Synthese (2016) 193:2367–2388
DOI 10.1007/s11229-015-0848-4

Conceptual analysis and x-phi

Mark Balaguer[1]

Received: 26 March 2015 / Accepted: 3 August 2015 / Published online: 26 August 2015
© Springer Science+Business Media Dordrecht 2015

Abstract This paper does two things. First, it argues for a metaphilosophical view of conceptual analysis questions; in particular, it argues that the facts that settle conceptual-analysis questions are facts about the linguistic intentions of ordinary folk. The second thing this paper does is argue that if this metaphilosophical view is correct, then experimental philosophy (or "x-phi") is a legitimate methodology to use in trying to answer conceptual-analysis questions.

Keywords Conceptual analysis · Experimental philosophy · Meta-philosophy

1 Introduction

I have two aims in this paper, one that's relevant to a very traditional philosophical concern and one that's relevant to a currently fashionable trend in philosophy. My first aim is to argue for a metaphilosophical view of conceptual-analysis questions—of what these questions are really *about*—and my second aim is to show that if this metaphilosophical view is right, then experimental philosophy (or "x-phi") is a legitimate methodology to use in trying to answer conceptual-analysis questions.

When I talk about x-phi in this paper, I'll be talking about (a) the attempt to gather experimental data about the intuitions of ordinary folk regarding philosophical questions of a certain kind, and (b) the idea that this data is relevant to the job of determining the *answers* to these questions. (This could also be called the "positive project of x-phi," but I usually won't bother to qualify the term in this way. It's important to note, however, that this isn't all there is to experimental philosophy; there are other projects

✉ Mark Balaguer
mbalagu@calstatela.edu

[1] Department of Philosophy, California State University, Los Angeles, Los Angeles, CA, USA

that x-phi can be used for (and has been used for).[1] I will have a bit to say below about one of these other projects (namely, the so-called "negative project"[2]), but I will mostly be concerned with the positive project—i.e., with the use of experimental data about folk intuitions to help answer philosophical questions; this is why I'll use the term 'x-phi' to refer to just the positive project.)

It's very common to see professional philosophers roll their eyes at the mere mention of x-phi. There are probably numerous reasons for the eye rolling, but one simple thought process here goes something like this:

> *Eye-rolling response to x-phi*: The idea that we can determine the answers to substantive philosophical questions by asking ordinary folk what they think is just ludicrous. It's analogous to claiming that we can answer open questions in physics or mathematics by polling ordinary folk. Who cares what the folk think?

In connection with lots of philosophical questions—e.g., ontological questions like 'Are there any abstract objects?'—I think this response is dead right. That's because with these questions, there's no plausible story to tell about how folk intuitions could be tracking the relevant facts. But I'm going to argue that there's one kind of philosophical question (I'll call them *type-C* questions, but we can also call them conceptual-analysis questions) for which folk intuitions can plausibly be seen as tracking the relevant facts; more specifically, I will argue that folk intuitions can be viewed as empirical data points that are relevant to the answers to type-C questions.

Let me start by defining the notion of a type-C question. A *type-C question* is (I hereby stipulate) a certain kind of question of the form, 'What is C?' (or 'What is a C?'), where 'C' denotes or expresses a concept. To give a few examples, I'm thinking of questions like 'What is free will?', 'What is a person?', 'What is knowledge?', and so on. Now, it's important to note that *not all* questions of the form 'What is X?', or 'What is an X?', are type-C questions. Neuroscientists might ask "What is a decision?", and they might answer that decisions are neural events of a certain kind; but this is different from the type-C question 'What is a decision?' The type-C question is about the *concept* of a decision, whereas the neuroscientific question is about the physical nature of the (actual) referents of that concept. Thus, when I talk about type-C questions, I'm not talking about *all* questions of the form 'What is X?'; I'm talking only about those questions that ask about the nature of the given *concept*. In particular, we can think of a type-C question as asking about the *decompositional structure* of the given concept, i.e., about the way that it's composed out of other concepts. For instance, one way to answer the type-C question about *knowledge* is to say that that concept is identical to the concept *true justified belief*.

Note that the distinction I'm drawing here—between conceptual and non-conceptual 'What is X?' questions—holds even when the given concept is a natural-kind concept. Consider, e.g., the case of *water*. Chemists have taught us that water is H_2O; but this isn't the answer to the *type-C* question about water; the type-C

[1] For discussions of the kinds of things that x-phi can be used for, see, e.g., Knobe and Nichols (2007) and O'Neill and Machery (2014).

[2] The negative project of x-phi is (roughly) to criticize the use of intuitions in philosophy; see, e.g., Alexander et al. (2010).

question asks what the *concept* of water is; i.e., it asks what concept is expressed by the word 'water'. And the answer, of course, is that 'water' expresses a certain natural-kind concept; in particular, following Chalmers (1996) and Jackson (1998), we can say that the concept *water* is identical to the concept *the actual watery stuff of our acquaintance* (or something like that). This is why 'water' is a rigid designator. We learned this by doing conceptual analysis, in particular, from Putnam's (1975) Twin Earth thought experiment; if not for Putnam, we might have thought that the relevant concept was *clear, odorless, tasteless liquid*, or something like that. (Note, too, that what I'm saying here is perfectly consistent with semantic externalism; it may be that the semantic content of ordinary uses of 'water'—i.e., what these uses contribute to the propositions expressed by sentences containing 'water'—is the property *being* H_2O; but it doesn't follow that the concept *water* is identical to the concept H_2O, and it obviously isn't—that's why 'Water is H_2O' isn't analytic. You can think of the point here like this: when philosophers try to answer type-C questions, they're not after semantic content (if it's distinct from meaning); they're after *meaning* (or Fregean sense, or Kaplanian character (see Kaplan 1989), or some such thing).

Just to be clear, my reason for introducing the notion of a type-C question is not because I think these questions are somehow different from conceptual-analysis questions. I don't; I think these two sets of questions are identical. But I don't want to justify this claim—I don't want to get into an argument about what conceptual analysis really is—and that's why I'm introducing the notion of a type-C question and stipulating the definition. So while my official goal in this paper is to argue for a thesis about *type-C* questions (in particular, I'll argue for a metaphilosophical view of these questions that entails that x-phi is a legitimate methodology to use in trying to answer them), I think that everything I say here applies equally well to *conceptual-analysis* questions—again, because I think that the set of type-C questions just *is* the set of conceptual-analysis questions. Moreover, in what follows, I will sometimes talk as if these two sets are identical. But I will not argue for this point, and nothing important depends on it, for again, my official thesis is about type-C questions, not conceptual-analysis questions.

In Sects. 2–3, I set things up; in Sects. 4–6, I argue for my metaphilosophical view; and in Sects. 7–9, I respond to objections and clean up some loose ends.

Let me make two more points before moving on. First, when I talk about x-phi, I'm not talking about gathering data about folk intuitions about the *answers* to type-C questions (i.e., about what C is, or what 'C' means); rather, I'm talking about gathering data about folk intuitions about whether some actual or possible object is an *instance* of C. For instance, we won't ask people to give us their intuitions about what free will is; rather, we'll ask them to give us their intuitions about whether some specific decision is free. I'll say more about this later.

