Experimental Philosophy, Conceptual Analysis, and Metasemantics*

Jason Turner

In David Rose (ed.), Experimental Metaphysics, *Bloomsbury.*

Empirically informed philosophy is nothing new. Despite aspersions about armchairs, the best philosophers in every era have made serious efforts to produce philosophy informed by the science of their day. In the last decade and a half, however, a new form of empirically informed philosophy has arisen: *experimental philosophy*, which probes folk judgments about philosophical concepts.

Why probe folk judgments? The basic idea runs something like this: When we theorize, we use thought experiments. In these experiments we make snap judgments about philosophically interesting concepts. These judgments are supposed to provide a sort of pre-theoretical constraint on theorizing; if a proponent of a counterexampled theory rejects the counterexample by appeal to the theory itself, he's not doing philosophy right. But if the judgments are supposed to be *pre-theoretical*, then they had best not be subtly influenced by other theoretical commitments. Since philosophers have had long exposure to the relevant theory, maybe their judgments shouldn't be relied on either. Better instead to rely on 'the folk', people with no prior philosophical exposure to taint their judgments.

Foes of experimental philosophy often object to this line of thought as follows: "If we probe the folk about, say, free will, we get judgments about what the folk think free will is like. But why do we care about *this* at all? We want to know what free will *really is* — not just what people *think* it is. After all, we wouldn't dream of probing folk judgments about simultaneity in a quest to determine the truth of special relativity. The theory of special relativity is a theory about what time is *really* like, and its truth or falsehood doesn't depend on what ordinary people think. Why should free will be any different?"

Experimental philosophers have often responded *tu quoque*: If, like time, the nature of (say) knowledge or free will isn't constrained by pretheoretical judgments, why do philosophers persist in appealing to their own judgments about cases while theorizing? But while the *tu quoque* may be dialectically effective, it is philosophically unsatisfying. It would be nice to have some sense as to *why* pretheoretical judgments have probative force when it comes to free will but not when it comes to time.

My aim here is to sketch a picture of philosophical theorizing that can answer that question. The idea, roughly, is that what free will or time *really* is just is whatever content our concepts of *free will* or *time* actually pick out. And what these

*Thanks to Ross Cameron, Mark Balauger, Elizabeth Barnes, Michael McKenna, Shaun Nichols, Robbie Williams, and especially Uriah Kriegel for helpful comments and conversation.

concepts pick out will ultimately be determined by a blend of our own cognitive and linguistic behavior with how the world is. Since our cognitive and linguistic behavior partially determines what our concepts pick out, judgments (whether folk or our own) matter. But since that behavior isn't the whole story, the judgments aren't the whole story, either; and sometimes when the world is recalcitrant (as it is when it comes to simultaneity) our concepts pick out things that violate these judgments.

1 Concepts and Conceptions

Let's start, as philosophers often do, with a distinction. This one will be between *concepts* and what I will call *conceptions*. Our concept of a thing is the mental vehicle for our thoughts about that thing, whereas our conception is made up of beliefs about what that thing is (essentially) like. My concept *horse* is a vehicle for my thoughts about horses. My *conception* of horse, on the other hand, is a cluster of thoughts about horses, such as that they are animals, have four legs and a mane, and so on. (Of course, not every thought I have about horses is part of my conception. I think that Seabiscuit is a horse, and this is of course a thought about horses, but it is in some sense incidental to my conception. Roughly, my conception consists of the features I take horses to generally have in common, not just actually but in any situation I can conceive of.)

Philosophers often engage in what they are pleased to call 'conceptual analysis', giving necessary and sufficient conditions associated with a concept. But this project can itself be understood in one of two ways. In what we might call the *descriptive* way, the analysis merely unpacks our common *conception* of something. On this way of thinking, when we analyze 'bachelor' as 'unmarried eligible male', we find out what our conception of bachelors is like. In other words, we find out *how people think* about bachelors — what features people think someone has to have in order to be a bachelor.

On an alternative *content* picture of analysis, analyses give necessary and sufficient conditions for something to *fall under* one of our concepts. On this way of thinking, when we analyze 'bachelor' as 'unmarried eligible male', we find out which people *in fact* fall under our concept *bachelor*.

If it turns out that something falls under one of our concepts if and only if it matches our conception, then the two projects collapse into one. But the projects remain distinct if our conception comes apart from the content of our concept. Familiar cases involving so-called 'natural kind terms' show that these projects can, in principle, come apart. Suppose that (to take one of Putnam's (1962) examples), unbeknownst to us, the entire species of cats are in fact a complex form of robot planted here long ago by aliens. A 'descriptive' conceptual analysis of *cat* will presumably say (among other things) that cats are animals. But the things that actually fall under our concept ‘cat’ won’t be animals; they’ll be robots. So a correct ‘content’ analysis of ‘cat’ will come apart from the descriptive analysis.

It would be nice if, in general, content and descriptive analyses always agreed with each other. Thanks to the work of Kripke (1972) and Putnam (1975) we know they won’t. Still, we might hope that cases where they come apart will be limited to a special kind of ‘natural kind’ concepts. We can then set these concepts up on a shelf and get back to analyzing the rest. Descriptive analyses can be done (more or less) from the armchair, and since the concepts are determined by the conceptions, we get content analyses in the bargain, too.

Unfortunately, I doubt we can hive off natural kind terms in this way. Consider the concept of *solidity*. While it is a concept that applies to things in the natural world, it doesn’t seem anything like a ‘natural kind’ — being solid isn’t relevantly similar to being water, or being a tiger, or so on. But there is good reason to think that the content of *solidity* comes apart from our conception of it. For there is good reason to think that our conception of *solidity* includes, as a necessary condition, that solid things be devoid of empty space. Contemporary science tells us (more or less) that most things we would think of as solid — tables, rocks, and so on — are in fact made up largely of empty space. They consist of lots and lots of tiny particles which are at a vast distance from each other in proportion to their size.[1] But contemporary science does *not* tell us that tables and rocks aren’t solid, or so it seems to me. It tells us, rather, that solidity is much different than we thought it was. It tells us, in fact, that our *conception* of solidity was wrong.

Of course, if the world had been much different — if we had lived in a world where rocks and tables and so forth consisted of matter spread continuously throughout a region — then presumably our concept of *solidity* would have picked out all and only the things that matched our conception of it. That our conception doesn’t match what falls under the concept depends, in part, on what the *world* is like.

The lesson? Even setting aside ’natural kinds’ and their ilk, we cannot in general move from a descriptive analysis to a content one. That is: We cannot infer from our *conception* of something that all and only things matching that conception fall under the concept. Maybe sometimes they do; but whether they do or not may very well depend on whether the world has been kind to us. And until we know whether the world has been kind to us, we won’t know whether our conception matches the content of our concept.

[1] At least, science used to tell us this; I’m not so sure it tells us this anymore, thanks to the difficulty making sense of ’empty space’ when a particle is superposed across it. For the purposes of this example, though, we will pretend contemporary science really does tell us this.

2 Ideal Interpreters

2.1 *The View*

That, anyway, is the phenomenon. What's the theory behind it, though? If the content of my concepts aren't automatically determined by my concepts but depend upon the world, how do they know what part of the world to pick out?

