# Metaphysics and Conceptual Analysis

*Experimental Philosophy's Place under the Sun*

Uriah Kriegel

## Introduction

What is the rationale for the methodological innovations of experimental philosophy? This paper starts from the contention that common answers to this question are implausible (section 1). It then develops a framework within which experimental philosophy fulfills a specific function in an otherwise traditionalist picture of philosophical inquiry. The framework rests on two principal ideas. The first is Frank Jackson's claim that conceptual analysis is unavoidable in "serious metaphysics" (section 2). The second is that the psychological structure of concepts is extremely intricate, much more so than early practitioners of conceptual analysis had realized (section 3). This intricacy has implications for the activity of analyzing concepts: while the central, coarser, more prominent contours of a concept may be identified from the armchair, the finer details of the concept's structure require experimental methods to detect (section 4).

## 1. The obscure rationale of experimental philosophy

In this section, I raise a question regarding the rationale for experimental philosophy (section 1.1) and argue that the most natural takes on what that rationale is do not withstand scrutiny (section 1.2).

## 1.1 Knobe and Gettier

Perhaps the best-known experimental philosophy study is Knobe (2003). Knobe presented seventy-eight random subjects (non-philosophers spending time in a Manhattan park) with one of two vignettes. The first involved a person's aware-ness of her actions' *negative* side effect ("harm condition"):

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment."
>
> The chairman of the board answered, "I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program."
>
> They started the new program. Sure enough, the environment was harmed.

The second involved a person's awareness of her action's *positive* side effect ("help condition"):

> The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. It will help us increase profits, and it will also help the environment."
>
> The chairman of the board answered, "I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program."
>
> They started the new program. Sure enough, the environment was helped.

The crucial results were: 83 percent of respondents said that the chairman harmed the environment intentionally while only 23 percent said that the chair-man helped the environment intentionally. This is the so-called Knobe Effect. It demonstrates, according to Knobe, a normative element in our folk concept of intentional action.

This paper is motivated by two observations. The first is that by far most philosophers share the majoritarian intuition in both of Knobe's cases, so that Knobe *could* have just written a traditional conceptual analysis paper. He could have just presented his two scenarios *to the reader* and entreat *the reader* to pass a verdict on them; that verdict would have supported the idea of a normative aspect in our concept of intentional action just as much as the non-philosophers' verdict. Indeed, I would surmise that among Knobe's article's first seventy-eight readers, a greater proportion shared the majoritarian intuitions than among Knobe's seventy-eight park-goers. There are various possible reasons for this (more on that later), but one suspects that many park-goers who offered the

minority verdicts did so because they did not fully understand the scenarios (or the task of reporting an intuitive verdict on them).[1] Such "performance errors" were probably avoided by Knobe's first seventy-eight philosopher-readers.

The second observation is that once Gettier had devised his well-known thought-experimental scenarios (Gettier 1963), he could have presented them to his undergraduates or some other non-philosopher subjects. But it is not immediately obvious what intellectual benefit this would have brought to epistemology; I daresay the history of epistemology would look virtually no different. The results Gettier would have obtained would have probably been just noisier versions of the results he did in fact obtain among his philosophical readership.

These two observations raise the question: how much of the significance of Knobe's (unquestionably important) article is really due to Knobe's decision to incorporate an experimental part in it? Since philosophers tend to share the lay intuitions on his cases, it is not immediately clear what extra epistemic weight was obtained from consulting the non-philosophers. So the question presents itself: what was the point of the experimental part of Knobe's paper? And we may generalize: what is the point of experimental philosophy?

Approaches to this question can be placed on a spectrum from more inflationary to more deflationary answers. On the deflationary end are positions that take experimental philosophy (henceforth, x-phi) to *have* no point. On the inflationary extreme are positions that tout x-phi as the only future of philosophy and the end of traditional armchair philosophy. Here I will propose a more moderate answer, though one that leans in the deflationary direction. It may be summarized, very roughly, as follows:

> RATIONALE: There *is* a rationale for experimental philosophy, but it is one that (i) gives it limited importance and (ii) nestles it in a very traditional conception of the philosophical project.