Second, even if I'm right that x-phi is a legitimate methodology to use in connection with type-C questions, it doesn't necessarily follow that we shouldn't roll our eyes at x-phi. For you might think that what my argument shows is that we should roll our eyes at type-C questions. I'll say more about this in Sect. 9.

2 What is a concept?

There are two main views of the metaphysical nature of concepts. On the psychologistic view, they're mental objects, existing in the heads of actual human beings; and on the platonistic view, they're non-spatiotemporal abstract objects, in particular, something like *meanings*. I will proceed as follows: throughout the bulk of the paper, I'll argue for my view under the assumption that concepts are abstract objects (or that they're *supposed* to be abstract objects—more on this in a moment); then at the end (in Sect. 8), I'll explain why my argument still goes through even if we assume that concepts are mental objects.

Let me make two quick points about the platonistic view of concepts. First, platonists don't deny that there are things in our heads that *correspond* to concepts; but on their view, the things in our heads aren't concepts; they're *mental representations* of concepts. Second, the kind of platonism that I have in mind is a *plenitudinous* platonism according to which every concept that we can dream up already exists. So on this view, for every analysis of a concept C, there's a concept that corresponds to that analysis. E.g., in the case of free will, there's a concept of *Hume-freedom*, and *libertarian-freedom*, and *reasons-responsiveness-freedom*, and so on. We can think of these concepts as existing "side-by-side" in platonic heaven. Thus, on this view, the type-C question 'What is free will?' asks *which* of these concepts is the concept of free will.

(For whatever it's worth, I think there are good arguments for thinking that the platonistic view is right in the following sense: it's built into the way that we think and talk about concepts that they're *supposed* to be abstract objects, so that *if concepts exist at all*, then they're abstract objects. But this doesn't commit us to actually believing in abstract objects. For just as mathematical fictionalists maintain that there are no such things as numbers (because they're supposed to be abstract objects, and there are no such things as abstract objects), we can say that there are no such things as concepts. And this doesn't commit us to saying that talk of concepts (and conceptual analyses) is worthless; for just as mathematical fictionalists can say that mathematics is a useful fiction, so too we can say that talk of concepts is a useful fiction.)

3 Three metaphilosophical views

Consider the following question:

> *Meta-question*: What sorts of facts settle our type-C questions? I.e., what sorts of facts determine which answers to our type-C questions are correct? Suppose, for instance, that we're trying to figure out what *free will* is, and suppose that Smith says that it's *Hume-freedom* (i.e., roughly, the ability to do what you want) and Jones says that it's *libertarian-freedom* (i.e., roughly, the ability to make decisions that are simultaneously undetermined and appropriately non-random). Then the meta-question here is this: What sorts of facts determine whether Smith or Jones (or neither) is correct?

One obvious answer you might give to this question is the following:

The decompositional view: An answer to a type-C question is correct just in case it captures the decompositional structure of the given concept—i.e., the way that concept is composed out of other concepts. Thus, our type-C questions are settled by the actual decompositional structures of the relevant concepts. Consider, e.g., the type-C question about free will; this debate is settled by the decompositional structure of the concept of free will—i.e., the one that exists in platonic heaven. If that concept has a Humean structure, then Humeans are right; if it has a libertarian structure, then libertarians are right; and so on. Thus, type-C questions are similar in a certain way to *mathematical* questions. If we ask what kinds of facts settle our arithmetical questions, the obvious answer is that they're settled by facts about *numbers*. Likewise, type-C questions are settled by facts about *concepts*.

This view has an initial appeal to it, but it isn't the only way to answer the meta-question. Another answer is suggested by reflecting on the fact that type-C questions can be thought of as *semantic* questions, in particular, questions about *meaning*. The question 'What is free will?' is essentially equivalent to the question "What concept is expressed by the term 'free will'?", or more simply (taking concepts to be meanings), "What is the meaning of 'free will'?" Given this, one obvious answer to the meta-question is the following:

The ordinary-language view: An answer to a type-C question is correct iff it captures the ordinary-language meaning of the corresponding expression 'C'—i.e., iff it picks out the concept that ordinary folk have in mind when they use the term 'C'. Given this, we can say that the facts that determine the answers to our type-C questions are facts about folk meaning, and these are facts about *us*—about our usage, intentions, conventions, practices, and so on. A second way to articulate the ordinary-language view is to say that an analysis of a concept C is correct iff it captures *our* concept of C, i.e., the one *we* have in mind when we think of C, so that the facts that settle the given type-C question are facts about *our mental representation* of C. Indeed, we can get even more specific than this; we can say that the facts in question are facts about the *decompositional structure* of our mental representation of C. So, the ordinary-language view is similar in a certain way to the decompositional view; the difference is that according to the latter, the relevant facts are about the decompositional structures of *concepts themselves*, rather than our mental representations of those concepts.

Finally, here's a third view one might endorse in response to the meta-question:

The hybrid view: Facts about the folk (about their usage, intentions, conventions, and so on) are certainly relevant to determining the answers to our type-C questions, but they aren't the whole story. When we do conceptual analysis, we're trying to "clean up" folk usage in a certain way, and so other kinds of facts are relevant here, in addition to facts about folk usage and intentions. Suppose, e.g., that (a) the ordinary usage and intentions associated with 'C' pick out a concept C1 that's incoherent, or contradictory, and (b) C2 is a coherent concept that, so to speak, "cleans up" ordinary usage and intentions in the best overall way; then you might think that C2 is a better candidate than C1 for *being C*, or for being the

> meaning of 'C'. And so you might think that facts about logical coherence are relevant to determining the answers to our type-C questions. Now, it's important to note that advocates of the hybrid view do not commit to any *specific* view concerning the kinds of facts that are relevant to type-C questions. They commit to the idea that facts about folk usage and intentions are relevant, but beyond that, all they say is that *other* kinds of facts are relevant as well. Facts about logical coherence are one possible example here, but (a) the hybrid view doesn't commit to the claim that facts about logical coherence are definitely relevant to our type-C questions, and (b) it leaves open the possibility that *other* kinds of facts are relevant as well—e.g., facts about Lewisian naturalness, or the decompositional structures of concepts, or whatever. (In Sect. 6, I'll discuss a number of different kinds of facts that one might want to add to the list here.)

Some advocates of the hybrid view might claim that the overall process of arriving at an acceptable conceptual analysis involves a sort of migration toward a *reflective equilibrium* between the desire to satisfy our ordinary intuitions and other kinds of considerations. But this needn't be part of the view.

The above three views are obviously not the only views one might endorse in response to the above meta-question. We can arrive at a more exhaustive list of views by moving from the decompositional view to a more general view that it's an example of, namely, the following:

> *The folk-irrelevance view*: Our type-C questions are settled by objective facts about some part of reality that has nothing to do with us—i.e., nothing to do with our usage, intentions, mental representations, and so on.