The driving idea behind the theory I'm about to sketch is a kind of functionalism about mental content.[2] Roughly, the thought is this: Our concepts are supposed to play a role in our mental lives. But that role isn't wholly confined to the mental; it affects how we interact with the world. I don't just deploy the concept *solid* in order to deploy *other* concepts (such as concepts about empty space and so on). I also deploy it in the presence of certain things (rocks, tables, and so forth). I might do this as input, coming into a new room and deploying concepts as I map my environment; or I might do it as output, deploying concepts in order to decide which piece of furniture to buy. That I deployed a certain concept will be (for instance) part of what justifies certain of my inferences or what rationalizes certain of my decisions.

Whether or not a concept can do this work will depend in part on how it relates to other concepts, but also on how it relates to the world. If my purchase of a hardwood table is rational in part thanks to my deployment of the concept *solid*, then that concept had better cover that table. On this picture, my concepts will have the contents that best fit the functional role those concepts play in my mental life; which potential contents best fit that role will depend not only on what the role itself is, but on what is out there in the world to realize this role.

It isn't easy to think about this functional role directly. One way to think about it *indirectly* is to imagine contents fixed by an 'ideal interpreter': a godlike being omniscient of all non-semantic facts. This interpreter then assigns contents to our concepts by following general rules — rules such as 'interpret people in such a way that their actions are by-and-large rational,' and 'interpret people in such a way that their beliefs are by-and-large true and their errors are understandable and expected.' (Lewis 1974: 112–113) These and other rules together code up the functional role of concepts generally.

The rules can't always be perfectly satisfied. Sometimes, for instance, an ideal interpreter will be unable to interpret an agent's concepts in a way that makes them both rational and responsive to their evidence. In such cases, the interpreter is supposed to assign contents in a way that minimizes departures from rationality and evidence-responsivenes. This reflects the fact that a concept's content is what

[2] The picture is largely derived from Lewis 1974 and 1975, although I will deviate from Lewis's view in ways to be spelled out below. Lewis, of course, was building on the work of Davidson (1974) and Quine (1960).

best realizes its cognitive role, not what *perfectly* realizes it — for it may be that nothing realizes it perfectly.

An ideal interpreter is aware of my conceptions. For instance, she is aware that my concept *solidity* is closely tied up with my concept *empty space*. One of her jobs is to interpret me so that my concepts match my conceptions: she should make it so that, by and large, I think about what I think I'm thinking about. But this is only one constraint among many, and she must best balance out all of these constraints. It may be that, by assigning contents to concepts in a way that completely vindicates my conceptions, she makes me wildly and unnecessarily wrong about lots of other things. Since she has to balance this constraint among others, she would do better to vindicate much (if not all) of my conception while reducing my errors elsewhere.

Arguably, this is what happens in the case of *solidity*. My conception has it that solid things are without empty space. But I also have lots and lots of other beliefs: that my table is solid, that white bread is not, and so on. The ideal interpreter can make me right about what solidity is like only by making me wrong about lots of other things. But she can make me just a *little* bit wrong about solidity, by making me wrong about the no-empty-space constraint — while still making me right about a lot of other things, including my judgments about individual things being solid.

2.2 *Refinements*

First: I have talked as though our conceptions are an all-or-nothing affair. That almost certainly isn't right. We have various beliefs about things; we are more attached to some, and less attached to others. Some seem more central to what we think of as 'a conception'; others less so. Perhaps *what is known is true* and *what is known is believed* are both part of our conception of knowledge, but the former more central, and less negotiable, than the latter. If so, an ideal interpreter will try to preserve both; but if she finds she must make one of them false, she will preserve the first and jettison the second. Likewise, some beliefs may be in the penumbra of our conceptions: Not clearly in, but not clearly out, either.

Second: Interpretation is a *communal* effort. We are not interpreted in isolation; the ideal interpreter wants to make sense of us, of our interactions with each other as well as our interactions with the world we live in, as a community. But concepts are private vehicles. As a result, the interpreter has to do some work to figure out how to coordinate between us, so that we by-and-large share concepts and coordinate on them. Suppose you have a concept c, with an associated conception, and I have a concept c^* with an associated conception. If we each deploy our respective concepts in roughly the same circumstances, and if there is a functional description of your cognitive economy in which c plays roughly the same role as c^* does in a similar functional description of my cognitive economy, the ideal interpreter will interpret your concept *c* as 'the same' as my c^*; and she will extend this throughout the entire interpreted community.

As a result, there may be interpretative pressures on my concept that do not come from my conception. Suppose c^* is a concept I deploy when trying to decide whether to hold someone morally responsible, and is sensitive to features involving whether they were coerced and so on. As a result, the ideal interpreter is liable to think of this as my *free will* concept. Suppose, though, that I do not deploy c^* in a way sensitive to the truth of causal determinism, and it is no part of my conception that c^* should be so sensitive. And suppose further that the ideal interpreter sees that the best candidate for everyone else's *free will* concept *is* sensitive, given how they deploy it, to the truth of causal determinism. She will thus have good reason to interpret my concept c^* as having a content incompatible with free will, despite there being nothing in my head suggesting it get such a content.

Third: As I've presented it, it sounds as though the ideal interpreter simply looks at what people in fact do and say and assigns content on that basis. Not so. The ideal interpreter is omniscient of *all* non-semantic facts, including dispositional ones. These dispositional facts figure in the assignment of contents. For example, as I've described it, the ideal interpreter is under pressure to assign a concept *c* a content which makes confident, unhesitating deployment of *c* accurate. But sometimes we make mistakes. When we discover our mistakes, we retract such deployment. When we fail to discover our mistakes, we do not retract, but we were disposed to upon coming to be better informed. The ideal interpreter ought to balance charity to the deployments we in fact make with charity to our retractions, both actual and dispositional. (Cf. Hirsch 2005: 73–74)

Fourth: The interpretative project, as I've described it, involves the ideal interpreter assigning contents to *concepts* — something more-or-less like words in a language of thought. The original interpretative idea was no such thing; it involved assigning *propositional attitudes* — conceived of as relations to coarse-grained propositions, or sets of possible worlds — to agents. Interpretativists wanted an account that worked for any rational agents whatsoever, and did not want to build it into their account that a concept- or language-of-thought-based mental structure was required for rational thought.

We might hope that, once the interpreter assigns propositions to agents, we could get a correlation between particular propositional attitudes and 'sentences' in a 'language of thought', and then reverse-engineer sub-propositional contents for the concepts in those 'sentences'. This won't work, though; Quine's (1960: ch. 2) indeterminacy-of-translation arguments can be used to show that subsentential content isn't determined by (coarse-grained) propositional content. (Cf. Lewis 1975:

175–178) Something else is needed.

But the original, proposition-first conception of the project isn't well-motivated. The interpretative project is functionalist at heart. In principle, there should be no objection to interpreting sub-agential functional apparatus alongside interpreting the agents themselves. Even stronger: The functional description of us as agents includes not just belief- and desire-talk but concept-talk, too, and connections between the two. The ideal interpreter's job is to find the best realizer of this functional role. If she is interpreting agents whose mental apparatus does not involve concepts, the best realizer of the functional role may leave that bit out and only assign propositional attitudes to agents. But if the agents' mental apparatus *does* have a conceptual structure, the best interpretation should assign contents to the concepts as well as to the beliefs and desires they figure in. If we think with concepts, ideal interpreters ought to interpret those concepts as well.