I start, in the next subsection, by considering a number of "natural" answers to the question of the rationale for x-phi. I argue that by and large they do not withstand scrutiny. To restrict the domain of possibilities to something manageable, I focus on views that take x-phi to be specially pertinent to our understanding of concepts, indeed as claiming to replace armchair conceptual analysis by a more modern, more empirically sophisticated investigation of how folk concepts behave. Not everybody agrees that this is essential to what x-phi is about (Knobe forthcoming), much less that it *exhausts* what x-phi has to contribute (Rose & Danks 2013). But since I cannot hope for a completely exhaustive consideration of potential approaches to the "point" of x-phi, and since my own view of the

relevance of x-phi to metaphysics goes through concepts, I focus here on views that tie x-phi to conceptual analysis.

## 1.2  Failed rationales

One rationale often offered for the "experimentalization" of conceptual analysis is that concepts are psychologically real entities, and so whatever else we do, we must study them "empirically" (Knobe, personal communication). The immediate question is what it means to say that an inquiry is "empirical." The topic is vast, of course, but two central symptoms of empirical research are observation and testing. At a minimum, then, given that concepts are psychologically real entities, we should (i) *observe* their causal effects (in practice, intuitions and intuition-reports regarding their extensions) and (ii) *test* hypotheses about their intensions in a way that produces confirmation or disconfirmation (and ultimately verification or falsification).[2] This, it might be suggested, is what the "experimentalization" of conceptual analysis brings to the table.

I want to argue that Gettier's traditionalist paper did not vary significantly from Knobe's x-phi paper along these dimensions. I speculate that Knobe *formed* his hypothesis about the concept INTENTIONAL ACTION simply by noticing his own intuitions about cases such as the one he ultimately presented to his subjects. We may think of this "noticing" as a kind of observation—perhaps introspective observation in a suitably loose sense. In any case, Gettier must have proceeded to his own hypothesis formation through a similar procedure of "observing" his intuitions about cases such as the ones he ultimately presented to the reader. So there is no real difference between Knobe and Gettier as far as the role of observation is concerned. One might think that the real difference pertains to testing: Knobe ran his thought-experimental scenarios by seventy-eight subjects before reporting on them to the community of inquiry, whereas Gettier skipped that step. However, the real intellectual test of both papers, accounting for the impact they have had on epistemology and philosophy of action, was in *the encounter with the readership*. It is because the readers shared immediately Gettier's intuitions that his paper was so influential. Had it been met with mixed or uncertain intuitions, it would be less influential. To that extent, Gettier's hypothesis withstood a crucial *test* of convergence with others' intuitions.[3] More interestingly, I would surmise that Knobe's paper's real influence is *also* due to his readers' reactions rather than his subjects'. Suppose the intuitions reported by Knobe's park-side subjects were the opposite of what they actually were, with 80-odd percent saying that the chairman intentionally *helped* the environment in the help condition and only

20-odd percent saying that he intentionally *harmed* the environment in the harm condition. I speculate that most readers of the ensuing article would be inclined to suspect that something went wrong in the experimental design, the subjects' understanding of the task, the statistical analysis, or some other nonsubstantive aspect of the study. They would *not* conclude—not purely on the basis of the reported results, at any rate—that the folk concept of intentional action is wildly different from what they thought it was. Thus the deeper reason for the Knobe Effect's survival and flourishing on the marketplace of ideas is not that Knobe consulted seventy-eight faculties of intuition rather than one, but that the intuitions he consulted converged with his readership's. To that extent, the hypotheses of both papers are equally "empirically tested."[4]

A second potential rationale for experimentalization might be that it purges inquiry from reliance on such dubious epistemic sources as intuition and introspection and replaces them with a much more respectable methodology.