When we set this more general view along side the ordinary-language view and the hybrid view, we get a lot closer to an exhaustive list of views.[3]

You might wonder who holds the above views. This is not an easy question. For while lots of philosophers work on type-C questions, it's difficult to find discussions of the above meta-question. However, I have been interested in this question for years, and I have put the question to many, many philosophers who work on type-C questions. My sense from these discussions is that most philosophers who work on type-C questions endorse either the hybrid view or some version of the folk-irrelevance view (usually something like the decompositional view, but not always). The reason for the popularity of the folk-irrelevance view, I think—or one of the reasons—is that a lot of philosophers seem to have a deeply held view that type-C questions are completely objective in the sense that they're not about our ways of conceptualizing things like free will and justice and knowledge; rather, on this view, the what-is-free-will question is about *free will itself*; and the what-is-justice question is about *justice itself*; and so on. Philosophers who talk this way—or who distinguish the project of analyzing our concept of C (e.g., our concept of free will, or justice, or whatever) from the project of uncovering the true nature of C itself (e.g., free will itself, justice itself, etc.)—include Sosa (2007), Kornblith (2013), Williamson (2007), and Stich and Tobia (forthcoming). Another

[3] It's still not entirely exhaustive because you might endorse a non-factualist view of type-C questions; i.e., you might claim that type-C questions aren't settled by facts at all and that they don't have factual answers.

example of someone embracing the folk-irrelevance view of at least some type-C questions is McKenna (2014), who says, "I do not think that even part of what we philosophers are doing in settling on what free will is should be construed in terms of what ordinary folk mean by…the expression 'free will'." And in private conversation, he says that he's "embarrassed to admit" that he leans toward something like the decompositional view. Finally, for some examples of people who endorse something like the hybrid view, see Goodman (1955) and Rawls (1971), and for a more recent example, see Murray and Nahmias (2014).

4 Why the decompositional view is false

I'm eventually going to argue that the ordinary-language view is right. But I want to start by arguing that the decompositional view is false, i.e., that our type-C questions are not settled by the decompositional structures of concepts. More specifically, I want to argue that even if we assume that (a) the concepts that we're asking about in type-C questions really exist in platonic heaven, and (b) these concepts *have* decompositional structures, and (c) an answer to a type-C question is correct iff it captures the real decompositional structure of the given concept, we still shouldn't say that type-C questions are *settled* by the decompositional structures of concepts. They're not. Indeed, they're not even *partially* settled by this, and so advocates of the hybrid view should not put the decompositional structures of concepts on their list of things that are relevant to type-C questions.

To bring out my argument here, let's focus on the type-C question about free will, and to simplify things, let's assume that we've narrowed the list of possible answers down to two, the Humean answer and the libertarian answer. Given this, we can construct the following initial argument for my thesis:

> *Initial argument for the claim that type-C questions are not settled by the decompositional structures of concepts*: If we assume a platonistic view of concepts, we get the result that *both* concepts exist, i.e., that in platonic heaven, there's a concept of *Hume-freedom* and a concept of *libertarian-freedom*. So the idea that the dispute is settled by the decompositional structure of *the* concept—i.e., the one that exists in platonic heaven—is misguided. Moreover, notice that we already know the decompositional structures of the two concepts; *Hume-freedom* is (roughly) the ability to do what you want, and *libertarian-freedom* is (roughly) the ability to make decisions that are simultaneously undetermined and appropriately non-random. (Indeed, this is all *stipulated*; *Hume-freedom* and *libertarian-freedom* just *are* the concepts that correspond to the definitions that we've given of them.[4]) But if we already know the decompositional structures of the two concepts, then since we still don't know the answer to the what-is-free-will question, it's hard to see how that question could be settled by the decompositional structures of the two concepts. In order to settle the

[4] Of course, different philosophers give slightly different definitions of these terms, so there are actually more than two concepts that can be treated as Humean and libertarian kinds of free will. But we can ignore this complication here.

> what-is-free-will question, we don't need to gain a better understanding of the decompositional structures of the two concepts; what we need to figure out, it seems, is *which* of the two concepts—or better, *which of the two decompositional structures*—is the concept of free will.

One might respond to this initial argument by saying something like the following:

> When we think about free will, we so to speak "grab hold" of a certain concept, and we do this *without specifying its decompositional structure*. Then when we do conceptual analysis, we try to uncover the decompositional structure of *that* concept—i.e., the one we've got hold of, or the one we're thinking of—and so the dispute *is* settled by the decompositional structure of the relevant concept.

This is confused. To see why, notice that (a) the question, 'What is the decompositional structure of *that* concept, i.e., the one we have in mind?' is just equivalent to the question, 'Which of the two concepts (i.e., which of the two decompositional structures) is the one that we have in mind?'; and (b) the latter question is a question about *us*. In particular, we can think of it as a question about our mental representation of the concept of free will. If our mental representation has a Humean structure, then the concept that we have in mind (or that we've "got hold of") is Hume-freedom, and so "that concept" is Hume-freedom; and if our mental representation has a libertarian structure, then the concept that we have in mind is libertarian-freedom, and so "that concept" is libertarian-freedom. So which concept counts as "that concept" is literally *determined* by what's in our heads; if our mental representation is, e.g., Humean, then that *makes it the case* that the concept of Hume-freedom counts as "that concept".

Now, of course, people who endorse the hybrid view will want to say that which concept counts as "that concept" is determined not just by what's in our heads, but by other things as well, like logical coherence, or Lewisian naturalness. But it's hard to see how the decompositional structures of the two concepts could be relevant at all here. How could the decompositional structures of *Hume-freedom* and *libertarian-freedom* be relevant to the question 'Which of these two decompositional structures is the one that's picked out by our mental representation of free will?'?

Think of it this way: the concepts themselves have been sitting in platonic heaven forever, and before we came along, neither was singled out as "that concept", or "the one we've grabbed hold of". Thus, if today one of them *is* singled out in this way, this is presumably because we started having free will thoughts of a certain kind. It's not because of the decompositional structures themselves. Neither concept is, in itself, singled out as *the* concept of free will. And this suggests that the decompositional structures of the two concepts aren't what determines—indeed, aren't even part of what determines—which of the two structures counts as the concept of free will.

Here's another argument for the same conclusion: if what I'm saying here wasn't right, we would have no way of answering our type-C questions. The reason we can figure out which concept counts as "that concept" is that this is determined by facts about us. Here's an analogy: philosophers of mathematics sometimes ask how humans could know which platonic structure counts as the *intended* structure for arithmetic; but there's an easy answer to this question—we can know which structure counts as the intended structure because this is determined by facts about *us*, in particular, our

intentions. Likewise, we can know which concept counts as *the* concept of free will because this is determined by facts about us, in particular, our semantic intentions regarding 'free will', or what comes to the same thing, our mental representation of the concept of free will.

