakens the results (i.e., contrast with Sosa’s definition).²⁹

But Sosa has a line of defence, however, that does not involve a Chomskyan analogy. It is simply the claim that careful second-order reflection obviates any worries about extraneous factors biasing intuitions. Perhaps demographic variation reflects culturally optional criteria or standards; perhaps biasing factors surreptitiously introduce such optional criteria or standards too. Despite this, Sosa argues that

Once having discerned the optional criteria, so as to hold them up separately for consideration on their own, the question will remain whether to adopt them. To say that intuition speaks in favour of doing so, either directly or via the deliverances of reflective equilibrium, is now separable from mere ethnocentric xenophobia. For the appeal to intuition here, once we are holding the criteria or standards themselves in focus, is quite distinct from any conservative appeal to community consensus. If I believe that $2+2=4$ because this is obvious upon consideration, then the reason why I believe it, its obviousness upon consideration, is quite distinct from the fact that everybody else also agrees.³⁰

Presumably this means one can recognize intuitively ‘optional’ criteria as such, which is certainly not obvious given the nature of the handling problem. Over and above this consideration though it seems we have only set the problem at one remove. The question is: If our intuitions can potentially reflect such ‘optional’ criteria, how are we to decide whether they are salient? By consulting our intuitions *on these*. Yet if we grant that first-order intuitions may require such second-order reflection, why should we grant that these second-order intuitions are not in a precisely similar position? Sosa must remain steadfast in his insistence that, like turtles, it is intuitions all the way down, in order to avoid what I termed earlier the *reasons pitfall*. This is why he insists on the ‘obviousness upon consideration’ of a proposition such as $2+2=4$.

There are nearby examples that can be used to make the opposite point, though: is it intuitively obvious that $n^0=1$? Or again, what is the intuitive meaning of 0^0 ?³¹ Why should intuitions settle such a question? To take up a simpler example, consider whether 0 is even or odd, a question that has perplexed generations of school-children. It perhaps *becomes* intuitively obvious that 0 is even upon consideration of a set of reasons: it is a multiple of 2, it is evenly divisible by 2, it is surrounded by odd numbers.³² It may be fiendishly difficult to separate intuitive obviousness from internalization of strongly compelling rationale—the usefulness of experimental philosophy is really in its reminder of this. Other times it is simply clear that intuition is no guide whatsoever to truth. The Banach-Tarski paradox, which states that it is possible to decompose a sphere and re-assemble the pieces into two further spheres identical to the first, flies in the face of intuition; but hardly anyone believes this casts aspersions on the axiom of choice. It simply would be mistaken to defend the truth or falsity of a derived mathematical proposition solely on the grounds of intuitive obviousness or lack thereof.³³ There seems no principled difference between *any* mathematical result, only psychological facts about what strikes us as easier to understand at first glance.

²⁸ Machery, Stich et al., [forthcoming] 16.

²⁹ Hintikka’s critical reading of Kripke’s assumptions and limitations are the perfect example; limitations of space force me to omit the blow-by-blow, but it is well worth reading in Hintikka (1999)140-143.

³⁰ Sosa (2008), 5.

³¹ 0^0 is an interesting case study: definition is a clear-cut matter of convention and context, both as a matter of sociological fact and mathematical principle. See Knuth (1992), 491-492.

³² Reaction-time experiments show that the parity of 0 is psychologically ambiguous; pedagogical materials emphasize the discussion here provides an excellent opportunity to stress the importance of reasoning from definition. The equivalence $.99999\dots = 1$ is another classic ‘counter-intuitive’ classroom example.

³³ Perhaps a case could be made that axiom choice is guided by intuition. This seems contradicted by the historical record. On this see Maddy (1999).

