Loose Constitutivity and Armchair Philosophy

Jonathan M. Weinberg[a], Stephen J. Crowley[b]

[a] Department of Philosophy & Program in Cognitive Science, Indiana University at Bloomington
[b] Department of Philosophy, Boise State University

Standard philosophical methodology which proceeds by appeal to intuitions accessible "from the armchair" has come under criticism on the basis of empirical work indicating unanticipated variability of such intuitions. Loose constitutivity—the idea that intuitions are partly, but not strictly, constitutive of the concepts that appear in them—offers an interesting line of response to this empirical challenge. On a loose constitutivist view, it is unlikely that our intuitions are incorrect across the board, since they partly fix the facts in question. But we argue that this ratification of intuitions is at best rough and generic, and can only do the required methodological work if it operates in conjunction with some sort of further criteria of theory selection. We consider two that we find in the literature: naturalness (Brian Weatherson, borrowing from Lewis) and charity (Henry Jackman, borrowing from Davidson). At the end of the day, neither provides the armchair philosopher complete shelter from extra-armchair inquiry.

Keywords: experimental philosophy, armchair philosophy, intuitions, constitutivity, semantics, conceptual analysis

1. The Experimental Restrictionist Challenge

The perennial debate about the standing of the method of intuitions (aka armchair philosophy)[1] in philosophy took a turn for the empirical in the last decade, as a number of self-described "experimental philosophers" reported

[1] Just how far armchairs extend is a matter of some recent contention; see (Williamson 2008) and (Nolan 2009). For the argumentative purposes here, we will grant that the armchair may contain in its epistemic ambit a moderate amount of non-specialist empirical knowledge.

experimental findings which, they claim, should cast doubt on the overall trustworthiness of that method.² This position has elsewhere (Alexander and Weinberg 2007, 61) been termed the "restrictionist" program in experimental philosophy, and we will follow that usage here; and we will refer to the central argument of these philosophers as "the experimental restrictionist challenge", or simply "the challenge". Our interest here is in canvassing one possible line of response to the challenge. In order to do so, we will begin by devoting this introductory section to a brief articulation of the dialectical structure of that challenge.

The experimental restrictionist challenge, as we understand it, has three main moving parts. First, there are the experimental results themselves, which are purported to indicate that the intuitions of their subject population is worrisomely sensitive to such factors as ethnicity, affective content, or order of presentation (see, for example, Weinberg et al. 2001, Nichols and Knobe 2007, and Swain et al. 2008). Second, while the method that the restrictionists want to restrict is that of professional philosophers, nonetheless the experimental studies have thus far been conducted almost entirely on untrained or relatively untrained undergraduates and random persons-waiting-for-a-bus. The challenge thus requires a further claim that there is a relevant similarity between the experimental subjects on the one hand and philosophers on the other, sufficient to license an ampliative inference from the population studied to the population ultimately targeted. Third, experimental restrictionists require the sensitivities that they experimentally document to be philosophically problematic—the differences in intuitions must be ones that lead away from the relevant philosophical truths, and not simply track them. (For example, intuitions attributing knowledge to a given agent are likely to be sensitive to the quality of the evidence possessed by the agent. Surely this would not be a *problematic* sensitivity for such intuitions to have!)

If they have all three pieces in place, then the experimental restrictionists do seem to have the makings of a substantive challenge to the philosophers' method of intuitions: they have evidence that the intuitions of the philosophers themselves may well be problematically sensitive to factors that will lead them away from the truths they aim for. Those looking to defend the armchair may thus want to respond to one or more of those pieces.

---

² For overviews, see (Alexander and Weinberg 2007) and the papers cited therein; and (Knobe and Nichols 2008) and the papers contained therein.

1.1   Responses to the Challenge

If the restrictionist challenge is successful, then armchair philosophy looks to be in significant trouble, displaying unanticipated liabilities to error. However, before consigning their armchairs to the flames (along with Hume's volumes of sophistry and illusions), defenders of the armchair will want to canvass all their options thoroughly. In principle, three broad classes of response to the restrictionist challenge are available that would enable the challenge to be avoided, corresponding to each of the three parts of the challenge. However, neither the first nor the second parts of that challenge can be successfully countered from the armchair. Responding to the experimental results and their interpretation is clearly a scientifically substantive project. And it turns out that the question of who can be expected to have expertise about what and under what circumstances is a pretty ticklish empirical question. And even if *some* sorts of philosophical expertise may be expected unproblematically, such as the expertise of historians of philosophy regarding the views and arguments of their preferred philosophers, nonetheless whether or not philosophers are expert *intuiters* is in particular something that would require some substantial empirical investigation to discern.[3]