However, x-phi as such appears to rely on intuition and (broadly introspective) noticing-of-intuition just as much as traditional conceptual analysis. Knobe's study does not ostensibly rely on *his own* intuition and introspection, but it does rely on his random subjects producing intuitions in response to his scenario and noticing the intuitions they produced. Moreover, to repeat, the study's intellectual impact is due primarily to his *readership's* intuition and introspection. And in any case, it is doubtful that Knobe's own intuition played no role in the inquiry. At the very least, it played a role in the *context of discovery*, even if not in the *context of justification* (to use Reichenbarch's old distinction). The reason Knobe ran the experiment he did, with the scenarios he did, rather than another experiment with another pair of scenarios, is presumably that he realized, using his own intuition, the interesting effect he later appealed to seventy-eight random faculties of intuition to reproduce. Those subjects' intuitions are relevant to the context of justification, but his own intuition was relevant to the context of discovery.

A third potential rationale for experimentalization might be the risk of divergence between philosophers and non-philosophers. Although in Knobe's original study the two groups' intuitions converged, in other cases they may not. Indeed, Starmans and Friedman (2012) have argued that the folk attribute knowledge in Gettier cases! Where such divergence occurs, we should want to *know about it*, and arguably, it is x-phi's task to expose where this happens. Moreover, it would be reasonable to suggest that wherever such divergence occurs, x-phi discovers what the folk concept *really* is, as opposed to what the philosophers' concept is, or what the philosophers have made the folk concept out to be.

This rationale is more plausible than those considered thus far. But if this were the only rationale for experimentalization, the overall impact of the latter would be limited—perhaps minuscule. For one thing, divergence between intuitions reported in surveys and philosophers' intuitions is quite limited (Dunaway, Edmonds, & Manley, 2013); Starmans and Friedman's claims about folk reactions to Gettier cases, for example, have come under forceful critique (Nagel, San Juan, & Mar, 2013). More deeply, where such divergence in intuition-reports does occur, there is an open question as to what lesson we should extract from it. As noted, folk intuitions about Knobe's two cases are just noisier versions of philosophers' intuitions, and one natural explanation of this is that non-philosophers' results involve more performance errors. As Jackson (2008) puts it, Gettier cases are just hard to process immediately, and the most philosophically interesting reactions to them are the *considered* responses. On this line of thought, divergence between non-philosophers and philosophers may simply reflect the performance/competence distinction.[5] Where this is the right explanation of the diagnosis, the x-phi results are actually *less* instructive than the philosophers' armchair intuitions. Now, there are certainly other possible explanations for divergences of this sort, ones that give a more central role to the x-phi results. Thus, one alternative explanation of a given divergence might be that philosophers' intuitions lie downstream of their substantive philosophical commitments—whereas the layperson's intuitions are presumably "purer," lying upstream of theory.[6] In all likelihood, sometimes the philosopher-favoring explanation of the divergence will be the right one and sometimes the philosopher-disfavoring explanation will be. But given that divergence is anyway a relatively rare affair, if x-phi's significance was exhausted by the cases in which divergence occurs and is best explained in a philosopher-disfavoring way, it would be very limited indeed.[7]

* * *

A fourth potential rationale for experimentalization is the following. It might be suggested that even if armchair results reflect more than just the philosopher's concept of intentional action, or knowledge, it may still reflect only a comparatively small group's concept—a group of very specific socioeconomic or ethnic or cultural profile, say. Consider Machery et al.'s (2004) claim that Kripkean intuitions undergirding causal theories of reference are not shared by East Asians, or for that matter Machery et al.'s (forthcoming) claims that Gettier intuitions are *stable* across cultures. The fact that we cannot know in advance of experimental research whether such robust intuitions as were elicited by Kripke and Gettier will extend to other cultures demonstrates the value of x-phi.