Where lies the demarcation between actual cases of belief *based on* obviousness—cases whose clear-cut existence Sosa is committed to—and nearby cases requiring explicit justification, truths that are not obvious in the same manner? A great deal of work is being done in the background of even the most basic proposition of arithmetic and yet the cases considered by philosophers are far more complex, involving semantic notions like ‘truth’, ‘knowledge’, ‘mind’, ‘natural kind’. What I wish to say now is rather prosaic: *Reasons are inevitable*. What experimental philosophy shows us is not that the proper way to resolve philosophical perplexities is by taking an opinion poll, but that our intuitions are shot through with premises, inferences and ratiocinations, with contextual sensitivities and definitional ambiguity. We’ll return to this.

A rather different rebuttal to the critiques of experimental philosophy has been pressed by Max Deutsch. Directly addressing the work of Machery, Stich et al. which criticized the reliance upon intuitions exhibited by philosophers working on theories of reference, Deutsch claims that intuitions in fact play no role at all—thus the experimental criticism is severely misguided. Philosophers of language do not rely on the ‘method of cases’ nor need make any predictions about the intuitions of competent speakers:

Nothing in Kripke’s famous argument against the descriptivist theory of reference for proper names hinges on assuming anything about people’s intuitions [...] to show that [it] is false, Kripke simply describes counter-examples [...] there is no explicit appeal to intuitions either in my brief rendition of Kripke’s argument, or in Kripke’s original presentation in *Naming and Necessity*.³⁴

This is almost certainly a revisionist reading. While Deutsch is right that presenting a clear counter-example in the manner described is a methodologically innocuous practice, and that re-reading Kripke in this fashion potentially deflects the experimental critique, a *re*-reading it is. The first lecture of *Naming and Necessity* is clearly explicit in its reliance on intuitive content, openly defending it—as we’ve seen—as ‘conclusive evidence ... ultimately speaking.’ Earlier in the same passage, Kripke contrasts intuition with philosophical theory, squarely siding with the former:

Suppose that someone said, pointing to Nixon, ‘That’s the guy who might have lost’. Someone else says ‘Oh no, if you describe him as “Nixon,” then he might have lost; but of course, describing him as the winner, then it is not true that he might have lost.’ Now which one is being the philosopher here, the unintuitive man? It seems to me obviously the second. The second man has a philosophical theory.³⁵

In other words, we ought not let prior commitment to philosophical theory override our intuitive judgments; rather, theory is constrained by intuitive content, just as we diagnosed in the previous section. But let’s grant this reading. Does it salvage intuitions? No.

Deutsch’s revisionary reading of Kripke, while ostensibly a broadside against experimental philosophers, concedes much in order to block their conclusions. With Machery, Stich et al. and against Sosa he agrees that intuitions aren’t evidence, which certainly is something.³⁶ Contra the experimentalists, however, Deutsch claims this does not imperil the elaboration of theories of reference. This is because Deutsch claims that Kripke presents genuine counter-examples to descriptivist theories of reference, and how these counter-examples strike the average person is irrelevant: “the predictions of a theory of reference concern terms and their referents, not competent speakers and their intuitions.”³⁷ There is a fact of the matter about the referent of a proper name that is not responsive to the intuitions, beliefs, or usages of speakers. The reference relation is independent; so we can legitimately ask, as Deutsch does, “to

³⁴ Deutsch (2009), 445, 446, 447.

³⁵ Kripke (1980), 41.

³⁶ “I also won’t argue that the intuitiveness of *p* is not evidence of any kind for *p*, though this is something I happen to believe.” Deutsch (2009), 450.

³⁷ Ibid., 448.

whom does this speaker's uses of 'Gödel' refer, Gödel or Schmidt?" and provide an answer. The *correct* answer is Gödel, not Schmidt, irrespective of what any competent speaker thinks.