If the interpretation of concepts can only be settled by reverse-engineering them from propositional attitudes, there will be radical indeterminacy in their contents. But if the ideal interpreter is interpreting concepts *alongside*, rather than *after*, propositional attitudes, the situation is more promising. For instance, the ideal interpreter — knowing all non-semantic facts — will know the causal connections between different concepts and between concepts and the external world. If concept *c* is reliably triggered by and only by proximity to cats, for instance, the ideal interpreter has pretty good evidence that *c* means *cat* or something in its neighborhood.[3]

Granted, this makes the ideal interpreter's job harder. When faced with a community, she must first figure out the computational structure underlying their rational behavior before she can start assigning contents. Harder; but not impossible. Just as a clever engineer might come to understand the computational structure of a computer without knowing what its programs are designed to do, the interpreter can come to understand the computational structure of our brains before figuring out just what each concept is supposed to be about. If such a structure is there, that will be her first step; she can then use what she learns about how our brains compute as part of the data for figuring out what our concepts mean.

[3] This won't be the whole story. There are, essentially, two kinds of semantic underdetermination. One happens at the *propositional* level, and one happens at the *subsentential* level. Propositional semantic underdetermination leaves it open which proposition one is believing. But the point is that, even if that sort of underdetermination is removed, there will be no unique recovery of subsentential content from propositional content (which is essentially what Quine's arguments, made suitably modal, show). It is true that Lewis (1984) later suggested a resolution for propositional semantic underdetermination that is often taken to also resolve subsentential underdetermination; but notice that it does this by in part backing away from the proposition-first model of interpretation. It is less clear that a fully proposition-first model that resolves propositional underdetermination can also resolve subsentential underdetermination. See Schwarz 2014 for relevant discussion, including detailed Lewis exegesis.

Fifth: To make sense of the project as I've described it, the ideal interpreter needs to be able to identify certain beliefs as being more or less central to the conception of a certain content. The notion of 'conception' is not just graded — one claim might be more central to a conception than another — but it is also *concept-relative*. That bachelors are male is plausibly central to our conception of *bachelor*, but not at all to our conception of *male*. The ideal interpreter needs a way to figure out which claims are more or less central to which conceptions. The tricky part is that she has to do this *before* she as interpreted either the 'claims' or the 'conceptions' themselves.

I don't know how, in full generality, to make sense of this for her. But I can make a start. The start will take the form of a just-so, simplistic story. Since the story is most likely far simpler than reality, the story can't be the ultimate explanation of how the ideal interpreter does her job. Hopefully, though, the story will make it plausible that a more realistic story will also give the interpreter the wherewithal to suss out our conceptions.

Suppose that, when determining our brains' computational structure, the ideal interpreter sees that our concepts cluster into 'sentences', and these sentences are each given a status — call them *Status B* and *Status D*. She notices that our Status-B sentences tend to be 'world-tracking': The causal mechanisms that make sentences gain or lose this status tend to be those that operate when gaining information from the environment, whereas Status-D sentences seem less sensitive to information-gathering. On the other hand, Status-D sentences tend to play a certain kind of causal role in the production of action — a role different from the Status-B sentences, especially in that the Status-D sentences are less sensitive to information about the environment.

On the basis of this evidence, the ideal interpreter treats Status-D sentences as (mental representations of) *desires* and Status-B sentences as (mental representations of) *beliefs*. Notice: She does this before assigning contents to anything. She then may observe the following: There are some mental sentences S which use a concept c where (i) the sentences have Status B and (ii) the agent is strongly disposed to stop deploying c if S lose Status B. In other words, the agent is tempted to give up on c-thoughts entirely if he stops believing S. The temptation may be stronger or weaker; it need not be an all or nothing affair. This is evidence of the S-sentences being parts of the agents' *conception* of c, because it is evidence that, if the agent stopped believing S, he would also stop using c.

The real story is of course going to be much more messy; there will be confounds and problems of all sorts, and I can't make a start on them here. But I don't need to. So long as the ideal interpreter has some way to figure out which mental sentence-like vehicles of representation count as central to a concept, she will be able to figure out what claims count as parts of a concept's conception, and to what degree. The fact that I can't figure it out doesn't really matter. The ideal interpreter is much smarter than I am.

3 Conceptual Analysis Revisited

3.1 Descriptive and Content Analyses, Again

If something like this metasemantic picture is right, what role is left for conceptual analysis? There is a very clear role left for something like *descriptive* analyses. On this picture, the job of a descriptive analysis is roughly to suss out our conception of a given concept. The relevance of this project should by now be clear: Our conception of a given concept places direct and important pressure on its interpretation. Ultimately, we are interested in the *content* of our concepts. That will be given by the best interpretation: the interpretation that makes the most sense of what we do with it. That interpretation will be constrained by how the concept is embedded in its larger conception.

On the other hand, there is little hope for any project of content analysis, and for several reasons. First, although the interpretation of a concept is constrained by conceptions, they aren't determined by it. Analysis alone cannot tell us what content a concept gets; we need to know a lot more about how the rest of the world is, too.

Second, conceptions do not have the right 'shape' to deliver conceptual analyses. Suppose the world treats us kindly, and every component of our conception of c turns out right. It doesn't follow that we can give a classical 'analysis' of c from this, because our concept may not give us individually necessary and jointly sufficient conditions for something's falling under c.

Consider, for instance, our concept *before*. Plausibly, it is a deep and important part of our conception that, if x is before y, and y is before z, then x is before z. It is very unlikely, though, that our conception includes further claims that would let us (non-trivially) fill in the blank in

x is before y iff __________.

So even if every component of our conception of *before* is right, we won't get a content analysis out of it.

Third, content analyses are supposed to be *exact*. That is, for a given concept c and necessary condition N, either N is definitely a part of c's content analysis or it is not. But conceptions are inexact. Whether or not a claim is part of a conception isn't an all-or-nothing matter.

Suppose, for example, that almost everyone accepts the following claims about bachelors:

(1) All bachelors are unmarried.

(2) All bachelors are of marriageable age.

(3) All bachelors are male.

(4) All unmarried eligible humans are bachelors.

(5) All bachelors are messy.

(6) No bachelor is a plant.

Some of these claims will be more central to our thinking about bachelors than others, and so will be more deserving to be counted part of our conception. (5), for instance, is clearly *not* part of our conception, whereas (1) is clearly a central part. But what about (6)? It correctly rules out bachelorhood for the plants in my yard. But suppose that we came across a race of sentient plants. Like (some) trees, these sentient plants divide into the biologically male and the biologically female, and they are cognitively capable of forming and participating in practices of marriage (whether or not they in fact bother to do so). Are the unmarried male members of this race bachelors? It seems that we have no clear sense about how to answer this.