Thus those who would pursue a strictly armchair response to the challenge are left with door number three: that the putative variation in philosophical intuition need not be deleterious, because such variation does not diverge unmanageably from the philosophical truths. Resistance of this kind has manifested itself in a variety of ways. Ernest Sosa (2006) and Timothy Williamson (2008) have each suggested that as long as we deploy intuitions carefully all will be well; a pious but not particularly practical suggestion, since precisely what is at stake here is whether philosophers actually know *how* to be careful in the requisite way. At a minimum such an approach is hostage to an, as yet undeveloped, substantive account of the underlying competence in question. Other theorists, such as Bealer (see for example Bealer 1999), have suggested that good or genuine intuitions bear some phenomenological mark that careful investigation can identify. But there is a great deal of high-level disagreement about the existence of such phenomenology, with such prominent authors—and would-be deployers of intuitions—as Williamson (2004, 117), Sosa (2009, 54), and Goldman (2007, 11) denying it in their own cases. (And, for what it may be worth, the authors of this paper cannot even agree between the two of them as to the exact nature of the phenomenology of intuitions!) This general dissensus indicates that even if any such mark of "real" intuitions exists, it is not yet something that can be put to good intersubjective methodological use.

---

[3]   See (Weinberg et al. forthcoming).

What these two forms of resistance share is the view that the appropriate form of modification to current philosophical practice is one where we appeal to some armchair-available resources—such as intellectual care or phenomenology—to separate the good intuitions from the bad. And they all suffer from the same problem, namely, that it does not turn out to be that easy to tell, from the armchair, just which intuitions are wheat and which ones are chaff. We do not pretend to have offered any sustained argument against such approaches here, but we do note that they face a particular kind of burden, one with both theoretical and methodological dimensions. The theoretical challenge is to find a sign of the trustworthy intuitions, and the methodological one to find a way for us to make use of that sign from the armchair. What is particularly challenging is meeting *both* burdens at *once*.

We highlight this double-edged burden in order to motivate an entirely different way of approaching the challenge. This alternative reverses the direction of the relation between philosophical intuitions and philosophical truths. Rather than treating intuitions as more-or-less good sources of evidence about some intuition-independent philosophical truths, this alternative approach construes the intuitions as in part *constitutive* of the relevant truths. On this account what one of our terms means is not, in general, something that can come totally unstuck from our use of that term, and as such, how we would apply or withhold the term across a range of hypothetical cases may be better viewed not as *claims* about how the term might apply in those situations, but as *constraints* on any account of that term's meaning. Such a reconfiguration of the relationship between intuitions and the relevant truths can become attractive once we see certain problems that can afflict the evidentialist view. For if intuitions are supposed to track the truths, then the question arises: how is it that such a tracking relation is maintained?[4] And the experimental restrictionist challenge raises the worry that that relation is, in fact, not maintained. But the constitutivist response seems to foreclose on the possibility of any radical breakdown of that connection. It is hard to miss when you first wait to see where the dart goes, and then draw a bullseye around it. If intuitions constitute the class of things to which a philosophical theory must be true, it follows straight away that those intuitions cannot be faulty in the way that the restrictionist challenge suggests. If intuitions are not faulty—are *necessarily* not faulty—then there is no reason to worry about the effectiveness of the armchair methodology driven by those intuitions.

This type of constitutivist response to the restrictionist challenge, however, comes at a cost. If intuitions are both strictly constitutive *and* various

---

[4] Posing this question is a central concern of (Goldman 2007), and it motivates his rejecting intuition as a source of evidence for extra-mental reality.

(as indicated by the experimental findings), then there must be a similar variety of philosophical concepts. For example those who disagree about the attribution of knowledge in a Gettier case must thereby mean different things by 'know'. In short, a strict constitutivism guarantees the accuracy of intuitions at the cost of considerable additional complexity in our philosophical theories. Instead of a single account of knowledge, reference, etc., we will have several, indeed perhaps a great many. And, more to the point here, a completed epistemology would thus depend on the contingent matter of who has which intuitions—not a question that can be answered from one's own armchair.

1.2   A Better Response: Loose Constitutivity

Although one could perhaps choose to embrace the extreme relativism thereby entailed, an attractive alternative is to develop a variety of constitutivism that does not require that every individual's intuitions are singularly and definitively constitutive. Such a view might be consistent with both the variety of philosophical intuitions and there being nonetheless a single concept under consideration. Call this a "loose constitutivism", to be contrasted with the "strict constitutivism" that we are suggesting is unpalatably relativistic. Such an account can be found in the work of Brian Weatherson (2003, 10), who points out an additional benefit of avoiding strict constitutivism: we thereby would avoid the odd state of affairs in which we turn out infallible about our own conceptual maps.

Weatherson develops an account of the role of intuitions in philosophical theorizing with the primary explicit aim of mitigating the force of counterexamples in epistemological theorizing in order to make such theorizing more tenable. As a result using Weatherson's work in the context of the empirical restrictionist challenge constitutes a substantive redeployment, or perhaps an exaptation of Weatherson's work, for Weatherson's goal does not explicitly include a defense of the armchair against the restrictionist challenge. Nonetheless, he does seem interested in defending something very much *like* current philosophical practice, so we hope we are not making *much* of a distortion of his views by utilizing them in this way, but we want to emphasize that at least some of the criticisms that we will be developing here may still be entirely consistent with Weatherson's main arguments and goals. With that disclaimer in place, in the balance of this section we will set out a Weathersonian account of loose constitutivity with reference to Weatherson's own example concerning Gettier cases and the JTB account of knowledge.