Such a move only brings us back to the opening methodological question. Machery, Stich et al. claimed that the correct theory of reference must accommodate speaker's intuitions about reference. Deutsch says intuitions have nothing to do with it. A theory of reference must make true predictions concerning terms and their referents; but how do we come to know the genuine referent of a term? Experimental philosophers are not—or at least shouldn't be—interested in intuitions *qua* intuitions. They are interested in the judgments philosophers and lay folk make about the cases they consider. What makes judgment reliable in these matters? 'Intuition' was just a word that pointed to our sense of 'obviousness upon consideration' that accompanies such a judgment. Deutsch's criticism is that majorities or pluralities of intuitions aren't constitutive of truth, and that's right; but surely experimental philosophers can agree to this. Intuitions don't *constitute* what is true. Yet for him *some* kind of sense of intuitive obviousness underlies our evaluations of responses to cases. How are we to resolve disagreement, should it arise?³⁸ *That's* the methodological problem.

Unfortunately Deutsch does not have a very convincing story to tell about the way philosophers make true judgments:

Suppose we philosophers know that a subject in a Gettier case does not know. How do we know this? The traditional answer is: By *thinking* about the case. Giving this answer appears to commit one to the existence of *a priori* knowledge.³⁹

So it does. Though such a commitment doesn't do a very good job at deflecting methodological concerns if the question just *is* what justifies our thinking a particular way about a case. Deutsch thinks it does:

For the philosopher who is not a skeptic about *a priori* knowledge, and who conceives of the philosophical method as including a significant *a priori* component, the results of the experimental philosopher's surveys are irrelevant. For the surveys to be relevant, the philosophical questions upon which they allegedly bear must be conceived as being answerable via *a posteriori* methods. But non-skeptical philosophers do not conceive of the questions this way.⁴⁰

This is seriously mistaken. Of course *surveying* intuitions is *a posteriori* if anything is. But their relevance does not rest on a thesis that reference is *determined* by the results of the survey. Rather the problem is that *if* judgment does genuinely differ, *if* there is genuine variability in the deliverances of 'thinking about the case' it is not clear which judgments are right. Deutsch is entirely correct that a commitment to the *a priori* entails a commitment that *one* answer is the right one, and not necessarily the most popular one at that. How does this help escape the problem that 'thinking about the case' gives us multiple conflicting answers? The problems we set out earlier involved the *prima facie* justification of 'intuition,' but the deliverances of *a priori* thinking are, under Deutsch's account, indistinguishable. Does he believe respondents to a survey aren't 'thinking about the case'? Maybe so, because he thinks it is a fair argument to claim that "philosopher's judgments about conceptual contents or word meanings are more reliable than those of the folk."⁴¹

Such a position is difficult to motivate. One could equally claim that the peculiarity of the discipline of philosophy reinforces specific intuitive responses. As Hansen notes, the philosophical thought experiment must certainly

³⁸ And it *does* arise, as the recent PhilPapers survey results show us. Not all cases are driven by intuition-pumps *per se*, but certainly many by what strikes philosophers as 'obvious upon consideration': <http://philpapers.org/surveys/results.pl>

³⁹ Ibid., 459.

⁴⁰ Ibid., 460.

⁴¹ Ibid., 459.

be the only type of experiment where it is perfectly alright to announce the desired result from the outset.⁴² Given the breadth of the variability and handling problems one cannot rest content with a mere claim of reliability: we're owed an argument, and Deutsch musters none to support his claim.

4. Theory Integration (and its consequences)

Judgments, intuitions—the deliverances of thinking about a case—ought to be argued for and integrated into theories. So seeming 'obviousness' is not the final arbiter, but just another part of analysis. Examples of this practice can be easily found in the literature; if it needs to be brought out explicitly, it is only because in the fractious conflict between experimental philosophers and armchair analysts, both have missed the already-practiced methodological compromise.