5 A quick argument for the relevance of facts about the folk and, hence, for the legitimacy of x-phi

Having dispensed with the decompositional view, I now want to argue against the more general view that it's an example of—i.e., the view that in Sect. 3 I called the "folk-irrelevance view". More precisely, I want to argue that either the ordinary-language view or the hybrid view is correct; or to put the point differently, I want to argue for the following claim:

> *Folk Relevance*: Facts about the folk (most importantly, their *heads*, in particular, their semantic intentions and mental representations) are at least *among the facts* that determine the answers to our type-C questions.

The arguments of Sect. 4 are already highly suggestive of this result, but I want to provide a more direct argument for it. Indeed, I'll give *three* arguments here.

My first argument for Folk Relevance is that we need to endorse this view to account for the methodology that philosophers actually use when they try to answer type-C questions. One of the main things that philosophers do here is use our intuitions (about the applicability of our concepts in real and imagined scenarios) to confirm and falsify theories. For instance, if a given theory of free will—i.e., a given answer to the what-is-free-will question—flies in the face of our intuitions about when the concept of free will applies, then this is seen as falsifying evidence. But this methodology would make little sense if we weren't at least partially engaged in trying to uncover facts about what we mean by our words. For while it's plausible to suppose that our intuitions reliably track facts about what we mean by our words (i.e., about our semantic intentions), it's not plausible to suppose that they reliably track other kinds of facts.

The reason our intuitions reliably track facts about our semantic intentions and mental representations is that our intuitions are *generated* by our intentions and mental representations. My mental representation of the concept C (and my intentions about how to use the term 'C') guide my usage of 'C'. And the intuitions we're talking about here are just intuitions about how to use 'C', i.e., about whether 'C' applies in some imagined scenario. It's well known that you can be a competent user of a word 'C' even if you can't articulate a definition of 'C'; in other words, you can know how to apply 'C' without being able to say exactly what we mean by 'C'. But given this, our intuitions (about when 'C' does and doesn't apply) can be seen as windows into what we mean by 'C'; and so they can be used as data points to confirm and falsify theories of what we mean by 'C'.

My second argument for Folk Relevance proceeds as follows:

(i) Type-C questions are questions about the meanings of the relevant words (e.g., the type-C question 'What is free will?' is essentially equivalent to the question "Which concept is the meaning of the term 'free will'?"). But

(ii) Facts about the folk (in particular, about their semantic intentions and mental representations) are, at the very least, among the facts that determine what our words mean. Therefore,
(iii) Facts about the folk are at least among the facts that determine the answers to our type-C questions.

Premise (i) seems essentially definitional to me, and it's hard to see how it could make any sense at all to deny premise (ii). If we tried to do this, we'd wind up saying things like this: "Well, of course, we use the term 'cat' to refer to those cute little animals, but it turns out we're wrong about this—'cat' really refers to bowling balls." This is, of course, nonsense; it can't be that our semantic intentions are completely out of touch with what our words actually mean, because our semantic intentions at least partially *determine* what our words mean.

Finally, my third argument is that if Folk Relevance weren't true—if the answers to our type-C questions weren't at least partially determined by facts about folk intentions and mental representations—then these questions wouldn't be constrained by facts at all. To see why I say this, consider the following theory:

FW: A person P has *free will* iff P has the ability to dunk a basketball.

This theory is obviously false. But *why*? After all, there's nothing *inherently* wrong with FW. So why is it an obviously false non-starter? The only plausible way to answer this question is to say that the reason FW is obviously false is that it's completely out of touch with what we mean by 'free will'. But if we say this, then we're committed to saying that facts about what we mean by 'free will' are relevant to the what-is-free-will question. If these facts *weren't* relevant to that question, then we couldn't object to FW in the above way, and it's hard to see how else we could object to it. Indeed, it seems to me that if we don't endorse the idea that facts about folk usage and intentions are relevant to our type-C questions, then we won't be able to maintain that these questions are *factual* at all. Anything would go in this scenario. I could just stipulate any definition of 'free will' that I happened to like, and it's hard to see how you could argue that I was wrong.

So we have good reason to endorse Folk Relevance, i.e., to think that facts about the folk—about their semantic intentions and mental representations and so on—are relevant to our type-C questions. But this suggests that x-phi is a legitimate methodology to use in connection with type-C questions because x-phi is just an attempt to gather data about the folk (in particular, it's an attempt to gather data about folk intuitions regarding the applicability of our concepts, and these intuitions can be used as data points to confirm and falsify theories about what the folk mean by their words—i.e., about their semantic intentions, mental representations, and so on).

So we now have the result that x-phi is a legitimate methodology to use for type-C questions. But I want to argue for a stronger result. I want to argue that the ordinary-language view is true, so that facts about folk intentions and representations and so on are the *only* facts that determine the answers to our type-C questions.

6 Why the ordinary-language view is correct

Let's think of ourselves as trying to construct a list of all the kinds of facts that are relevant to determining the answers to our type-C questions. So far, we know that facts about folk intentions and representations belong on the list. The hybrid theorist thinks that other kinds of facts belong on the list as well. I disagree. I think the ordinary-language view is right; I think the *only* facts that belong on the list are facts about *us*—about our intentions, representations, and so on. I will argue this point by running through the various kinds of facts that you might think are relevant (i.e., that you might think are among the facts that determine the answers to our type-C questions) and arguing that they are all in fact *not* relevant. Now, I've already argued that the decompositional structures of concepts aren't relevant; but in this section, I will consider various other kinds of facts as well. (I don't have an argument for the claim that the list of fact types that I consider here is exhaustive; I just can't think of any other kinds of facts that one might reasonably think are relevant to our type-C questions.)

Before I start, let me make one point very clear. I am not suggesting here that actual working conceptual analysts *believe* the ordinary-language view. I'm fully aware that lots of conceptual analysts would reject this view. But what I would claim is that these people are simply mistaken in their meta-views about their own work.

6.1 Facts about common nature

It's a trivial truth that an answer to the type-C question 'What is C?' is correct iff it captures what all and only (actual and possible) Cs have in common. And given this, one might think that the facts that settle our type-C questions are facts about the common nature of the (actual and possible) Cs. But this can't be right. If we help ourselves (for the sake of argument) to Lewisian realism about possibilia, we can bring this point out very clearly. The problem is that (a) every analysis of C provides a theory of what the instances of C have in common, and (b) for each such theory, there's a set of objects of the relevant kind, i.e., a set of (actual and possible) objects that have the given trait in common.[5] Consider, e.g., a dispute between Hume and Frankfurt about what free will is. The facts that determine which of these views (if either) is correct can't be facts about what the various free choices have in common, because we don't have an independent handle on what the free choices *are*. There's a set of (actual and possible) Hume-free choices (and what they have in common is being Hume-free), and there's a set of (actual and possible) Frankfurt-free choices (and what they have in common is being Frankfurt-free), and we don't have an independent handle on which of these sets (if either) is the set of *free* choices. Indeed, that's precisely what we're trying to determine when we ask the what-is-free-will question; the question 'What is free will?' is just *equivalent* to the question, 'Which set of (actual and possible) choices is the set of *free* choices?' Put differently, the point here is that facts about what the Cs

[5] Of course, if an analysis picks out an incoherent concept, call it C1, then there won't be *any* (actual or possible) objects that instantiate C1, and you might think that facts like this could be relevant to whether C1 is identical to the relevant folk concept C, i.e., the one we're trying to analyze. I will consider this suggestion below, in Sect. 6.5.

have in common can't determine which answer to the what-is-C question is correct, because the what-do-the-Cs-have-in-common question isn't, so to speak, "far enough away" from the what-is-C question; these two questions have to be answered together.