Plausibly, that's because (6) is in the penumbra; not far enough away to be clearly outside the conception, but not central enough to be clearly in. As a result, if we try to get a content analysis out of our conception there may well be no fact of the matter about whether the proper analysis is

x is a bachelor iff x is an unmarried eligible male

or rather something else that rules out the sentient plants.[4]

3.2 *Alternatives I: Classical and Inferential Theories of Concepts*

That, anyway, is the picture I want to explore. You might fairly wonder how it compares with other ways of understanding concepts and conceptual analysis. The next two sections will contrast it with various alternatives.

The current picture of conceptual analysis is diametrically opposed to the *classical theory of concepts*. That theory is essentially 'molecular.' There are some basic, 'atomic' concepts, and some story (perhaps a causal story, perhaps something else) connecting them with their contents. Further 'molecular' concepts are made from logical combinations of these atoms. Something falls under a molecular concept

[4] In discussing this case with others, they tend to hem and haw a bit before tentatively concluding that the male plants are probably bachelors after all. I suspect what's going on is that, prior to considering the case, most people's conception of *bachelor* is simply silent about whether plants can be bachelors, but after considering the case people decide that the simplest way of extending the concept to the newly considered case lets the plants in. I don't think that undermines the main point, which is that in principle we may have some beliefs which are neither clearly in nor clearly outside of our conception of some concept.

if and only if it satisfies the property determined by the logical combination of its atoms.

The classical conception has long been in trouble, and from a number of directions. For one thing, it has trouble dealing with the sorts of 'natural kinds' cases discussed above. It also runs counter to our actual classificatory practices: When asked to classify things into (say) 'trees' and 'non-trees', we do not seem to check whether a given sample meets some set of necessary and sufficient conditions in our heads, but instead check to see how close the samples are to a prototypical tree. (See e.g. Rosch and Mervis 1975)

On the classical theory, since the concepts themselves were molecular in nature, their application conditions could be discovered simply by introspection. Conceptual analysis (both a descriptive and content-determining endeavor, on that view) was thus a purely *a priori* matter.

Some successor pictures of concepts keep the *a priority* while rejecting the molecularity. For instance, on a strong inferentialist picture, a concept's content is exhausted by its inferential role. So long as its inferential role is something that can presumably be ascertained *a priori*, its content is *a priori* as well. The job of conceptual analysis is then not that of 'unpacking' the conceptual molecule into its constituent atoms, but rather of mapping the concept's inferential connections with other concepts.

On the current interpretationist picture, there is no guarantee that content can be accessed *a priori*. A proponent of the classical conception might think that *is devoid of empty space* is a component of *solid*, and an inferentialist might think that we can *a priori* infer *x is devoid of empty space* from *x is solid*. On these highly *a priori* pictures, the scientific discovery that almost nothing is devoid of empty space would have to be a discovery that almost nothing falls under the concept *solid*. The interpretationist picture disagrees; we did not discover that nothing falls under *solid*, but rather that the things which do are different than we thought they were.[5]

Interpretationists need not give up entirely on inferentialism. It may be, for instance, that the computational role of certain concepts is exhausted by their inferential role. If so, the ideal interpreter will presumably give those concepts a content that *best fits* their role. If reality is kind to us, there will be a content that perfectly fits that role, and the ideal interpreter will assign it. If reality is not kind, there may be no perfectly-fitting content for her to assign.

[5] Uriah Kriegel reminded me that the inferentalist need not say this, since they may think the conceptual role of *solid* constituted by non-deductive inferences. For instance, *x is solid* may merely probibilify *x is devoid of empty space*. I don't see that makes things a whole lot better, however, since the scientific discovery also seems to show that something being solid makes it at best only infinitesimally more likely to be devoid of empty space. I will stick with deductive inferentialism in the text for ease of exposition.

Interpretationists need not give up entirely on the *a priori*, either. Suppose we come to believe P solely by correctly deploying the inferential role of P's constitutive concepts. Then, if reality is kind to us, P will be true. Perhaps if reality is kind to us our belief in P counts as *a priori* knowledge — even though we cannot tell 'from the inside' that reality was kind to us. If this is the right way to think about *a priori* knowledge, then we can have it even if it is not 'indubitable', and even if it is in part thanks to our good luck in how reality turned out.[6]

An example might bring these points home. Suppose we have two concepts, $\supset$ and $\neg$, with computational roles exhausted by their inferential roles. $\supset$'s inferential role is the one logicians assign to the (material) conditional, and $\neg$'s the one logicians assign to negation.[7] Given these rules, the only viable interpretation of $\neg$ will be as negation: for any proposition P, $\neg P$ will be true if and only if P is false. Furthermore, the rules together will license:

Material: From A and B, conclude $A \supset B$.

RAA: From $A \supset B$ and $A \supset \neg B$, conclude $\neg A$.

If reality works the way our logic assumes it works, everything will be fine. $\supset$ will be interpreted as a (material) conditional and $\neg$ as negation. Furthermore, the constitutive inference roles will be truth-preserving, and the knowledge we get by using them and them alone might as well count as *a priori*.

But *that* reality works the way our logic assumes is itself a metaphysical hypothesis, and one that has been challenged. Graham Priest (2006*a*, 2006*b*, and elsewhere), for instance, has mounted a considerable defense of the view that there are true contradictions. Suppose that Priest is correct, but that the constitutive inferential role of our concepts does not recognize this. Then there *is* no assignment of contents to concepts that perfectly validates their inferential role. If Priest is right, some Q is both true and false. Priest does not think that every proposition is both true and false; some P will be true and only true. Since Q is both true and false, $\neg Q$ is true as well as Q. So if **Material** were truth-preserving, both of $P \supset Q$ and $P \supset \neg Q$ would be true. And if **MP** were truth-preserving, this would mean that $\neg P$ is true, in which case P is false. By hypothesis, P is *not* false, so **Material** and **MP** cannot both be truth preserving — which means the inferential rules constitutive of $\supset$ and $\neg$, from which they were derived, cannot be either.

If reality allows for true contradictions, it has not been kind to us, and the best interpretation of our logical concepts won't validate their constitutive inferential roles. But if the world *has* been kind to us, our inferences will work perfectly, and may even give us *a priori* knowledge. The point is not that the interpretationist

[6] See Jenkins (2008) for a worked-out version of this basic idea.

[7] $\neg$'s rules need not be classical; intuitionistic rules will suffice for the example, as RAA is derivable in intuitionism.

metasemantics removes any role for inferences or *a priori* knowledge. It doesn't. The point is, rather, that neither the inferential role nor the *a priori* status is sacrosanct. Both are hostages to how the world is.

3.3 Alternatives II: The Canberra Plan

The interpretationist picture I've been sketching bears a lot of similarity to the philosophical method known as 'the Canberra Plan' (Jackson 1998 its manifesto). The plan runs like this:

Phase One: Figure out all of the platitudes governing philosophically interesting concepts. For instance, it's a platitude that what is known is true, so 'If *S* knows that *p*, then *p*' is written down as one of the platitudes. Once all the platitudes have been written down, roll them up into one big conjunction. Call this Platitude.

Phase Two: Take Platitude, and turn it into a Ramsey sentence by existentially quantifying into the positions of all of its philosophically interesting terms. Call this Ramsey.