My response is twofold. First, even if we accept such claims at face value—and many do not (Systma & Livengood 2011)—this would at most justify the cross-cultural (or more generally the variability) component of x-phi research; it leaves untouched more "modest" conceptual-analysis claims, such as that *Western civilization's* concept of reference involves a causal dimension and that *Western civilization's* concept of knowledge requires more than Justified True Belief. Experimentalization has not yet been shown to be needed for the assessment of such claims. More crucially, cross-cultural research of this kind involves a principled difficulty concerning the perfect match between translations. The difficulty may be summarized as follows: if concepts individuate sensitively to their conceptual role, then in all likelihood non-English words used to translate "refers" into such languages as Cantonese, where Kripke intuitions are claimed not to be sustained, do not express the same concept that "refers" does. Accordingly, the relevant cross-cultural studies do not reveal a difference in *the concept of reference*; they only reveal that other cultures find it more useful to work with a closely associated concept than with the concept of reference. Let me expand on this a little.

Anyone who has engaged in translation work knows that the operative task is to find the *closest* word in the translation language to the word in the original language; there is typically no expectation, nor hope, of finding a word whose entire web of connotations—its full conceptual role, as philosophers would put it—is strictly *identical* to that of the original word.[8] For anyone who thinks that concepts individuate sensitively to their conceptual role, this means that the translating word and the translated word do not express a perfectly *identical* concept but only the two closest concepts the translator could pinpoint. Thus, while it is unquestionable that the *best* French translation of "S knows that *p*" is "S sais que *p*," it is highly unlikely that the concepts expressed by "knows" and "sais" are identical; for example, the fact that the verb "connaitre" in French is also best translated as "to know" means that some of the conceptual work performed by "knows" is *not* likewise shouldered by "sais." The conceptual roles are thus different, which plausibly entails that the concepts themselves are different. So any study that claimed to look into French speakers' concept of knowledge would be missing out on the fact that "the concept of knowledge," that is, the concept expressed by the English word "knowledge," is probably *not* the concept expressed by the French "savoir." And if this sort of conceptual daylight opens when we move from English to French, presumably the conceptual distance when moving from English to Cantonese is even starker. If so, the aforementioned studies do not quite reveal the intuitions that subjects in other cultures

have regarding the concepts of *reference* and *knowledge*, that is, the concepts expressed by the English words "reference" and "knowledge"; what they suggest, rather, is that subjects in the relevant cultures work in semantic and epistemic contexts with slightly different concepts than we do.

It is *epistemically possible*, of course, that (i) the concept expressed by the best translation of "knows" in *some* language is strictly identical to the concept expressed by "knows" (that is, the concept of knowledge) and (ii) competent speakers of that language will consistently return different verdicts on Gettier cases. This is possible, but for any given language the probability that this is what is going on is overwhelmingly smaller than that the two concepts are ever so slightly different. Certainly this is what the canon of translation instructs us to surmise, as Davidson (1974) has already shown. Suppose I hold up a coffee mug and exclaim "Look at my new tiger—it's a very nice tiger, and more than all other tigers I've had before it keeps the coffee inside warm while ensuring my fingers don't get burned." In trying to understand me, you face two interpretive options: (a) I believe that I have a new tiger, and take the word "tiger" to express the concept of a tiger; (b) I believe that I have a new coffee mug, and take "tiger" to express the concept of a coffee mug. What is striking here is that any remotely competent interpreter would go for interpretation (b). By the same token, when less dramatic divergence from our web of beliefs occurs, it is incumbent on us to start with the presumption that the words used by the other side are not perfect translations of the words we use. This is the very meaning of the principle of charity, as a constraint on the competence of an interpreter. In assuming that the words that best translate "refers" in various languages in which Kripke intuitions do not hold up express the exact same concept "refers" expresses, Machery et al. (2004) effectively violate the principle of charity, thus failing to interpret competently their subjects.[9]