Recall the situation the loose constitutivist confronts. Intuitions on some

topic of philosophical interest are constitutive of the concept[5] but they also vary in a variety of ways. How could such a collection pick out a single property? The trick is to consider which, among all the properties that at least roughly answer to the set of intuitions, are the most *natural* ones. Given some collection of intuitions, what they all refer to is the particular property that best balances between maximizing closeness to the collection, and metaphysical naturalness. This framework offers the promise of accommodating the diversity and instability of intuitions within a single philosophical theory, because *different* sets of intuitions might all nonetheless be maximally close to the *same* natural property. The referent property must be one whose extension the intuitions do a more-or-less good job of mapping, and this gives the theory its constitutivist element. However, the referent property can also get arbitrarily far away from any one intuition, or any sufficiently small and/or unsystematic subset of intuitions, so long as that loss is compensated by a commensurate increase in naturalness, and this factor allows the theory to be only *loosely* constitutivist.

The basic Weathersonian strategy clearly depends on having at least some sort of answer to the question: what makes a property natural? Characterizing (in the sense of providing a clear example of the kind) the notion is pretty straightforward, though a definition has proved elusive; on this topic Weatherson (2003, 11) refers us to Lewis, who in turn switches between treating 'naturalness' as primitive and offering accounts of the term that are sufficiently challenging to be the subject of substantial on-going exegesis (e.g., Lewis 1983, 1984, 1992, 2001). The general "feel" of the notion is often cashed out in terms of the notion of 'carving nature at its joints', and it is something that the property *green* has a fair amount of, though not as much as *electron*, and vastly more than *grue*. Natural properties mark off objective resemblances in the world, and are the sort of things appropriate to figure in inferences that require projectible predicates. Though this is a far from precise characterization, it is perhaps enough to let us begin to canvass how it might begin to serve the role needed in a loose constitutivist approach to armchair philosophy.

Consider the various mostly-overlapping patterns of intuitions that different people might have about what cases fall under the concept KNOWS. Loose constitutivists would need to ask themselves how such sets of intuitions map onto the landscape of properties and the varying degrees of naturalness of those properties. A number of outcomes are possible in principle.

---

[5] Weatherson frames his discussion more as a matter of assigning properties as the meanings of terms in a language, but we do not think that anything in our discussion here will turn on the distinction between concepts, and terms in a language, or indeed any other similar way of carving up this territory.

It could be the case that there is simply no natural property in the vicinity of any of these clusters of intuitions, in which case it will turn out that KNOWS-thoughts are non-referring for members of that community. Alternatively, it may be the case that there are a number of equally-maximally-natural properties in the vicinity. Should that occur, then it might be that this particular pattern will lock onto one such property, whereas another, only slightly different pattern might lock onto another. In this case it could turn out that the diversity of intuitions argued for by experimental philosophers would reveal that there are distinct knowledge concepts out there. A third possibility, the one preferred by Weatherson and the one which makes loose constitutivity so attractive as a response to the restrictionist challenge, is that there will be a single maximally natural property in the vicinity of *all* of these clusters of intuitions, and that will therefore be the property picked out by the relevant sets of intuitions.

2.   Troubles for the Natural Armchair

Let us grant, for the sake of exploring the promise here offered by loose constitutivity, that it is this possibility of convergent reference that obtains. We now face two methodological questions, if we wish as philosophers to uncover what property it is that is the shared referent. First, how do we decide, for some case that is a matter of particular theoretical concern or otherwise contested, whether it itself really is a case of knowledge or not? We cannot simply trust our intuition about the case and let that be the end of the story, for the kind of validation of our intuitions that loose constitutivity provides is both *rough* and *generic*. The validation is rough, in that we know at most that any given intuition is somewhat likely to be true, but there is no intuition that is guaranteed true. And it is generic, as it applies to all of the intuitions equally. We have no useful suggestions as to how, within the framework of loose constitutivity, we can extract a more precise degree of validation for our intuitions themselves. If intuitions carried with them any sort of phenomenological or other indication of their degree of divergence from the natural, then of course we would have a way of weighting our intuitions accordingly. But that is just the wheat-from-chaff problem that we were hoping loose constitutivity would allow us to avoid having to confront.