6.2 Facts about relations to other concepts

You might think that facts about whether a given concept is related in some way to some *other* concept can be relevant to whether it's identical to some folk concept. For example, it's plausible to suppose that free will is required for moral responsibility, and so you might think that facts about whether Hume-freedom and libertarian-freedom and so on are required for moral responsibility are relevant to determining which of these concepts *is* free will.

But this just moves us back to a different type-C question. On this view, the answer to the question 'What is free will?' depends on the answer to the question 'Which kinds of freedom are required for moral responsibility?' But this is just to say that it depends on the answer to the question 'What is moral responsibility?' There's a concept of *Hume-responsibility* (that requires Hume-freedom); and a concept of *libertarian-responsibility* (that requires libertarian-freedom); and so on. We need to determine which of these concepts provides the right definition of 'moral responsibility'. But, of course, this is just another type-C question, and so we have to ask our meta-question all over again. In other words, we need to ask what kinds of facts settle the question 'What is moral responsibility?' (or, in particular, 'Which kinds of freedom are required for moral responsibility?')—and when we do this, we'll be right back where we started. In particular, we'll know that facts about us (about our mental representation of the concept of responsibility, about our usage and intentions concerning 'responsible', and so on) are relevant, and we'll need to determine whether any other kinds of facts are relevant.

Now, I suppose you might say that the what-is-moral-responsibility question is a *moral* question, not a semantic question. In other words, you might think that the question here is this: 'Which concept of moral responsibility (and which concept of free will) *should* we employ?' But which *should* are we talking about here? There is presumably a *Hume-should* (and we can presumably say that we Hume-should employ the concept of *Hume-responsibility*); and there's a *libertarian-should* (and we libertarian-should employ the concept of *libertarian-responsibility*); and so on. To figure out what the *real* should is, we need to answer another type-C question. Etc., etc., etc.

6.3 Facts about our environment

You might think that facts about our physical environment could be relevant to determining the answers to certain type-C questions. For instance, you might think that the fact that water is made of H_2O is relevant to the question, 'What is water?' But as I've already explained, facts like this are *not* relevant to type-C questions. Type-C questions are *conceptual* questions; e.g., the type-C question about water doesn't ask

what water is made of; it asks what the concept *water* is, and the environmental fact that water is made of H_2O isn't relevant at all to that question.

6.4 Facts about pragmatic usefulness

Suppose that C is a folk concept and that for some pragmatic reason, we need to provide a precise definition of the term 'C' (e.g., we might want to use 'C' in our legal system, or in a theory we're constructing, or we might just want to make sure we're not talking past each other in our use of 'C'). And suppose that we have two competing definitions, one that takes 'C' to express the concept C1, and another that takes 'C' to express C2. Given this, you might think that if the C1-definition does a better job of satisfying our pragmatic needs (e.g., if it leads to a better legal system, or a better theory, or whatever), then all else being equal, C1 is a better candidate than C2 for *being C*, or for being the meaning of 'C'.

But it's hard to believe that pragmatic considerations of this kind are relevant to type-C questions. On the contrary, it seems that such considerations are best thought of as giving us reasons to *stipulate* certain definitions of our words, counterintuitive consequences be damned. Suppose, for instance, that we're trying to come up with a definition of 'reasonable doubt' for use in our legal system, and suppose we're trying to decide whether to define it as RD1 or RD2, where these are precisely defined concepts. Finally, suppose that if we use RD1 in our instructions to jurors, we'll have a better or more just legal system than if we use RD2—in particular, suppose that juries will do a better job of convicting guilty people and not convicting innocent people. Is this reason to think that RD1 captures what it *means* to have a reasonable doubt? Surely not. It may be reason to use this definition in our legal system, but that's another matter. There's the question of what we should tell jurors in order to have the most just legal system, and there's the question of what a reasonable doubt really *is*, and the two questions might not have the same answer.

6.5 Facts about coherence and instantiation

Suppose that (a) the ordinary usage and intentions associated with 'C' pick out a concept C1 that's incoherent (or uninstantiated), and (b) C2 is a coherent (and instantiated) concept that, so to speak, "cleans up" ordinary usage and intentions in the best overall way. Then you might think that, all else being equal, C2 is a better candidate than C1 for *being C*, or for being the meaning of 'C'. And so you might think that facts about logical coherence (or instantiation) are among the facts that determine the answers to our type-C questions.

This view—that facts about coherence and/or instantiation are relevant to our type-C questions—is implausible. The problem, in a nutshell, is that it makes perfect sense to say that a concept of ours is incoherent and uninstantiated. For instance, it makes sense to say that the concept of free will is incoherent (or uninstantiated) because (a) free will is libertarian-freedom, and (b) libertarian-freedom is incoherent (or uninstantiated). But if it were built into the correctness conditions of analyses of free will that the correct analysis has to pick out a concept that's coherent and instantiated, then this

view—i.e., the (a)–(b) view—*wouldn't* make sense. In this scenario, the claim that (b) was true (i.e., that libertarian-freedom was incoherent or uninstantiated) would undermine the claim that (a) was true (i.e., that free will was libertarian-freedom). But in point of actual fact, (b) *doesn't* undermine (a); the two are perfectly compatible.

It's not just that the (a)–(b) view *make sense*; the truth is that it's a *standard* view in the literature. Numerous philosophers have argued that the concept of free will is incoherent or uninstantiated because the libertarian analysis is right and libertarian-freedom is incoherent or uninstantiated.[6] But when was the last time you saw someone argue like this: "Libertarian-freedom is incoherent (or uninstantiated); therefore, free will isn't libertarian-freedom"? I don't know of anyone who has ever given this argument. And the reason, I think, is obvious—this just isn't how you play the conceptual-analysis game. The fact that an analysis takes a folk concept to be incoherent or uninstantiated is simply not a reason to think that it's not the right analysis of that concept.

Now, it may be that in some cases, the fact that some folk concept C is incoherent or uninstantiated gives us good reason to *revise* the given concept, i.e., to change what we mean by 'C'; e.g., this is what Jackson (1998, pp. 44–45) and Vargas (2009) say about free will. But there's a difference between (i) suggesting that we should change the meaning of 'C' and (ii) trying to figure out what 'C' actually means right now; and the point I'm making here is that facts about coherence and instantiation aren't relevant to questions about what our words actually mean right now.