To give but one example: while commenting on a clash between intuition and the semantic theses he advocates, Scott Soames carefully writes that

On the one hand, the intuitions are persistent and widespread. On the other hand, the semantic theses I have adopted are highly motivated; it is not at all easy to see how they could be wrong, or what modified theses could be put in their place. In order to resolve this dilemma, we will have to look very carefully at the sources of intuitions about [these] examples, and to review the arguments given for our central semantic claims.⁴³

The particulars of the examples can be put aside for our present purposes.⁴⁴ Rather here Soames illustrates in practice the point I am making in theory. There is no reason to think that the intuitions he alludes to are the final word on the issue. 'Highly-motivated' theses sometimes need to *correct* erroneous intuitions. These intuitions might persist even after acquiescence to the theses Soames presents, which are carefully argued for in central chapters of *Beyond Rigidity*—a contrast to *Naming and Necessity*, despite Deutsch's protests to the contrary.⁴⁵ Soames has an explicit account of intuitions that prioritizes theory over the reactions of competent speakers:

...it is not surprising that the answers that speakers give to semantic queries differ from those provided by a correct semantic theory. [...] The totality of information the language user associates with sentences containing the name or natural kind term will vastly outstrip what the theorist recognizes as the semantic content of the sentence. [...] we ought to give up the assumption that individual speakers have internalized semantic theories that provide them with the means of identifying the propositions semantically expressed by sentences and distinguishing them from other propositions the sentence may be used to assert or convey.⁴⁶

By itself the passage provides support for Deutsch's criticism of experimental philosophy and of its methodology—I have passed over his salient critique of the inexcusable neglect of the speaker/semantic reference distinction in the experimental surveys. The difference lies in how Soames and Deutsch approach the problem, with Deutsch asserting the unproblematic use of *a priori* thinking and Soames clearly preferring to combine conceptual analysis with careful

⁴² "Experimenter bias is an effect generated when experimenters disclose (even unconsciously) their own beliefs about the 'correct' response to an experiment," Hanson [unpublished], 3. Such biases are to be avoided at all costs in psychology, but not philosophy—it would seem. In any event, the Phil Papers survey shows widespread intra-disciplinary disagreement, despite philosopher's sophistication.

⁴³ Soames (2002), 142.

⁴⁴ The very basic summary is that the sentence "Peter Hempel lived on Lake Lane in Princeton" is purported to have the same semantic content as "Carl Hempel lived on Lake Lane in Princeton." The argument is partially based on considerations related to normal contexts of utterance and a more formal principle "about the identity of propositions," a tour-de-force of carefully motivated conceptual analysis. See *ibid.*, 142-146.

⁴⁵ Space limits prevents a fuller contrast between the two, though the interested reader is warmly invited to do so.

⁴⁶ Soames (2002), 68, 70.

argument. I contend Soames is right, on grounds that the mere acknowledgement of the divorce between intuitions and correct theory is no inoculation from the former's distorting influence on the latter, which only theory and argument can properly correct. This methodological approach I will term *theory integration*: philosophical theories are not constrained or created wholly by intuitions. Theories are motivated by the usual criteria that speak for them: predictive power, explanatory depth, normative probity, simplicity, and so on. In some cases theories correct stubborn intuitions. Other times an intuition might legitimately up-end a theory—but even in these cases, I will argue, the grounds for the intuition should be well-understood. The insight *simpliciter* shouldn't be doing the work.

Which is not to say thought-experiments are not useful! We can distinguish between 'Empyrean' and 'Empirical' uses of intuitions (this distinction might form a continuum, not discrete categories). Empirical thought experiments are those which can either be experimentally realized or are extrapolations of physical laws where relevant concepts are 'held steady.' People *can* be put in Gettier scenarios. Galileo *could* attach two weights to each other. Empyrean⁴⁷ thought-experiments are *not* realizable or obvious extrapolations of laws, and involve new applications and extensions of concepts. While the Chinese room is thought to be *nomologically possible*, it is unclear that we understand what is involved in a Turing-satisfying translation book. The properties of such a book might very well incline us to revise current concepts, but we cannot anticipate such revision in advance of its creation (unlike the Galilean case). The twin-Earth case is similar: it remains an open question how we would want to conceptualize actual XYZs. It should be noted that scientists themselves express conceptual caution in similar cases (e.g., atomic super-clusters).