Of course, if we could indentify the correct *theory* of knowledge then we could resolve the status of particular cases.[6] Achieving that identification constitutes the second methodological challenge for loose constitutiv-

---

[6] Note that this increased reliance on philosophical t is, in its way, a feature and not a bug of loose constitutivity. Part of what Weatherson is looking to do is pry philosophical practice away from what he sees as an over-reliance on counterexamples.

ity. For each set of intuitions we may find a philosophical theory jostling for acceptance that maps it precisely. But this brings us to our second methodological question: how do we tell which of these competitor theories is *the* correct one? The constitutivity move just tells us that none of them will likely be radically wrong, but it does not by itself tell us which one is actually right. Something more is needed to take us from candidates assembled by intuition to the unique winner of the relevant philosophical debate. Given that we want a theory of knowledge that will select among the various contenders, how do we go about choosing it?

2.1   Detecting the Natural

Weatheron's version of loose constitutivity has further resources for making use of intuitions, as roughly and generically ratified as they are. So far, the only candidate armchair-deployable methodological resources that we have considered are the individual philosopher's intuitions about cases themselves. But suppose that we could *also* tell from the armchair which, among candidate sets of competitor analyses, is the most natural? And suppose further that our armchair epistemic purview included how to make trade-offs between fit of intuitions and degree of naturalness, in selecting between such competitors? Under such suppositions, we might have ready-to-armchair all the tools we would need to select a *best* theory.

We contend that only if the following two significant claims are true, then we may be able generally to recover philosophers' intuitive practices. The first is a metaphysical claim, about the relation between more and less fundamentally natural properties:

> *Constructability (C)*   *Philosophically important properties are natural in virtue of their location in a matrix of relations to other, more natural properties.*

If (C) holds, then the philosopher analyzing some concept can seek out the more natural properties that underwrite the naturalness of the property that concept picks out, and in virtue of which, presumably, we are able to recognize such properties (e.g. knowledge) in the real and hypothetical cases where they obtain. The more natural the construct, the better the theory. Weatherson suggests that theories involving comparatively short constructions of fairly natural properties are thus to be preferred over other theories (Weatherson 2003, 8–9).

For any of the above to be of methodological import, though, a further claim needs to be made:

> *Detectability (D)*   *For any two competitor analyses C1 and C2 for a philosophically important property, we can discern from the*

armchair which is the more natural.

(C) tells us that an analysis is possible, but we need (D) to render it likely that we can tell easily enough from our armchairs, without special scientific investigation or anything like that, how different proposed analyses rate in terms of their naturalness. (D) would be clearly true were we to have some access to some canonical list of the natural properties. Sadly though, we have no such list. So (D) will as a matter of fact require some other way of distinguishing the natural from the unnatural—or, more typically, the more-or-less natural from the just-plain-less natural.

So, (C) and (D) together would enable philosophers to infer from a pattern of intuitions involving a concept to a theory that captures how that concept is situated in the constellation of the natural.

The reader will perhaps not be surprised that our intention here is make trouble for both (C) and (D).[7] We will raise doubts about both *seriatim*, and then raise a further concern that even if (C) were true, it would likely thereby make it likely that (D) is false.

2.2   Contra (D)

Here is what the question at hand is not: can we *ever* tell, comparing two conceptual matrices, which one picks out the more natural property? We probably do have at least *some* general sense for such comparisons; if we had no such ability whatsoever, Lewis' and Weatherson's discussions would not even have gotten off the ground. And surely everyone not suffering from a fit of Goodmania can agree that *green* is oodles more natural than *grue*, if anything is. But what philosophers are going to need, if loose constitutivity is to have armchair-preserving methodological consequences, is something rather more demanding. Reflect on the theories of knowledge that epistemologists are looking for help in choosing between. We claim that it is not at all manifest to this rough-and-ready sense of the natural just which of tracking; safety; justification and truth; justification and truth and some further anti-Gettier-condition; etc. is most natural. Even if we can sometimes just plain see which of two constructs is more natural, *this* task far outstrips *that* capacity. If (D) is to hold true here, we will need something more to appeal to than just our most basic sense of naturalness. That basic sense fails, so to speak, to measure naturalness out to as many significant digits as the theoretical task demands.

[7]   Let us be clear—our goal here is *not* to problematize naturalness itself but rather its ability to do a very particular philosophical job; that of underwriting something like the current philosophical practice of armchair analysis.

So let us survey what available materials might possibly be adverted to, in addressing a question of comparative naturalness. One can divide the relevant naturalness-conferring aspects of a property into *inherited*, *structural*, and *holistic* factors. Inherited factors are those that it receives simply from its relation to the other properties that are configured in the matrix of its analysis. For example, one might feel that a property cannot be more natural than the least natural property it is configured in terms of. Structural factors derive from the structure of the locating matrix. For example, overall logical complexity may generally be inversely correlated with naturalness. Finally, holistic factors are those that apply only to the property as a whole, and not in virtue of the character of its matrix. For example, there may well be laws at higher levels than fundamental physics. If that is so, then a complex matrix may configure a property that figures in such laws, and it would attain some amount of naturalness in virtue of that nomicity. Suppose there were laws about red apples that did not derive from the laws pertaining to red things and to apples more generally; and suppose there were no such laws about green apples; and suppose that *red* and *green* are equally natural. Then *red apple* would be more natural than *green apple*, even though the inherited and structural determinants of their respective naturalness would be equivalent.