Another problem with the view I'm discussing here—the view that facts about coherence and/or instantiation are relevant to type-C questions—is that it seems to involve a kind of semantic *cheating*. As we've seen, type-C questions are semantic questions. There's no difference between asking what free will is and asking which concept is expressed by the term 'free will'. But given this, it would seem very odd to build coherence and instantiation into the correctness conditions of an answer to a type-C question. It would seem to be a kind of cheating. On this view, we could never mistakenly talk about objects of a kind that didn't actually exist; for the very fact that they didn't exist would make it the case that we weren't talking about them. "How convenient," a critic might say. While we're at it, maybe we should also stipulate that whenever we utter a sentence, the proposition that's expressed is the "nearest *true* proposition"—i.e., the proposition that, among the true propositions, is "nearest" in content to the proposition that actually corresponds to our semantic intentions concerning the sentence. This would be very convenient as well—it would make it impossible for us ever to say anything false. But, of course, this is silly. We can make mistakes. We can utter sentences that express propositions that, unbeknownst to us, are false. And we can use terms that express concepts that, unbeknownst to us, are uninstantiated or incoherent.

6.6 Facts about Lewisian naturalness

You might think that facts about Lewisian naturalness are relevant to type-C questions. Lewis talks about naturalness in connection with properties, but most of what

[6] See, e.g., Broad (1952), Strawson (1986), and Pereboom (2001).

he says carries straight over to the case concepts. We can think of the naturalness of concepts in two different ways. On one way, to say that a concept C1 is more natural than another concept C2 is to say that C1 does a better job of "carving nature at the joints"; i.e., it's to say that the instances of C1 *resemble* each other more than the instances of C2 do, or that they have a higher degree of intrinsic qualitative similarity. On a second way of thinking about naturalness, to say that C1 is more natural than C2 is to say that it's less *arbitrary*, or less *gerrymandered*, i.e., that its decompositional structure is simpler, or less disjunctive, or some such thing, so that it can be defined in a shorter passage, with fewer terms, and so on. So, to provide an example, on both conceptions of naturalness, we can say that the concept *green* is more natural than the concept *green or a Tuesday in May or a prime number between 8 and 24*.

(Lewis writes as if the two conceptions of naturalness are equivalent; I think this is false, but for our purposes, it won't matter.[7] Also, while the term 'natural' is supposed to be suggestive of natural kinds, facts about naturalness are not contingent upon facts about the actual world; on the contrary, as Lewis points out in connection with properties, concepts are natural and unnatural *in themselves*, independently of any contingent facts.)

In any event, the suggestion I want to consider now is that facts about naturalness are relevant to type-C questions. For instance, you might think that if C1 and C2 fit equally well with the folk usage of 'C', and if C1 is more natural than C2, then all else being equal, C1 is a better candidate than C2 for being C (or for being the meaning of 'C').[8]

I have two reasons for rejecting this view. First, the idea that facts about naturalness are relevant to our type-C questions involves a semantic *cheat*—in the same way that it would be a cheat to use facts about coherence and instantiation. If our usage and intentions regarding the term 'C' fail to zero in on a unique concept—if they fit equally well with C1 and C2—then the right thing to say is that 'C' is imprecise and that there's some semantic indeterminacy surrounding that term. The idea that 'C' automatically expresses the more natural concept is a cheat. Why should it refer to the more natural concept if *we* don't do anything to *make* it refer to that concept? After all, it's *our* word. Suppose, for instance, that our usage and intentions regarding 'rabbit' are neutral between the concept *rabbit* (or *unified, entire rabbit*, or some such thing) and concepts like *rabbit stage* and *simples arranged rabbitwise*. I doubt this is actually the case—I think our mental representation corresponds more closely to the former concept than to the latter two concepts—but that's another matter. The point I want to make is that if our intentions and representations *did* fail to zero in on a unique concept, then the right thing to say would be that there's some semantic indeterminacy here. It would be a straightforward cheat to claim that 'rabbit' expresses the concept

[7] Consider the concept *green and a Toyota*; this seems more gerrymandered than *green* is; at any rate, it's more arbitrary and it has a more complicated decompositional structure; but green Toyotas resemble each other more than green things do.

[8] The idea that facts about naturalness can be relevant to questions about meaning is sometimes called the *best-candidate theory of meaning*. This idea was first introduced by Merrill (1980) and developed by Lewis (1983, 1986). See also Sider (2001).

rabbit (or *unified, entire rabbit*) because that concept is more natural than the other two concepts.[9] How could it be right to say that *our* term 'rabbit' picks out a unique concept if *our* intentions concerning that term are genuinely neutral between multiple concepts? If *our* intentions don't zero in on a unique concept, then we're imprecise in connection with 'rabbit' and so there's some semantic indeterminacy regarding that term. Using an appeal to naturalness to avoid this result is a kind of cheating; it's analogous to using truth as a means of deciding which propositions our sentences express.

The second reason for rejecting the idea that facts about naturalness are relevant to our type-C questions is that it's out of touch with the actual methodology used by real conceptual analysts. In other words, conceptual analysts don't actually use claims about naturalness to settle their disputes. Contemporary metaphysicians have put the concept of naturalness to several different uses, but I don't know of anyone who's used it in a type-C dispute. After all, there isn't a real type-C dispute in the philosophical literature about what a rabbit is. And if we look at *real* type-C disputes—about concepts like *free will*, *knowledge*, *personhood*, and so on—we just don't find people appealing to facts about naturalness to argue for their views.

6.7 Running out of ideas

I can't think of any other kinds of facts that might settle (or partially settle) our type-C questions. So I want to tentatively conclude that the only facts that are relevant here are facts about *us*, about our semantic intentions and mental representations and so on. In other words, I want to tentatively conclude that the ordinary-language view is right.

Now, to repeat something I said at the start, I'm aware that a lot of conceptual analysts would reject the ordinary-language view. But for the most part, I think these people are just mistaken about the nature of their own work. Philosophers often pay lip service to the idea that when we do conceptual analysis, we're moving toward a sort of reflective equilibrium, balancing the need to preserve as many ordinary intuitions as we can with other more theoretical goals. But I suspect that this is usually just an excuse to justify their need to bite the bullet with respect to certain intuitions that they can neither save nor explain away.

Finally, I want to acknowledge that it may be that *some* conceptual analysts really do use facts of some of the kinds I've considered here—e.g., facts about logical coherence, or pragmatic usefulness, or whatever. But if this occurs at all, I think it's very rare, and what's more, I think that if any philosophers use such considerations, then they're talking past the people they're arguing with, because most mainstream conceptual analysts just *don't* use facts of these kinds to justify their theories.

[9] I actually don't think it's true that *rabbit* (or *unified, entire rabbit*) is more natural than *rabbit stage* or *simples arranged rabbitwise*, but that obviously doesn't matter here.

7 Three worries

7.1 The objection from analyticity

If what I've been arguing is correct, then our conceptual-analysis questions are about folk meaning, and so they're *empirical*. But one might object to this by saying something like the following:

> The sentences that give the answers to our type-C questions—sentences like 'Bachelors are unmarried men', 'Water is the actual watery stuff of our acquaintance', and so on—are *analytic*, or *conceptually true*. But if they're analytic, then they're a priori, and so conceptual analysis is an a priori affair. Thus, our type-C questions are not empirical. (Jackson (1998) argues in a similar way for the a priority of conceptual analysis.)