Phase Three: Look at the world to find the best realizers for Ramsey. Then identify the philosophically interesting concepts with their best realizers. For instance, if 'knows' in Platitude was traded in for '*x*' in Ramsey, and if the best realizer for Ramsey assigns a certain state *K* to '*x*', then state *K* is knowledge.

The Canberra Plan thinks of 'conceptual analysis' as Phase One: the process of sussing out all the platitudes. This can be non-trivial. The platitudes are supposed to be common knowledge, but they may only be *implicit* common knowledge, and so considerable work may be needed to make that implicit knowledge explicit.

The Canberra Plan and the interpretationist metasemantics are both functionalist accounts of the content of philosophical concepts. One may suspect they are in fact the *same* account, merely presented in different ways. What I've called 'conceptions' the Canberra Planners call 'platitudes', and what I've called 'interpretation' is what the Canberra Planners call 'realization'.

If the interpretationist picture and the Canberra Plan are to deliver the same results, platitudes should not be identified with conceptions. Let me illustrate with a just-so story. First, background: In Gettier's (1963) famous thought experiment, we are asked to consider the case of Jones. Jones has no idea where Brown is right now, but in fact Brown is in Barcelona. Jones also has good reason to believe that Smith owns a Ford, even though Smith in fact sold his Ford yesterday. Jones concludes, from his belief that Smith owns a Ford, that either Smith owns a Ford or Brown is in Barcelona. Since Brown *is* in Barcelona, Jones's belief is true. We — Gettier's readers — are asked whether Jones belief constitutes *knowledge*. Gettier and most of his peers judge that Jones's true belief does not in fact count as knowledge.

Next, the story: As a matter of fact (goes the story) it is part of our conception that any true justified belief is knowledge. That's because the mental sentence 'known things are exactly those beliefs that are true and that you are justified in believing' gets the relevant mental thumbs-up to count as part of our coneption. But as a community we nonetheless judge Jones as not knowing in Gettier's case (and make relevantly similar judgments about relevantly similar cases). The ideal interpreter will have to decide whether to honor our judgments about cases or our conception, and may very well go with our judgments. In that case, *knows* gets a content that does not apply to Gettier's cases. But if platitudes just are conceptions written down as sentences, then the Canberra Plan will only care about conceptions and not about one-off judgments about cases. In that case, the Plan will deliver the verdict that Jones *does* know in the Gettier cases.

I doubt the story is true, in part because I doubt that 'justified true belief is knowledge' is really part of our conception. But it illustrates why, in principle, the Planner's platitudes should not just include our conception. Jackson's (1998: 31–37) own treatment of Gettier's cases and similar phenomena suggest that he would resist the identification anyway. When defending the Plan, Jackson speaks of our 'implicit theory', and takes our judgments about cases to be (at least partially) relevatory about this implicit theory. Insofar as Platitude is supposed to reflect this theory, it won't be simply a restatement of our conceptions, but instead an amalgam of both our conceptions and our judgments about cases.

There are several ways to implement this idea, but one in particular brings the Canberra Plan much closer to the interpretationist picture. The *hardwiring strategy* simply includes claims about cases in the platitudes. So, the platitudes may include 'knowledge is justified true belief' (if that is in fact part of our conception of knowledge), but it may also include 'If Jones believes that Smith owns a Ford or Brown is in Barcelona solely on the basis of his evidence that Smith owns a Ford, and if Smith does not own a Ford, then Jones does not know that Smith owns a Ford or Brown is in Barcelona.' (This will make the platitudes *very* long, but that was probably going to happen anyway.)

If platitudes are understood this way, the two pictures are rather similar. But they remain different in important ways. For one thing, the interpretationist picture allows for gradation in a way that the Canberra Plan does not. Both pictures allow for some gradation on the *output* side: the Plan because it looks for the *best* realizer of Ramsey, and the interpretationist because the interpreter looks for the *best* assignment of contents to concepts. But only the interpretationst picture allows for gradations on the input side. For one thing, conceptions aren't an all-or-nothing affair. Some claims may be borderline; there may be no fact of the matter, recall, as to whether our conception demands that bachelors be bipeds. For another, judgments about cases can come in degrees, too. Judgments themselves can be tentative; *that* we aren't quite sure what to say in a given case is useful data for the ideal interepreter every bit as much as that we *are* quite sure what to say in others. And when a population splits over a given case (whether they are tentative or confident), that is also data the ideal interpreter can use. It is, at least, an open question whether all this messy give-and-take can be coded up into a single Platitude sentence.[8]

A second difference: Although the ideal interpreter is sensitive to our conceptions and our one-off judgments about cases, that is not *all* she is sensitive to. Her job is to make sense of us, and doing that may require generalizing from patterns of behavior that are invisible to us. It's worth keeping in mind that we use concepts for a lot more than just making one-off judgments about cases.

This can happen internally. Perhaps we deploy our concept of *freedom* in rational deliberation in such a way that certain possibilities are ruled out before even being presented to the deliberative process. If so, the ideal interpreter will want to assign *freedom* a content that makes this procedure rational; but this feature may not be discoverable in a way that could be coded into a platitude.

This can also happen externally. Perhaps our concept *expert* figures in the process by which we evaluate testimony, and it might do it in such a way that comes apart from our explicit judgments about experts. Maybe when asked whether a certain person is an expert we employ a system that issues *expert*-judgments in light of certain criteria, but when we evaluate testimony we deploy the concept differently, so that some people who may have been judged an expert by the first system won't be treated-as-an-expert when testimony time comes around. The ideal interpreter needs to make sense of our *expert* concept, but we are inconsistent enough with it that she cannot do it perfectly. It may well be that she can make better sense of us by interpreting *expert* in line with how we deploy it in testimony-evaluation than we do in yes-or-no classifications.

Both of these functional roles for concepts are not happily captured in the Canberra Plan's platitudes. The idea behind the Platitudes, I thought, was that they were supposed to be available by first-person introspective access. There is plenty of empirical evidence, however, showing us that we are bad at introspectively assessing the way external social factors influence our judgments. If that's right, this sort of social effect won't show up in the Platitude, in which case the Canberra Plan's determination of content won't be affected by it the way the ideal interpreter's would.

The point is not that these are the *right* accounts of our concepts *freedom* or *expert*, but rather that there are various theoretical possibilities which the ideal interpreter will treat differently than the Canberra Plan.

[8] Though see section 3.2 of Kriegel's contribution to this volume for an attempt to do just this.

I suppose that a committed Canberra Planner could bring the two pictures into closer alignment, perhaps by fixing the PLATITUDES in such a way that it codes up each term's broader functional role. In the limit, the two pictures may converge. But even if they do, thinking in terms of an ideal interpreter makes salient content-fixing features that may not come to mind as readily when thinking in terms of 'platitudes'.

4 The Role for Experimental Philosophy

4.1 Intuitions

A classical picture of conceptual analysis gives pride of place to *intuitions*. Unfortunately, the classical picture had little to say about precisely what these were. If we look at how appeals to intuitions get used in philosophy, though, we can identify two different forms they tend to take.

One form tends to be as unreflective, snap judgments about particular cases — the kind of one-off judgments made about thought experiments such as Gettier's. This is perhaps the kind of 'intuition' most experimental philosophers have in mind when they use the term. But there is a second kind of 'intuition' prevalent in the literature as well, which appears as a kind of pre-theoretically compelling general claim.