Contrast this conceptual caution with Forbes' methodological prescription:

The status of a property as accidental or essential is settled by *a priori* principles ... to the extent that one finds the principles plausible, they seem forced on us by the nature of our concepts.⁴⁸

It is intuitive reflection that tells us what the 'forced' results are. I do not think this is a justifiable methodological practice. For example, deriving 'essentialist' conclusions from our intuitive evaluations of modal cases involves taking as given substantive and controversial principles which only a theory integration stance can uncover and evaluate. Holding *one* class of intuitions as correct in twin-Earth cases implicitly assumes philosophical positions that are equivalent to the conclusions they seek to establish.

Why? "If water is H₂O, then necessarily it is" seems to follow without controversy from the common intuition that there are no possible worlds where there is water, but not H₂O. In a recent discussion of Hughes' book *Kripke: Names, Necessity, and Identity*, Fodor outlines a problem that illustrates the circularity we're after: what about these intuitions, the ones we use to establish that indeed there are no possible worlds where water isn't H₂O?

What are modal intuitions intuitions of? ... the question I was trying to raise wasn't: 'What about possible worlds makes it necessary that water is H₂O?' My question was: 'What about *water* makes it necessary that water is H₂O?' There must be something about water that does because, notice, there are plenty of kinds of stuff for which the corresponding modal claim would be false.⁴⁹

⁴⁷ The highest plane of the Celestial Spheres, where God is said to reside (in the case of the Philosopher's God, true Knowledge).

⁴⁸ Forbes (1997), 529-530. It is hard to see how the intuitive obviousness of arithmetic is analogical to the principle expounded here.

⁴⁹ Fodor (2004), 8. Two-dimensionalists attempt to address such objections. I hope my own extension of Fodor's arguments sidestep the issues raised there, but I cannot go in further detail here. Yablo (2000) does a good job in addressing anti-physicalist arguments that rely on two-dimensionalism.

Surely there are possible worlds where Mountain Dew is made of XYZ; asks Fodor, then, what is the difference between water and soft drinks? Well, one is a natural (or ‘material’) kind concept: “water is a material kind concept because every sample is *ipso facto* required to have the same microstructure that actual samples do.”⁵⁰

Required by what? This is where, according to Fodor, the story gets hazy: the requirement holds because water just *is* a material kind concept. Fodor quotes Hughes’ own discussion of this very issue:

If it should turn out that only philosophers balk at classifying XYZ as water, I am ready to defer in my usage to non-philosophical majority and say that ‘water’, like ‘glue’, is not the name of a kind with a chemical essence.⁵¹

But making such a move, while seemingly unavoidable, defeats the purpose. For it turns out that, crucially, modal intuitions end up relying on ‘optional’ facts about concepts—just as the experimental philosophers worried. It is unclear what Sosa could say here: is it ‘obvious upon consideration’ that water *must* be a material kind concept? That any given concept *is* a material kind concept is *a posteriori* if anything is: so we ought to say “*if* ‘water’ is a material kind concept; and *if* water is H₂O, then necessarily water is H₂O.” Of course the *definition* of a material kind concept guarantees that all samples of the kind be made of the same ‘stuff’, trivializing the result. This is exactly the point where Nathan Salmon cries foul in *Reference and Essence*, arguing that Kripke and Putnam both rely on “a hidden essentialist premise that is quite independent of the theory of reference,” and thus the program to derive essentialism from reference (and hence intuitions of reference) is a failure.⁵² What is the premise? Salmon identifies it thus: “Being a sample of the same substance as something consists in having the same chemical structure.”⁵³

Salmon’s analysis of the modal argument demonstrates that the premise, in order to secure the validity of the argument, includes an explicit essentialist principle. As Salmon puts it,