I want to start my response to this objection by agreeing that sentences that give the answers to our type-C questions (i.e., sentences like 'Bachelors are unmarried men' and 'Water is the actual watery stuff of our acquaintance') are analytic and a priori. But we need to be very clear about what it means to say that a sentence is a priori. It's well known that whenever a sentence is true, the reason it's true is that there obtain certain *meaning-facts* and certain *world-facts*. For instance, the reason 'Snow is white' is true is that (a) it means that snow is white, and (b) snow *is* white; thus, e.g., if snow had been orange, then 'Snow is white' wouldn't have been true, and likewise, if that sentence had meant that, say, Mars is flat, then it also wouldn't have been true. Now, when a sentence is analytic, the given *world-facts* are knowable a priori. But the meaning-facts behind an analytic sentence *aren't* knowable a priori; indeed, meaning-facts of this kind are *never* knowable a priori. When we say that a sentence is a priori, what we mean is that the relevant world-facts are knowable a priori; or to put the point another way, what we mean is that we can know a priori that the sentence is true, *given knowledge of what it means*. Thus, in ordinary philosophical lingo, it's correct to say that the answers to our type-C questions (sentences like 'Bachelors are unmarried men') are a priori; but it doesn't follow that we can know a priori that, e.g., 'bachelor' means *unmarried man*, and in fact, we can't know this a priori, and so there's no problem here for my view that type-C questions are empirical. In short, the claim that type-C questions are empirical is perfectly compatible with the fact that the sentences giving the answers to these questions are a priori because (a) to say that these sentences are a priori is to say that the relevant *world-facts* are knowable a priori, and (b) type-C questions aren't about these world-facts—they're about the *meaning-facts* behind these sentences.

You might wonder why, if conceptual analysis is empirical, the methodology of conceptual analysts (or at any rate, the traditional methodology—i.e., the methodology we employed before x-phi came on the scene) isn't empirical. My response is that the traditional methodology *is* empirical. The traditional methodology is to use the intuitions of native speakers of the given language as data points and to construct theories (i.e., conceptual analyses) that account for these data. If a given theory (i.e., a given analysis) runs counter to the data (i.e., to intuitions), then this falsifies the

theory; and if a theory fits with the data (i.e., with intuitions), then this confirms the theory. If this isn't an empirical methodology, I don't know what is.

Doing conceptual analysis can sometimes *feel* a priori. But this is just because we have first-person epistemic access to some of the empirical data. As we saw above, the facts that determine what words like 'water' mean are facts about our psychology, about what we mean by our words. Now, facts about what we mean by our words aren't always accessible to introspection, but some of the *data* about what we mean—namely, our intuitions about the applicability of our words—*are* accessible to introspection. Thus, if I'm a competent user of 'C', then I can move from (a) facts to which I have first-person epistemic access (namely, my intuitions about whether 'C' is applicable in various scenarios) to (b) a theory of what 'C' means in English. So this can feel a priori in a certain way. But it's not a priori, it's empirical. Try figuring out what some word means in *Hungarian*, and then tell me it's a priori.[10]

7.2 Abstract objects

Here's a second worry one might raise about my view:

> *The abstract object worry*: If type-C questions are about concepts, and if concepts are abstract objects, then how could type-C questions be empirical?

On the view I'm proposing, we can state the answers to our type-C questions with sentences like the following:

> (O) Ordinary folk use the term 'water' to express the concept *the actual watery stuff of our acquaintance*.

This sentence is about an abstract object (in particular, a concept), but it's also about the physical world (in particular, ordinary folk), and so there's nothing wrong with saying it's empirical. We can appreciate this by looking at a different sort of case. Consider, e.g., the following sentence:

> (M) The Nile is 4,130 miles long (i.e., the Nile stands in the mile-relation to the number 4,130).

Like (O), this sentence is about the physical world (in particular, the Nile) and also an abstract object (namely, the number 4,130). But (M) is clearly empirical. The only reason we refer to numbers in length ascriptions like (M) is to help us state empirical length facts about physical objects. The reason it's convenient to do this is that there's a 1–1 correspondence between the positive real numbers and the possible length properties, and so, in essence, we can use the real numbers as *names* of the length properties. The same can be said about sentences like (O): the purpose of such sentences is to state empirical facts about the usage and intentions of ordinary folk; we just use talk of concepts to help us do this. And the reason this is convenient is that

[10] Even if I stop trying to figure out what 'C' means in English and just try to figure out what it means in my own private language, this is still empirical. For, again, we don't in general have a priori epistemic access to what we mean by our words.

there's an obvious correspondence between the set of concepts and the set of possible meaning intentions that people can have concerning their words.[11]

7.3 The negative project of x-phi

On the view I've argued for, type-C questions are settled by facts about what we mean by our words, and our intuitions about the applicability of those words are relevant data points in constructing theories of what the relevant words mean. But the so-called negative project of x-phi can be seen as raising worries about how much we should trust our intuitions. For instance, there are a number of studies that purport to show that people's intuitions about philosophical cases vary across different cultures; see, e.g., Weinberg et al. (2001) and Machery et al. (2004). Some philosophers question these results—see, e.g., Mortensen and Nagel (forthcoming)—but I just want to point out that even if it's true that there are cross-cultural differences in our intuitions, this wouldn't undermine my view, because we could simply take this to show that people from different cultures sometimes have different concepts (Jackson (1998) makes a similar point). Now, one might claim that this response won't work if the goal of conceptual analysis is to uncover the structures of concepts like free will, and knowledge, and so on, as opposed to the structures of some specific culture's mental representations of these concept (Stich and Tobia (forthcoming) say something roughly like this). But if the arguments of the present paper are right, then there is no substantive difference between uncovering the structure of some concept C and uncovering the structure of our mental representation of C; this is because, as we've seen, the structure of a given concept (free will, or knowledge, or whatever) is metaphysically *determined* by the structure of our mental representation of that concept. Now, of course, if the members of some other culture have a different mental representation, this doesn't mean that their representation is mistaken; rather, on the view I'm suggesting, they have an *accurate* representation of a *different* concept.

8 If concepts were mental objects…

I've been assuming that concepts are (supposed to be) abstract objects, and I've argued that (a) the ordinary-language view is right, and (b) x-phi is a legitimate methodology to use in connection with type-C questions. I now want to point out that if we abandon the platonistic view of concepts and endorse the idea that concepts are mental objects, my arguments still go through. On the view I've been presenting, even if concepts are abstract objects, the facts that determine the answers to our type-C questions are fact about *us*—in particular, our mental representations of concepts. But if concepts are mental objects, then what I've been calling mental representations just *are* concepts. Thus, on this view, we still get the result that the facts that settle our type-C questions

[11] This doesn't matter here, but we don't need to endorse the existence of abstract objects to endorse this view of sentences like (O) and (M). For we can endorse a fictionalistic view according to which these sentences are strictly speaking false (because they refer to non-existent objects) but useful fictions. I defend this view in Balaguer (1998) and (2009).

are facts about us. Indeed, they're the very same facts. It's just that now, instead of calling them facts about our mental representations of concepts, we'll call them facts about concepts. But they're still the same facts. And so the same methodologies will be relevant to uncovering these facts. And so we'll still get the result that x-phi is legitimate in this context. Indeed, this conclusion follows a lot more easily if we assume that concepts are mental objects.