Peter van Inwagen, for instance, when defending (roughly) the principle that if you have the ability to falsify a proposition, you have the ability to falsify whatever entails that proposition,[9] writes:

> I must confess that my belief in the validity of [the principle] has only two sources, one incommunicable and the other inconclusive. The former source is what philosophers are pleased to call 'intuition': when I carefully consider [the principle], it seems to be valid. (1983: 73)

(The second source, he goes on to say, is that he can think of no counterexamples to the principle.) By 'intuition' van Inwagen clearly doesn't mean snap-judgments about cases, but rather than he simply finds the principle deeply compelling, and has a hard time imagining how it could be false.

On the interpretationist picture, these two different sorts of 'intuitions' are best thought of as two different phenomena. The kind of 'intuition' van Inwagen reports

[9] Somewhat more precisely, the principle is

> If $P \rightarrow Q$ and no one has or ever had a choice about $P \rightarrow Q$, and if P anf no one has or ever had a choice about P, then Q and no one has or ever had a choice about Q.

Even more precisely, the principle is an inference rule which would be truth-preserving if and only if the just-stated principle is true.

will be understood as something like a report about a *conception*: his principle is part of his conception of *ability*. On the other hand, when we make snap judgments about cases we straightforwardly deploy or refuse to deploy a concept. When considering the Gettier cases, we deploy our concept *knowledge* in imagination just as we deploy it outside of imagination when trying to determine what our peers and associates know.

Both phenomena place significant pressure on an ideal interpreter. Arguably, charity requires that she interprets our concepts so that confident snap judgments using them tend to be correct, at least when there are no conditions present that, if known to us, would lead us to retract the judgment. Charity also requires that she preserve as much of our conception as possible. Insofar as that's right, we can use our intuitions to get a grip on the kinds of interpretations the interpreter could have charitably given our concepts, and that way use intuitions as evidence that our concepts have some particular content.

The present picture cannot give intuitions as central a status as the classical picture did, though. The classical picture treated intuitions (whatever they were) as something like inviolable constraints. The second worst thing that could happen to an analysis was that it be counterintuitive; the worst was that it be counterexampled — inconsistent with snap judgments about thought experiments. Intutions were treated as philosophical data, fixed points that had to be worked around.

On the present picture, nothing is sacrosanct, and for several reasons. First, the ideal interpreter is beholden to a *community's* conception, not an *individual's*. Even if van Inwagen treats his principle as part of his conception, if he is idiosyncratic enough — if enough of his community do not treat it as part of their conception — it may not be part of the community conception, and the ideal interpreter won't be under much pressure to make it true. Even if all of the philosophers have the principle as part of their conception, the philosophers make up only a small part of their community; if their less philosophical peers don't share this conception, the philosophers will presumably be simply outvoted, that part of their conception carrying less weight with the ideal interpreter.

By the same token, the ideal interpreter needs to interpret not just the handful of snap judgments philosophers make about individual cases, but *all* the concept-deployments made in the relevant community. If ordinary folk commonly and unhesitatingly deploy their *knowledge* concept in the presence of Gettier-like cases, then the mere fact that a few philosophers do not will carry little weight with the ideal interpreter. She cannot make us *all* right, and once again, if we philosophers are in the minority, we will simply be outvoted.[10]

[10] I'm glossing over an important point, which is that ordinary folk may simply be getting confused by the complexity of philosophers' cases. There are such things as performance errors, and the ideal interpreter will have to work around them somehow. If ordinary folk attribute *knowledge* in Gettier's own (admittedly hard to parse) cases, but not in simpler cases with the same structure,

Finally, the pressure from snap judgments and the pressure from conceptions might pull against each other. Plausibly, that's what has happened in the case of *solidity*: the ideal interpreter cannot make our judgments about the solidity of rocks and tables right without making the no-empty-space part of our conception wrong, and vice versa. This can happen in more philosophical cases. Suppose that van Inwagen is right about our conception containing his principle, but that we happen to live in a deterministic world. I've argued elsewhere (2013) that, in this sort of case, it *has* happened: pressure from our snap judgments has overridden pressure from our conception. Pressure on the ideal interpreter to charitably interpret our *freedom* concept forces her to give it a content compatible with determinism, which then leads to van Inwagen's principle being violated (or so I argued).

It can also happen in the other direction. Even if we confidently deploy some concept *c* in many situations, if our attachment to some part of *c*'s conception is *too* strong, and if the ideal interpreter cannot be charitable both to our deployments and this part of our conception, she may make us wrong about our one-off judgments.[11]

4.2 *Sussing out Conceptions, Tracking Judgments*

Much experimental philosophy has seen itself as firmly participating in an intuition-based methodology, either as full participants in that methodology (Knobe 2003, Nahmias et al. 2005) or as interlopers trying to subvert it from within (Weinberg et al. 2001, Mallon et al. 2009). The interlopers accuse traditional philosophy of an over-reliance on intuitions, and experiments are pulled out to show that, overall, intuitions (usually in the form of judgments about cases) are shaky across a large and diverse population, undermining the method that appeals to such judgments.

Since the interpreter-based methodological picture doesn't give intuitions the same status, it isn't undermined by experiments in the same way. Rather, experimental philosophy should be understood as an empirical investigation into precisely the sort of facts the ideal interpreter needs to know in order to do her job. For instance, when Weinberg et al. (2001) discover that Eastern students are more prone to count Jones as knowing in the Gettier case than Western students, we discover that the ideal interpreter has a messier set of judgments about cases to deal

we philosophers are probably right after all. If ordinary folk attribute *knowledge* in even the simplest and easiest-to-understand cases, though, then we are in trouble. Thanks here to Uriah Kriegel.

[11] Example: If each macroscopic physical object were a thin shell around a vacuum, and we only failed to notice this because an evil demon kept adding new surfaces whenever we cut one of these hollow objects in half, the ideal interpreter would interpret *solid* in a way to make it *false* that rocks and tables are solid. We can handle being pulled a little bit away from our conception of *solid* to allow some empty space, but we can't handle being pulled *that* far away. (See my 2013: 301–303)

with than we might have initially thought.[12] That means that she is under a less pressure to assign *knows* a content that does not apply to Jones in the Gettier case.

Even if we discover that most people judge Jones to know in the Gettier case, it does not follow that the ideal interpreter will assign *knows* a content that applies to Jones. Remember that the ideal interpreter must consider not only our conceptions and snap judgments, but also the role that the concept plays in our overall cognitive lives. The concept *knows* is central and important enough that it will undoubtedly have many computational connections with other concepts, whether we are aware of these or not. It may be that these connections will pressure the ideal interpreter into treating some of them as more important than others.

To take one example, it may be (as some have suggested; see Williamson 2000: ch. 11) that knowledge is a norm of assertion: We should not assert what we do not know. Suppose this is reflected in agents' cognitive makeup as follows: They are disposed to assert that *p* only when they take the concept *knowledge* to relate them to *p*, and if they assert that *p* and then come to retract the application of *knowledge* to themselves with respect to *p*, they are then disposed to retract the assertion. If actual folk in Gettier-style scenarios are disposed to retract their assertions upon finding out that the evidence they based their belief on is misleading, then even if a majority of observers 'from the outside' are inclined to judge them as knowing, pressure from this functional role for *knowledge* may lead the ideal interpreter into assigning it a content that does not apply to Gettiered people.