A further potential rationale for experimentalization might be that experimental studies have shown that intuitions about thought-experimental scenarios can vary with truth-independent factors (Swain, Alexander, & Weinberg, 2008), such as the font in which vignettes are displayed. This is claimed to debunk the very idea of using thought experiments in clarifying the structure of concepts. The problem with this proposed rationale is that it would not support Knobe's decision to add an experimental component to his article—it would support Knobe not writing his article at all. For what it suggests is that intuitions about presented vignettes are epistemically worthless. So, the proposed rationale is not really a rationale for experimentalizing the study of concepts via consultation of folk subjects' intuitive reactions to thought-experimental scenarios, but for

abandoning this kind of inquiry altogether. It would dismiss any research using vignettes and surveys in the style of social psychology to understand the structure of folk concepts. Yet the vast majority of x-phi research does just that. This is explicitly acknowledged by proponents of this rationale (Alexander, Mallon, & Weinberg, 2010), who have argued that the "negative program" of x-phi (exposing the truth-insensitive aspects of intuition reports) undermines the "positive program" of x-phi (empirically discovering the structure of folk concepts).[10] While this may be, the effect of the negative program would be to destroy a certain type of inquiry, not to "experimentalize" it. In this paper, however, my concern is to identify a rationale for experimentalization—a rationale, therefore, for x-phi's positive program.

Knobe's own current view is that x-phi is straight-up cognitive science that happens to be carried out by people whose salary is paid by philosophy departments (Knobe, forthcoming). Perhaps; but the significance of this notion should be fully appreciated by proponents of x-phi. On this view, experimental philosophy is as much a kind of philosophy as a rubber duck is a kind of duck. It is a historical accident that Stephen Stich was already a respected professional philosopher and stayed in Rutgers's philosophy department, when his interests started to migrate into areas that, at the level of substance, are nonphilosophical, and that he attracted many talented students within Rutgers's philosophy PhD program who went on to obtain positions in various philosophy departments around the United States and to propagate a type of nonphilosophical inquiry within the institutional structure of philosophy. On this picture, then, experimental philosophy has nothing directly to do with philosophy.[11,12] I find this line excessively dispiriting; my hunch is that there actually is something about philosophy's project that potentially calls, at some point in its pursuit, for the kind of research performed by x-phi. In the remainder of this paper, I explore a kind of contribution that x-phi could make (and make *in virtue of* its experimental dimension) to philosophy proper.

Doing so is complicated by the fact that it is far from clear what makes a given piece of inquiry "philosophy proper." Obviously, we cannot take this matter up here.[13] Instead, I will dogmatically adopt, without argument, one specific, quite traditional conception of the philosophical project. With this view of the *ends* of philosophical inquiry in place, I will develop—rather speculatively—a line of thought that secures a place for x-phi as a *means* for the pursuit of those ends. Since I will not argue for the relevant conception of philosophy, my main claim is better thought of as conditional: *if* we adopt that conception, *then* we can make sense of the role of experimental methods in philosophy proper. Warning: even

my case for this conditional will be highly speculative, involving many substantive assumptions about the structure of concepts that cannot be defended here.


## 2.  Conceptual analysis and serious metaphysics

There is a traditional way of doing philosophy that features centrally a cluster of familiar practices, including careful examination of arguments and objections; clarification and analysis of central concepts; rife appeal to thought experiments and the intuitions they elicit; marked sensitivity to parsimony considerations; a premium on armchair reasoning, and more. But what is the overarching project in the service of which these practices are deployed? This is not an easy question, and as promised, here I dogmatically adopt one particular answer to it. It is the answer articulated by Frank Jackson in *From Metaphysics to Ethics* (Jackson, 1998a) and further developed by David Chalmers in *Constructing the World* (Chalmers, 2012). Jackson calls the project he articulates "serious metaphysics"; he presents the main idea as follows:

> Metaphysics is about what there is and what it is like. But it is not concerned with any old shopping list of what there is and what it is like. Metaphysicians seek a comprehensive account of some subject-matter—the mind, the semantic, or, most ambitiously, everything—in terms of a limited number of more or less basic notions. (Jackson, 1998a: 5)

Serious metaphysics has two central features: "it is discriminatory, and it claims completeness" (1998a: 27). At the end of serious-metaphysical inquiry, if you will, we would have a *complete* theory of the world, a theory that leaves nothing unaccounted for. But despite leaving nothing unaccounted for, the theory would list only a *proper subset* of all truths. For example, while it might list "Jimmy is a boson," and perhaps also "No boson is a fermion," it will *not* list "Jimmy is not a fermion." Likewise, while it might list "Johnny is an electron" and "Electrons are particles with negative charge," it will not also list "Jimmy is a particle with negative charge." The reason is that it does not *need* to list these extra truths to ensure that nothing is left unaccounted for—that the *whole truth* is captured by it.