Given these factors, we can see how various sorts of more fine-grained comparative naturalness judgments would be possible: cases where all the possible factors except one are equivalent, and where furthermore there is a clear difference in the remaining factor. Just for the sake of illustration, the following seem to us to be *plausible* pairwise naturalness judgments, though we certainly do not take ourselves to be committed to them being right: *has a charge of +1e* may be more natural than *has a charge of +1 coulomb*, because *e* (the fundamental charge) is itself a more natural unit than is the coulomb. Moreover, *has a charge of +1e* may be more natural than *has a charge of +1e or +3e*, because the latter has greater structural complexity. And, even assuming that 44 and 43 are equally natural, *has a nucleus with a charge of +44e* may count as more natural than *has a nucleus with a charge of +43e*, even though all their parts are equally natural and they are structurally identical, because technetium (which, it is our understanding, only occurs synthetically) may count as less natural than ruthenium (which occurs, well, naturally).

Unfortunately, it is rare for the armchair philosopher to be comparing two so closely matched concepts. For example, as noted above, *justified true belief* and *true belief produced by a reliable mechanism* are just not obviously comparable by means of these resources. Nor is *true belief produced by a mechanism operating in accord with its proper function*. And so on. Again, even granting these more refined pairwise evaluations of naturalness, in many of the places where we would actually like to be able to make a com-

parative naturalness judgment, our powers to do so simply pull up short.

And it gets worse; the state of the art just does not tell us how to make careful trade-offs between these various factors. After all, *grue* isn't just supposed to be less natural than *green*—it's supposed to be less natural than *almost anything else at all*. Yet *grue* is a very simple construct from a small handful of fairly natural properties. (Even more so if one considers the commonly-mistaken variant, *green before time t or blue afterwards*, dropping out Goodman's original formulation involving observation. Our gut sense is that this is only marginally more natural than *grue*, if at all.) The analyst may of course attempt to find other factors to appeal to, and it would be a fascinating project to devise a plausible scheme for quantifying the various factors and weighing them against each other. But such methodological resources are not presently at hand in philosophical practice.

Finally, it seems that many cases of the holistic naturalness of a property will simply be invisible from the armchair. Our elemental example above is such a case. If many properties are ones whose degree of naturalness cannot be discerned from the armchair, then this entire methodological gambit may fail. Indeed, as we shall now argue, holistic naturalness poses a fundamental problem to using the Weatherson/Lewis framework to defend the armchair.

2.3   Contra (C)

(C) claims that the naturalness of any non-fundamental philosophical property is derived from its configuration in terms of other, still more natural properties. For (C) to be false, then, would mean that some natural properties that are not *maximally* natural, nonetheless are themselves *primitively* natural; their naturalness is not inherited from the naturalness of some properties that are in turn more natural than they are. That is, for (C) to be false, there must be philosophically-relevant properties which have a high degree of holistic naturalness. We acknowledge that, assuming that at least some minimal sort of supervenience holds, there will be interesting questions about how that property is related to arrangements of the world specified in terms of more fundamental properties. But our most important theorizing of such a property would have to look more like our theories of fundamental natural properties—more in terms of its relations to other properties at its own level, less in terms of relations to lower-level properties. (Of course, theories of the most fundamental properties will have exactly *no* features cashed out in terms of relations to lower-level properties.)

The existence of such properties as the would-be targets of philosophical analyses, ones that mark irreducible resemblances, would falsify (C). But do we have any reason to think that there are any such properties? Arguments

from the special sciences literature provide just such a reason.[8] Looking at the sciences themselves reveals that the laws governing such properties as *believing* and *desiring* (or, for that matter, such properties as *being a geological plate*) simply cannot be replaced with generalizations about lower-level properties and constructions thereof (Fodor 1974). In fact, all we need here is something somewhat weaker that what Fodor usually claims: we could grant that there is some, albeit highly contorted, identity of belief with the disjunction of its possible instantiators, but still insist that the unwieldy disjunction fails to capture the *resemblances* involved. (See Antony 2003.) From the point of view of the lower-level properties, belief is a wildly gerrymandered affair. And so, from the point of view of the lower-level properties, belief would be just as natural as any other equally-gerrymandered construct; which is to say, not very natural at all. Indeed, you might even be able to make a much *more* natural property by cutting off aleph-nought or so of the disjuncts. Rather than such an extreme collapse of naturalness, we think it more likely that properties of central philosophical interest will turn out to be holistically natural, and not constructable in the manner Weatherson's account would require.

2.4   The incompatibility of (C) and (D)

The last argument we wish to offer is not, in fact, an argument to show that one or the other of (C) and (D) is false, but rather that they cannot both be true. In particular, although the analyst needs both (C) and (D), we will argue here for the conditional: if (C), then not-(D).