Of course, this claim about dispositions is just as empirical as a claim about ordinary people's judgments about the Gettier cases. As such, it's a fit subject for philosophical experimentation, too. The point is not that experimentation is irrelevant! It is, rather, that given the large number of considerations that may influence the ideal interpreter in doing her job, no single data point, no single experiment — whether thought- or empirical — is going to settle things. Philosophy has always been a messy business, and the current picture that helps explain just how messy it is, and why.

The role of experiments, then, is to help suss out precisely what constraints the ideal interpreter is operating under while trying to interpret us. But there is a role for broadly *a priori* methods too, and in two directions. First, the sum total of experiments might eventually, in the limit, give us all the data that the ideal interpreter has to work with. But there is still the further question: what assignment of interpretations to concepts *best* fits the data? That is not an empirical question; it requires the kind of reflective judgment philosophical reflection can bring to it. So in the long run, even once all the empirical data is in, a tough and non-empirical question remains.

[12] The version of the case they presented respondents was slightly different, but that needn't concern us here.

In the short term, *a priori* methods help by offering the experimenters hypotheses worthy of testing. Experimental philosophers are well-positioned to check whether people tend to give the same judgments in Gettier cases cross-culturally; or whether we are disposed to retract a claim once we find our evidence for its inferential basis is misleading; or whether we tend to judge an action as free and its agent as morally responsible when it has been predicted by a computer from the state of the universe before the agent was born; and so on. But we didn't need empirical methods to come up with those hypotheses. *A priori* reflection led many of us to judge agents in the Gettier cases as not knowing, that knowledge is a norm of assertion, and that agents whose actions were in principle predictable with perfect accuracy before they were born are unfree and not morally responsible. Philosophical training makes us very good at coming up *a priori* with interesting and plausible hypotheses, and very good at tracking the consequences of these hypotheses. Furthermore, since we are *among* the target population, finding that these hypotheses reflect our own one-off judgments or conceptions gives us some evidence that the ideal interpreter will be trying to vindicate them. We simply need to remember that this individualistic evidence is fairly weak, and that empirical work may show us that we are in the minority.

5 Interpretation and Metaphysics

5.1 *Two Traditional Metaphysical Projects*

'Metaphysics' is less a single narrow project and more a conglomerate of loosely connected projects unified by a family resemblance.[13] Some of those projects are more closely allied with conceptual analysis than others. Debates about the nature of free will or of personal identity, to take a couple of examples, have largely been treated as conceptual projects, whereas debates about the nature of individuals (i.e., whether individuals are 'particulars' instantiating properties, or instead bundles of compresent properties, or something even more *recherché*) have not. Other debates are more mixed; debates about the nature of time, for instance, may include conceptual appeals as well as more traditionally 'metaphysical' concerns.

What are metaphysicians *doing*, then? On the present picture, traditional metaphysical projects can generally be understood as doing either of two very different things. These two different things correspond to two types of constraints on ideal interpretation. One is a constraint coming from *us*, and it concerns the sum total of our behavior, dispositions, conceptions, and so on. Another is a constraint coming from the *the world*, and determines what contents are available for the ideal interpreter to assign. One kind of metaphysical project is concerned with the balance:

[13] Or so say I. Others may disagree, but I lack the space to argue with them here.

Trying to find precisely where the world meets our concepts, to figure out their content. Debates about free will or personal identity are easily understood as this sort: we don't want to just know what our conception of free will or personal identity is, we want to know what *free will itself* or *personal identity itself* is. But these phenomena just will be whatever content the ideal interpreter assigns to these concepts. So debates of this sort are constrained by two different factors: What we are like, and what the world is like.

But what *is* the world like, before it is interpreted? The other sort of metaphysical project asks this directly. Of course, since we are doing this project we have to deploy the concepts we have in order to ask the question, perhaps coining some new ones along the way. But we are not asking the question in order to fit our concepts to it, but rather in order to figure out what it is like 'pre-conceptually', as it were. Debates about whether individuals are bundles of properties or instead 'bare' particulars are of this sort. We aren't asking about our concept of individuals at all, but rather what possible candidate interpretations of *individual* are out there to be meant.

Skepticism about the second sort of project is natural. ('If we have to *use* our concepts to undertake the project, how could we ever find out what it is like "pre-conceptually?"') I cannot hope to defend it here. But note that skepticism here will spill over even to more solidly 'naturalistic', science-driven metaphysics. Those who say that reality is in fact a wave in a very high-dimensional space (Albert 1996) or comprised of quantum states at regions of spacetime (Wallace and Timpson 2010) may have to use our concepts to describe reality, but are using our concepts to describe not the manifest image but instead the raw materials that (if they are right) the ideal interpreter has to work with. Debates about whether individuals are particulars or bundles of properties may be less constrained by science, but are in principle concerned about the same general sort of thing: what the world is like, prior to the interpretation of our concepts.

Given the difference between these two projects, we can see why the objection to experimental philosophy at the beginning of the paper is not really on target. When we wonder about the nature of time, we wonder to a large extent about what the ideal interpreter could find in reality to assign to our time-concepts. We are fully confident that people's *conception* of time (or their one-off judgments, or perhaps both) have it that some events are absolutely simultaneous. We are wondering whether reality has the right sort of structure for the ideal interpreter to make these judgments true. When we claim that there is no absolute simultaneity, we are effectively claiming that reality didn't give the ideal interpreter the resources she needed to vindicate these simultaneity judgments. And our reasons for thinking this have less to do with the shape of our thoughts using *simultaneity* and more to do with the nature of the world where those thoughts were being had.

By contrast, when we wonder about the nature of free will, we wonder in large part about the content of the concept *free will*. Traditional arguments to the effect that free will (the stuff itself) is, say, incompatible with determinism tend to rely on largely conceptual matters. These arguments should, on the present framework, be understood roughly as follows: 'Free will (the stuff itself) is the content of our concept *free will*; but our concept is determined in part by its conception, and our conception makes *free will* incompatible with determinism. Thus the ideal interpreter will not give it a content compatible with determinism, in which case free will and determinism are incompatible.' The contested premises here are the ones about the nature of our conception, which is what influences the ideal interpreter. Since the ideal interpreter is influenced by the *shared* conception of all of us (and not just the conception of this or that philosopher), experimental methods seem an apt way to settle just what interpretive pressure she was under.

But although the original objection was not entirely apt, it did catch a glimmer of truth. The glimmer is this: Even in the case of *free will*, conceptual contours aren't the whole story. What free will 'really is' gets determined not just by our concept but also by what contents are available to the ideal interpreter and other interpretative pressures. As I've argued elsewhere (2013), even if our conception is in fact incompatibilist, if we happen to live in a deterministic world the ideal interpreter may well assign *free will* a content compatible with determinism. If so, free will is compatible with determinism, our incompatibilist conceptual leanings notwithstanding.