In order for some liquid sample *x* in an arbitrary possible world *w* to be cross-world consubstantial with the actual paradigm water sample mentioned in the ostensive definition of water, sample *x* must have the same chemical structure in *w* that the actual paradigm has in the actual world, namely the chemical structure H₂O.⁵⁴

No matter how plausible a principle it may seem, the hidden presence of such a substantive claim packaged in a canonical and ‘obvious’ intuition-pump underscores the need for caution and critical care when dealing with arguments ‘from reference.’ Salmon is quite correct to claim it is not a principle of science, but of metaphysics. Indeed, when confronted with *actual* ‘xyz’-style cases, scientists generally express conceptual caution, knowing full well that a re-thinking of approaches may be required. Atomic ‘super-clusters’ have been discovered which mimic the properties of specific atomic types, and for predictive and classificatory purposes, scientists extend “the periodic table into a third dimension” by describing the electronic and not atomic structure of clusters.⁵⁵ Is an atomic cluster the same substance as that which it mimics? Intuitively, no; yet projecting the periodic table ‘upwards’ by consideration of electronic similarities muddles the issue sufficiently that any definite answer can be seen to invoke a version of the essentialist principles Salmon outlines and not be a question of mere ‘intuitive’ reference.⁵⁶

⁵⁰ Ibid., 9. This microstructure plays both a *semantic* role, in securing the reference, and a *causal* role, in being responsible for the kind’s properties—conceptual dualism again.

⁵¹ This is Hughes, quoted *ibid.*, 9. Hence Fodor concludes that “it’s ‘water’ being a material kind concept that vindicates the intuition that water is necessarily H₂O.”

⁵² Salmon (1979), 713.

⁵³ Salmon (2005), 166.

⁵⁴ *ibid.*, 180.

⁵⁵ Amato (2006), 1-2.

⁵⁶ For more on electronic versus chemical/atomic dissimilarities and classificatory controversies, see Rich (2005) 1761-1763.

Fundamentally, the reason we need to be careful in our use of intuitions is that conceptual grasp is far more slippery than commonly assumed, vulnerable to new information that shifts and extends understanding. Yablo senses this when he writes,

Am I the only one who feels the intuition of zombies to be vulnerable? I am *braced* for the information that is going to make zombies inconceivable, even though I have no real idea what form the information is going to take.⁵⁷

Yablo, reflecting on this, concludes that there is no “conceptually determined truth of the matter” what modal intuitions a “given evidential diet” *should* give rise to.⁵⁸ This is because there is no canonical response to evidence: “rational thinkers will ... draw different conclusions from the same evidence,” a thesis Yablo considers uncontroversial.⁵⁹ Our best arguments may rely to some extent on intuitions—of reference, conceivability, and so on—but these are limited by the vulnerability of the involved concepts *themselves* to non-monotonic change.

In contrast to this position is the seldom-articulated thesis of ‘semantic finality,’ which asserts that “conceptual grasp becomes complete and stable after a period of early acquisition.”⁶⁰ Once we *understand* a concept, e.g. ‘water’, we have acquaintance with a concept that carves up nature in determinate ways, independently of us. Scientific investigation may yield to us the facts that such a ‘carving-up’ reflect, but we don’t need to know this in order for the concept to do its job. *That* such a carving up is true, regardless of the explanation for it, secures our modal intuitions about natural kind terms. But it is at best unclear that concepts do function this way, even the prosaic ones used in the very examples deployed to buttress essentialist arguments.

Looking to the most ‘fundamental’ characterization of ‘water’ given by our best theories yields an eccentric substance whose classification and properties are negotiated in highly complex ways. Are ‘clathrate hydrates’—a form of hydrogen-bonded crystalline solid—ice, asks Wilson?⁶¹ The answer seems to depend on whom you ask, what literature you consults, what your epistemic interests are. ‘Water’ is an ideal example to pursue the point. It is only true in the roughest sense that samples of water *are* H₂O, even when considered as a necessary but not sufficient condition for kind membership. At best, the slogan serves as a kind of shorthand notation for a hugely complicated set of behaviours.