Suppose that (C) is true. That means that how natural any given philosophically important property X is depends on its correct analysis; in particular on the inherited and structural components of such an analysis. But note that *(C) will very likely apply again to the components themselves.* How natural each component of X ($x1$, $x2$, $x3$…) is will depend on each of *their* correct analyses. And of course how natural the components of each such analysis will depend on *their* analyses…and so on, until we reach an exhaustive analysis in terms of concepts of fundamentally natural properties. Only at that point will the requisite comparison of naturalness be possible, since only at that point will we truly be able to see what the inherited and structural contributions to its overall naturalness may be. But we take it that such an exhaustive level of analysis is one that is far out of practical philosophical reach, *a fortiori* out of reach of the armchair. Attempts to use a naturalness

---

[8]   A Weathersonian analysis can deal with a handful of very special things having this character (e.g. truth; Weatherson 2003, 28). However, it can *not* accept the quantity of primitively natural entities suggested by the Fodorian perspective. Bear in mind *that lots and lots* of things might be primitively natural in this way.

metasemantics to defend the armchair will have to thread a very tight needle here: they need enough by way of constructability to motivate philosophical analyses in the first place, but not so much that the raw materials for such constructions fly completely beyond our seated grasp.

This problem can be illustrated by considering how Weathson's analysis of the naturalness of theories of knowledge interacts with his prime candidate for basic naturalness: physical fundamentality. As Weatherson writes:

> I think we can dispense with [the claim that knowledge is primitively natural] rather quickly. It would be surprising, to say the least, if knowledge was a primitive relation. That X knows that p can hardly be one of the foundational facts that make up the universe. If X knows that p, this fact obtains in virtue of the obtaining of other facts. We may not be able to tell exactly what these facts are in general, but we have fairly strong opinions about whether they obtain or not in a particular case. This is why we are prepared to say whether or not a character knows something in a story, perhaps a philosophical story, without being told exactly that. We see the facts in virtue of which the character does, or does not, know this. This does not *conclusively* show that knowledge is not a primitively natural property. Electrical charge presumably is a primitively natural property, yet sometimes we can figure out the charge of an object by the behaviour of other objects. For example, if we know it is repulsed by several different negatively charged things, it is probably negatively charged. But in these cases it is clear our inference is from some facts to other facts that are inductively implied, not to facts that are constituted by the facts we know. (Only a rather unreformed positivist would say that charge is *constituted* by repulsive behaviour.) And it does not at all feel that in philosophical examples we are inductively (or abductively) inferring whether the character knows that p. (Weatherson 2003, 22–23; italics original)

We have already indicated one way in which this argument for (C) (at least with regard to *knowledge*) might be insufficient: if there are special-science laws in which *knowledge* figures. (There may also be other ways that holistic naturalness may arise, such as thorough entrenchment in a folk practice.) But let us suppose, as Weatherson does here, that any sort of property that occurs only in virtue of other properties' occurring, is only *natural* in virtue of its relation to those other, more natural properties. The problem for Weatherson is that at least *believes* and *justified* are such properties as well. Whatever naturalness they may have, they will only have in virtue of whatever properties they can be analyzed in terms of. And so on. But once our basic sense of naturalness runs out it will quickly become practically impossible to discern how much naturalness is possessed by different complex

concepts.[9]

For example, it is unclear even that we can discern which of JTB and JTB+ (i.e., JTB with some additional Gettier-eliminating clause) is more natural. One might have thought that *that* comparison would be as close to a straightforward pairwise judgment as we are likely to find between real competing analyses in epistemology. Furthermore Weatherson takes it pretty much for granted that JTB is the more natural of the two, and the operative question for him becomes whether or not that increase in naturalness can sufficiently offset the loss of matching up with our intuitions about Gettier cases. Nonetheless, we contend that the *complete* analyses of JTB and JTB+ could well reveal the latter to be the more natural. For any nontrivial "+", JTB+ will be a stronger condition than JTB—everything that is JTB+ is also JTB, but some things are JTB without being JTB+. But that means that the total set of physical realizer states for JTB+ will be smaller than the total set of physical realizer states for JTB. Thus, if it turns out that the ultimate version of the analyses are vast sets of disjuncts, enumerating all the possible instantiations of JTB and JTB+, then there will be *fewer* disjuncts in the analysis of JTB+. So JTB+ will be more natural, according to the factor of logical complexity.