5.2 Broadened Horizons?

The two projects — conceptual analysis and 'deep metaphysics' — are relatively familiar. The metasemantic picture discussed here provides a sort of rational reconstruction of what these projects are up to. The reconstruction doesn't completely vindicate both projects. Given what this picture thinks conceptual analysis amounts to, its practitioners have not been drawing on all the relevant evidence. Experimental philosophy provides a helpful corrective by bringing more relevant evidence to bear. Unfortunately, the evidence is in principle so wide, consisting of so many different, messy factors that constrain an ideal interpreter, that we will probably never get all the relevant evidence, and will have to be content to muddle along as best as we can with what is available to us.

But the current framework can make sense of more projects than just the traditional two. Consider, for instance, Sally Haslanger's (e.g. 2000, 2007) *ameliorative* analysis of gender. She gives an account of gender which intends neither to unpack our conception nor to try to locate gender as a deep feature of reality, akin to time or the nature of individuals. Its job, rather, is

> not to explicate our ordinary concepts, nor is it to investigate the kinds

> that we may or may not be tracking with our everyday conceptual apparatus; instead we begin by considering more fully the pragmatics of our talk employing the terms in question. What is the point of having these concepts? What cognitive or practical task do they (or should they) enable us to accomplish? Are they effective tools to accomplish our (legitimate) purposes; if not, what concepts would serve these purposes better? (2000: 33)[14]

Undertaking these questions, she comes up with an analysis of gender in terms of the way it interacts with and influences social interactions and structures. Nonetheless, she insists that her project is metaphysics, and her analysis tells us what gender *is*.

If the only metaphysical projects we recognize are those of conceptual 'unpacking' and deep metaphysics, Haslanger's ameliorative project will be hard to understand. (Barnes 2014) But it is relatively easy to understand the proposal given the current metasemantic picture. 'What gender *is*', we may assume, is just whatever is in fact the content of the concept *gender*. That concept's content is whatever the ideal interpreter assigns to it. But the ideal interpreter doesn't simply look at our conceptions, judgments, and the world to make an assignment. She also has to pay careful attention to precisely how that role functions in our overall cognitive and social economy. Her job, remember, is to make sense of us; and part of making sense of us is making sense of how our concepts in fact get employed in their cognitive and social settings. If our gender concept is in fact best understood as enmeshed in the sort of socio-structural properties that Haslanger describes, then that may well be the best interpretation of the concept, despite that being opaque to us. The best interpretation of our *gender* concept is the one that makes the most sense of us when we deploy it. The pressure from one-off judgments about cases, or our (perhaps strongly-held) beliefs about gender, are less central to making sense of us than the overall social and cognitive role that the concept plays in our society.

6 Conclusion

The ideal-interpreter-driven picture of philosophical theorizing that I have sketched here is attractive, though I expect (like any philosophical thesis) it will be controversial. I haven't tried to defend it. My aim has been instead to reconstruct what philosophers have been up to, and what they should have been up to, if that picture is right. If the picture *is* right, conceptual analysis is a messy business, and will involve all kind of empirical considerations, ranging from how non-philosophers respond to purported counterexamples to how a given concept is embedded in our

[14] On my reading, Haslanger's use of 'concepts' corresponds with what I am calling 'conceptions'.

larger socio-cognitive practices. On the other hand, despite its messiness, it suggests also that traditional philosophical methodology has been largely on the right track: It's been looking at the right *kind* of evidence, even if it has often ignored other points of data of the same sort. Philosophy doesn't need a revolution, overthrowing traditional methods, but a supplementation, mixing new evidence with the traditionally-gathered old.

References

Albert, David (1996). "Elementary Quantum Metaphysics." In J. Cushing, A. Fine and S. Goldstein (eds.), *Bohmian Mechanics and Quantum Theory: An Appraisal*. Kluwer.

Barnes, Elizabeth (2014). "Going Beyond the Fundamental: Feminism in Contemporary Metaphysics." *Proceedings of the Aristotelian Society* 144(3): 335–351.

Davidson, Donald (1974). "Belief and the Basis of Meaning." *Synthese* 27(3/4): 309–323.

Gettier, Edmund (1963). "Is Justified True Belief Knowledge?" *Analysis* 23(6): 121–123.

Haslanger, Sally (2000). "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34(1): 31–55.

— (2007). "What Good are our Intuitions?" *Aristotelian Society Supplementary Volume* 80(1): 89–118.

Hirsch, Eli (2005). "Physical-Object Ontology, Verbal Disputes, and Commonsense." *Philosophy and Phenomenological Research* 70(1): 67–97.

Jackson, Frank (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford Univerisity Press.

Jenkins, Carrie (Ichikawa) (2008). *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. Oxford: Oxford University Press.

Knobe, Joshua (2003). "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63(3): 190–194.

Kripke, Saul (1972). *Naming and Necessity*. Harvard University Press.

Lewis, David (1974). "Radical Interpretation." *Synthese* 23: 331–344. Reprinted, with postscripts, in Lewis 1983.

— (1975). "Language and Languages." In *Minnesota Studies in the Philosophy of Science*, volume 7, 3–35. Minneapolis, Minn.: University of Minnesota Press. Reprinted in Lewis 1983: 163–188.

— (1983). *Philosophical Papers*, volume 1. Oxford: Oxford University Press.

— (1984). "Putnam's Paradox." *The Australasian Journal of Philosophy* 62: 221–236. Reprinted in Lewis 1999: 56–60.

— (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Mallon, Ron, Edouard Machery, Shaun Nichols and Stephen Stitch (2009). "Against Arguments from Reference." *Philosophy and Phenomenological Research* 79(2): 322–356.

Nahmias, Eddy, Stephen G. Morris, Thomas Nadelhoffer and Jason Turner (2005). "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18(5): 561–584.

Priest, Graham (2006*a*). *Doubt Truth to Be a Liar*. Oxford: Oxford Univerisity Press.

— (2006*b*). *In Contradiction*. 2nd edition. Oxford: Oxford Univerisity Press.

Putnam, Hilary (1962). "It Ain't Necessarily So." *The Journal of Philosophy* 59: 658–671.

— (1975). "The Meaning of Meaning." In *Mind, Language, and Reality*, volume 2. Cambridge: Cambridge University Press.

Quine, Willard Van Orman (1960). *Word and Object*. MIT Press.

Rosch, Eleanor and Carolyn B. Mervis (1975). "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7: 573–605.

Schwarz, Wolfgang (2014). "Against Magnetism." *The Australasian Journal of Philosophy* 92: 17–36.

Turner, Jason (2013). "(Metasemantically) Securing Free Will." *The Australasian Journal of Philosophy* 91(2): 295–310.

van Inwagen, Peter (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

Wallace, David and Christopher G. Timpson (2010). "Quantum Mechanics on Spacetime I: Spacetime State Realism." *The British Journal for the Philosophy of Science* 61(4): 697–727.

Weinberg, Jonathan M., Shaun Nichols and Stephen Stitch (2001). "Normativiy and Epistmic Intuitions." *Philosophical Topics* 29(1-2): 429–460.

Williamson, Timothy (2000). "Existence and Contingency." *Proceedings of the Aristotelian Society* 100: 117–139.