9 Two lingering worries

9.1 X-phi versus traditional conceptual analysis

I've argued that x-phi is a legitimate method to use in connection with type-C questions. But I think the traditional methodology is also legitimate. Indeed, on the view I've presented here, they're essentially the *same* methodology—they're both based on the use of native-speaker intuitions as data points. But there's a problem with the traditional methodology. The intuitions of philosophers have (very likely) been affected by the doing of philosophy. More specifically, if you've thought about what some term means, and if you've come to accept a theory of what that term means, then once you've accepted the theory, we can expect that your intuitions will line up with that theory. But this doesn't mean that that's how your intuitions would have gone before you started your investigation. Maybe the accepting of the theory went hand-in-hand with an alteration in your intentions about what the term in question means. And if so, then your intuitions will have changed.

Unfortunately, there's a corresponding problem with x-phi. It takes a bit of training to teach someone how to play the intuition-thought-experiment game. When you give thought experiments to ordinary non-philosophers and ask for their intuitions, they often don't understand what you're asking them to do. And this problem is made worse in an experimental situation, where subjects are trying to answer the questions *correctly*. I think they often end up guessing what they think the objectively *correct* answer is; they often don't fully understand that the answer is *within* them.[12]

And so we have a sort of catch-22. You need to train people a bit to teach them how to play the game, but when you do that, you run the risk of altering their semantic intentions and, hence, their intuitions. I'm not saying that there's no way around this problem, just that it's tricky business.

9.2 Guilt by association

I've argued here that x-phi is a legitimate methodology to use in connection with type-C questions. But I haven't argued that you shouldn't roll your eyes at x-phi. What

[12] This is related to a widely discussed question, namely, the question of whether philosophers can be thought of as *experts* in relation to type-C questions. I would hesitate to put it this way; if they're expert at anything in this connection, it's knowing how to call up stable, reliable intuitions that accurately reflect their own concepts. Now, I certainly don't think philosophers are infallible at this, but it's plausible to suppose that they're better at it than most people. For other discussions of these issues, see, e.g., Knobe and Nichols (2007) and Kauppinen (2007).

I've argued is that you shouldn't roll your eyes at x-phi unless you're prepared to roll your eyes at type-C questions as well. So one way to respond to this is to roll your eyes at type-C questions. And this, I think, is a *legitimate* response. Now, of course, if you're interested in empirical semantics—if you're interested in knowing the contingent empirical facts about the meanings of ordinary-language terms—then you shouldn't roll your eyes at type-C questions. And I certainly don't think there's anything wrong with being interested in such questions. But if you're interested in *metaphysics*—if you're interested in discovering non-semantic facts about the nature of reality—then you shouldn't think you're going to make any progress by answering type-C questions. For instance, if you're interested in the problem of free will, and if your interest in that problem is based in a desire to discover non-semantic facts about the metaphysical nature of human decision-making processes, then you shouldn't think you're going to make any progress by answering the what-is-free-will question. You're not. That question is just about *words*, and so by answering it, you're not going to learn anything non-trivial about the nature of human decision-making processes. I can't argue this point here, but I've argued for it at length in Balaguer (2010).

Acknowledgments I would like to thank David Pitt and Talia Bettcher for commenting on an earlier version of this paper. Also, versions of this paper were read at the University of Arizona and at a conference at the University of San Francisco, and I would like to thank the members of those two audiences. Special thanks go to Michael McKenna, Shaun Nichols, Terry Horgan, Uriah Kriegel, Carolina Sartorio and Manuel Vargas.

References

Alexander, J., Mallon, R., & Weinberg, J. M. (2010). Accentuate the negative. *Review of Philosophy and Psychology*, *1*, 297–314.

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. New York: Oxford University Press.

Balaguer, M. (2009). Fictionalism, theft, and the story of mathematics. *Philosophia Mathematica, 17*, 131–162.

Balaguer, M. (2010). *Free will as an open scientific problem*. Cambridge, MA: MIT Press.

Broad, C. D. (1952). Determinism, indeterminism, and libertarianism. In *Ethics and the history of philosophy* (pp. 195–217). London: Routledge & Kegan Paul.

Chalmers, D. (1996). *The conscious mind*. New York: Oxford University Press.

Goodman, N. (1955). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). New York: Oxford University Press.

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations*, *10*, 95–118.

Knobe, J., & Nichols, S. (2007). An experimental philosophy manifesto. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 3–14). Oxford: Oxford University Press.

Kornblith, H. (2013). Is there room for armchair theorizing in epistemology. In M. Haug (Ed.), *Philosophical methodology: The armchair or the laboratory?* (pp. 195–216). New York: Routledge.

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

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics, cross-cultural style. *Cognition*, *92*, B1–B12.

McKenna, M. (2014). The metaphysical importance of the compatibility question: Comments on Mark Balaguer's *Free will as an open scientific problem*. *Philosophical Studies*, *169*, 39–50.

Merrill, G. H. (1980). The model-theoretic argument against realism. *Philosophy of Science*, *47*, 69–81.

Mortensen, K., & Nagel, J. (forthcoming). Armchair-friendly experimental philosophy. In W. Buckwalter & J. Sytsma (Eds.), *Blackwell companion to experimental philosophy*.

Murray, D., & Nahmias, E. (2014). Explaining away incompatibilist intuitions. *Philosophy and Phenomenological Research*, *88*, 434–467.

O'Neill, E., & Machery, E. (2014). Experimental philosophy: What is it good for? In E. Machery & E. O'Neill (Eds.), *Current controversies in experimental philosophy* (pp. vii–xxix). New York: Routledge.

Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.

Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam (Ed.), *Language, mind and reality*. Cambridge: Cambridge University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Sider, T. (2001). *Four dimensionalism*. Oxford: Oxford University Press.

Sosa, E. (2007). Experimental philosophy and philosophical intuitions. *Philosophical Studies*, *132*, 99–107.

Stich, S., & Tobia, K. (forthcoming). Experimental philosophy and the philosophical tradition. In W. Buckwalter & J. Sytsma (Eds.), *Blackwell companion to experimental philosophy*.

Strawson, G. (1986). *Freedom and belief*. Oxford: Clarendon Press.

Vargas, M. (2009). Revisionism about free will: A statement and defense. *Philosophical Studies*, *144*(1), 45–62.

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

Williamson, T. (2007). *The philosophy of philosophy*. Oxford: Blackwell.