Alternatively, if a physical-fundamentality construal of (C) does not force matters all the way down to sets of realizer states, that would seem to mean that there are intermediate levels at which there is sufficient holistic naturalness to stop such a maximal analysis. But then we would face a different challenge to (D): is there any reason to be confident that the analysis of JTB would square better than JTB+ with such holistically natural intermediates? It might be that some additional conditions will combine with justification, belief, and/or truth in some way to produce a more natural property than any of the JTB properties have for themselves. E.g., maybe *beliefs produced in accord with the proper function of the agent's belief-producing capacities* will prove more natural than *just* belief simpliciter. That is, perhaps the various sorts of events and states of our neurological processes that occur when our cognitive hardware is operating as designed by natural selection have more in common with each other, than the whole set of events and states involved in belief-producing processes do in general; and then perhaps this internal similarity is sufficient to render the property shared only by that subset of

---

[9] By way of diagnosis, we suspect that Weatherson's account requires that two distinct metaphysical issues run together: the supervenience of one property's tokenings on the tokenings of other properties, and the supervenience of one property's naturalness on the naturalness of others. Our objection is in part that a property can supervene on physically more fundamental properties in the first sense, without its naturalness doing so as well. This may just be part and parcel of being a nonreductive physicalist, for example.

beliefs a more natural property than beliefhood itself. Maybe; maybe not. The devil will be in the details—and these details are not ones that can be detected from the armchair.

3.   Another path to loose constitutivity?

Given the inscrutable nature of the natural, one might wonder whether it is the only way for a loose constitutivist to go, or whether some other method-ological resource can be mustered to help select between candidate philo-sophical theories.

Henry Jackman (forthcoming) offers one such approach, deploying the *principle of charity* in an attempt to achieve the twin desiderata of loose con-stitutivity.[10] Philosophically undesirable sensitivities of the sort documented by experimental philosophers are just something to be interpreted around, without posing a threat to the interpretations themselves. Inter- and intrap-ersonal diversity will still be consistent with a general convergence of best in-terpretations of various persons and their timeslices (though Jackman does acknowledge the possibility that Easterners and Westerners may, as would be consistent with some recent experimental results, have divergent seman-tics for their terms).

So far, so good—but it is only so far. We still have the same rough-and-generic ratification problem that faced the naturalness-based account; so we likewise must ask what may be appealed to in order to select between differ-ent competitor theories.

Mere internal coherence will not do. There are many different internally-coherent stories that can be told about our intuitions, each one making sense of a different pattern of attributions. All of the different epistemological the-ories on offer are at least in the running, for starters. So how do we decide what 'knows' means in the mouths of those who do not share Gettier intu-itions, and whether it is the same as what it means in the mouths of those who do? Our concern is that there will be, in Jackman's Davidsonian frame-work, no resources available from the armchair to answer that question.

For if we are to maximize not just internal coherence, but *truth*,[11] then[12]

---

[10] These desiderate are: first, our intuitive applications of terms are part of what fixes the meaning of those terms (and so while we cannot be construed as wildly and across-the-board wrong, yet at the same time some amount of error or confusion is surely to be en-countered), second, acknowledging "noise" in our intuitions allows us to avoid massive relativism about philosophical concepts.

[11] or knowledge; see (Williamson 2008, ch. 8).

[12] Jackman (personal communication) objects that maybe there really could not be rival com-petitors that were all truly coherent, especially given strong weighting to our most central commitments. We suspect that part of the problem may be, however, that so many cases of philosophical import are not really of much import to the speakers we are interpreting. So

we will need to know what the truth is, in a way that will be decidedly un-armchairish. Consider a 'whale'-utterer. What content we should attribute to them will depend very significantly on *what we take to be true about whales*. And, as our armchairs are (thankfully) not located on the poop deck of the *Pequod*, we take it that very little indeed about whales can be discerned from the armchair; we are better off consulting with biologists.

Moreover, even that might not be enough, if somehow it is granted that substantial scientific knowledge is part of the armchair. Here is Davidson himself on the limits of interpretation:

> If a speaker utters the words, 'There's a whale', how do I know what he means? Suppose there is an object that looks like a whale in the offing, but I know it is not a mammal? There seems to be no abso-lutely definite set of criteria that determine that something is a whale. Fortunately *for the possibility of communication*, there is no need to force a decision. Having a language and knowing a good deal about the world are only partially separable attainments, but interpretation can proceed because we can accept any of a number of theories of what a man means, provided we make compensating adjustments in the beliefs we attribute to him. (Davidson 1980, 257; emphasis added)

For the purposes of communication, this degree of indeterminacy about meaning may be absolutely fine; but for the purposes of philosophical anal-ysis, it is a disaster. The reason that philosophers appeal to the sorts of in-tuitions that they do is to pull apart the different theories about knowledge, freedom, and so on—we take the philosophical truths to be determinate in a way such that all these different theories are *competitors*.[13] But it sounds as if, given how subtle the differences are between them, the principle of charity may not license our preferencing one over the others. Perhaps some would be happy to embrace the kind of quietism about much of contemporary epis-temology and metaphysics that results—but that constitutes a more radical alteration to philosophical theorizing than that aimed for by the experimen-tal restrictionists!