In any sample of water, self-ionization of H₂O molecules creates continuous streams of hydronium and hydroxide ions. These ions are an important part of water’s microstructure, giving rise to many of water’s properties—its slight conductivity and pH, for instance. Notes one research team,

A randomly chosen, intact water molecule will dissociate in liquid water, producing hydronium (H₃O⁺) and hydroxide (OH⁻) ions, within ten hours.⁶²

The form of the necessity is: “If water is H₂O, then necessarily it is.” So do we amend the basic structure to reflect self-ionization? Recall that some of the causal properties of water supervene on the presence of these ions; they are not ‘impurities.’ In a recent journal article devoted to this very issue, R. Hendry writes

I conclude then that samples of water are not mere assemblages of H₂O molecules ... compounds are more problematic and complex [than elements] so that micro-structuralism about compounds will need to be assessed case by case, and water is only one.⁶³

⁵⁷ Yablo (2000), 119.

⁵⁸ *ibid.*, 117.

⁵⁹ *ibid.*, 117. This is simply the nature of inductive logic.

⁶⁰ Wilson (2006), 140. Wilson claims this is a central facet of a “classical picture” of concepts that shares deep affinities from Russell to Kripke, despite surface dissimilarities.

⁶¹ Wilson (2006), 55-56.

⁶² Geissler et al., (2001), 2121.

Self-ionization is but one wrinkle when dealing with ‘water,’ though in itself it certainly demonstrates the need for scientific sophistication in dealing with naive ‘essentialist’ claims: intuitively obvious first principles are never a match for nature’s messiness.⁶⁴ (Isotopic variability presents another challenge, one that leads scientists to create ‘standard’ waters by direct stipulation!)

5. Conclusion

Perhaps the more natural position to take in regard to this ‘messiness’ would be to say that ‘water is H₂O’ is *motioning* towards a larger body of knowledge and does not involve ‘reference’ at all. This could be one consequence of taking theory integration seriously: any theory involving terms with complex semantic underpinnings ought to take these complexities into account. Thought experiments cannot be considered in isolation from their scientific brethren.

Wilson has been the most lucid recent advocate of such a position, writing that

We employ a rich variety of evaluative platforms that involve both ‘water’ and ‘H₂O’ often with a local preference for one or the other, but accompanied by a basic willingness to substitute one term for the other as long as contextual clashes do not arise. So neither linguistic term designates any clear or fixed form of attribute ... the loose schema ‘water = H₂O’ reports upon this terminological interchangeability, without providing evidence that *either* predicate has become definitely tuned to some fixed chemical attribute.⁶⁵

Theory integration suggests we must investigate the complex negotiations of concept and world not only in *use* but in the constraints placed on use. This involves an extension of ‘experimental’ methodology to include thorough engagement with empirical data. Indeed, viewed in this light, intuitions become a useful tool for diagnosing *potential* conceptual confusion—a kind of signpost during the course of careful and painstaking investigation. This meshes well with deflationary positions towards reference such as the one elaborated by Hartry Field:

on the deflationist viewpoint, though, the observations [of Kripke] aren’t at the most basic level about reference, but about our inferential practice.⁶⁶

Philosophy can easily carry forth without a substantive theory of reference and without methodological reliance on indubitable intuitions. Mapping out this path ought to figure among our chief future concerns.

⁶³ Hendry (2006), 871, 873. In the end Hendry concludes that microstructuralism fails for some kinds (e.g., acids) and is relative to “epistemic interests” of chemistry in other cases (water): “many substance terms are used in contexts that are not governed by these interests,” (ibid., 874). The similarity to the present argument is evident.

⁶⁴ Further interesting complications that count against too-simple reference and essence claims are can be found in Chomsky (1997b), 50-52; Dennett (1995), 408-412.

⁶⁵ Wilson (2006), 429.

⁶⁶ Field (1994), 261.