Perhaps naturalness could come in through the back door here. For a sufficiently large corpus, including also the deferential dispositions of the speakers, the charitable reading might have to converge with the natural reading, so long as the interpretee has any interest in and commitment to carving nature at its joints. But such an approach obviously would run right smack into the difficulties we presented in §2.

---

there are cases that philosophers are leaning on heavily, but about which speakers really are basically uncommitted. But Jackman may not be looking to defend such corners of armchair practice.

[13] We put aside current debates concerning the "merely verbal" nature of some philosophical disputes. But see (Chalmers et al. 2009) for a discussion of these issues.

Another possible response: Bealer argues that many concepts of philosophical interest do not have scientific essences, but are "semantically stable" (see Bealer 1999). The issues here are different from the one he is wrestling with, but there may be this similarity: maybe only armchair-accessible truths are relevant for interpretation of concepts of philosophical interest. We think this might be pretty unDavidsonian (but then again perhaps charity *does* begin at home!), but it is not an obviously incoherent position. Such a move however *drastically* reduces the amount of data that can be appealed to in selecting one interpretation over another, with an attendent increase in the degree of indeterminacy. Can this be compensated for by the generous use of hypotheticals? That is, suppose we consider a large range of hypotheticals, in each of which we just stipulate all the contingencies, and then see whether interpretees do or do not extend their concepts? We could try that—but then we're back at putting all the weight on the intuitions! Such an outcome would mean that, at worst, loose constitutivity does nothing to address the restrictionist challenge; at best, we would still be in need of some further armchair-accessible methodological resource for preferring one competitor theory over another. And that's a need that we have not yet found a way to satisfy.

4.   Conclusion

Loose constitutivity theories offer not just promising metasemantic accounts, but also suggest an interesting line of response to the experimental restrictionist challenge. We offer no objections here to the first aspect of such theories, but have been concerned that, since they offer only rough and generic ratification of our intuitions, it can only do methodological work for us if it operates in conjunction with some sort of further criteria of theory selection. We have considered two that we find in the literature: naturalness (Weatherson, borrowing from Lewis) and charity (Jackman, borrowing from Davidson), and both seem to have problems. If you've got an idea for another candidate, please let us know.

Acknowledgements

Bibliography

Alexander, J. and Weinberg, J. (2007). Analytic epistemology and experimental philosophy, *Philosophy Compass* **2**: 56–80.

Antony, L. (2003). Who's afraid of disjunctive properties?, *Philosophical Issues* **13**: 1–21.

Bealer, G. (1999). A theory of the a priori, *Philosophical Perspectives* **13**: 29–55.

Chalmers, D., Manley, D. and Wasserman, R. (eds) (2009). *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford University Press, New York.

Davidson, D. (1980). *Essays on Actions and Events*, Oxford University Press, New York.

Fodor, J. (1974). Special sciences, *Synthese* **28**: 97–115.

Goldman, A. (2007). Philosophical intuitions: Their target, their source, and their epistemic status, *in* C. Beyer and A. Burri (eds), *Philosophical Knowledge: Its Possibility and Scope*, Rodopi, Amsterdam, pp. 1–26.

Jackman, H. (forthcoming). Semantic intuitions, conceptual analysis, and cross-cultural variation. To appear in *Philosophical Studies*.

Knobe, J. and Nichols, S. (eds) (2008). *Experimental Philosophy*, Oxford University Press, New York.

Lewis, D. (1983). New work for a theory of universals, *Australasian Journal of Philosophy* **61**: 343–377.

Lewis, D. (1984). Putnam's paradox, *Australasian Journal of Philosophy* **62**: 221–236.

Lewis, D. (1992). Meaning without use, *Australasian Journal of Philosophy* **70**: 106–110.

Lewis, D. (2001). Redefining 'intrinsic', *Philosophy and Phenomenological Research* **63**: 381–398.

Nichols, S. and Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions, *Noûs* **41**: 663–685.

Nolan, D. (2009). The a posteriori armchair.

Sosa, E. (2006). Experimental philosophy and philosophical intuition, *Philosophical Studies* **132**: 99–107.

Sosa, E. (2009). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Vol. 1, Clarendon Press, Oxford.

Swain, S., Alexander, J. and Weinberg, J. M. (2008). The instability of philo-

sophical intuitions: Running hot and cold on truetemp, *Philosophy and Phenomenological Research* **76**: 138–155.

Weatherson, B. (2003). What good are counterexamples?, *Philosophical Studies* **115**: 1–31.

Weinberg, J. M., Alexander, J., Buckner, C. and Gonnerman, C. (forthcoming). Are philosophers experts? *Philosophical Psychology* (in press).

Weinberg, J. M., Nichols, S. and Stich, S. (2001). Normativity and epistemic intuitions, *Philosophical Topics* **29**: 429–460.

Williamson, T. (2004). Philosophical 'intuitions' and skepticism about judgement, *Dialectica* **58**: 109–153.

Williamson, T. (2008). *The Philosophy of Philosophy*, Wiley-Blackwell, Hoboken.