Let us call "fundamental" the statements our complete theory explicitly lists and "derivative" those it does not—but which are nonetheless true by its lights. And likewise let us call the notions appearing in the fundamental statements "fundamental notions" and those appearing only in derivative statements

"derivative notions."[14] Then the above characteristics of serious metaphysics may be stated as follows:

COMPLETE: Serious metaphysics attempts to produce a *total* theory of the world, that is, a theory of everything.

DISCRIMINATE: Serious metaphysics discriminates between fundamental and derivative statements (and notions).

Jackson argues for a third crucial characteristic of serious metaphysics as he views it, what he calls "entry by entailment." The idea is that any statement S not listed by a serious-metaphysical theory T qualifies as true by T's lights *only* if S is *entailed* by T's fundamental statements. In other words:

ENTAIL: The relationship between fundamental and derivative statements in the total theory of the world is that of entailment.

Serious metaphysics is fully characterized by the conjunction of COMPLETE, DISCRIMINATE, and ENTAIL.[15]

For Jackson (1998a: ch.2), conceptual analysis has a crucial role to play in serious metaphysics. Given a complete description of the world in fundamental notions, we may wish to know whether some statement involving other notions is true; but to consider this question, we need an account of the relevant other notion in terms of the fundamental ones. Within the framework of serious metaphysics, we might put this as follows. Given fundamental truths $A_1, \ldots, A_n$ and a statement S, if S involves only notions already appearing in $A_1, \ldots, A_n$, we can always use purely formal means to determine whether S is entailed by $A_1, \ldots, A_n$. But if S involves some notions foreign to $A_1, \ldots, A_n$, then in most cases there will be no purely formal tests for entailment.[16] We must appeal to a priori bridge principles that connect fundamental and derivative notions. As Chalmers (2012) shows, such bridge principles need not amount to definitions of the derivative notions in terms of the fundamental ones; all they require are analytic one-way conditionals. With such conditionals in place, we can consider whether S is entailed by (or at least consistent with) the conjunction of $A_1, \ldots, A_n$ plus the relevant analytic conditionals.

For example, suppose it is known that in 1900, 50 percent of thirty-year-old Spaniards were male, and that 20 percent of thirty-year-old male Spaniards were unmarried. Then the statement "In 1900, 10 percent of thirty-year-old Spaniards were bachelors" should be regarded as true. But this statement does not *formally* follow from the other two, since the notion of a bachelor does not appear in those two. To derive it, we must have in place the *conceptual* claim that
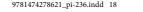
unmarried adult males are bachelors. It is the mandate of conceptual analysis to supply this extra claim. Without conceptual analysis, we cannot evaluate the truth of any non-fundamental statement that features any term not used in the fundamental statements.

The starkest illustration of this is with non-fundamental *existentials*. Presumably, the fundamental truths about the world will not mention policemen and ghosts. Yet, "There are policemen" is among the truths they will entail, whereas "There are ghosts" is not. How can we account for this pattern of entailment? The answer is that the fundamental truths guarantee the existence of policemen but not that of ghosts *given what it takes for some chunk of the world to qualify* as a policeman or a ghost. It is the job of conceptual analysis, now, to tell us what it takes for some chunk of the world to qualify as an N, for any derivative N (i.e., for any notion that does not appear in any fundamental truth). Thus insofar as metaphysics is supposed to tell us, for any putative entity mentioned only in non-fundamental truths, whether it exists or not, there is no way to decide the question without engaging in conceptual analysis. Some putative entities mentioned in non-fundamental truths should be grounded in or reduced to (combinations of) entities mentioned in fundamental truths; some should be eliminated from our picture of the world. The choice between reductivism and eliminativism about such non-fundamentals requires doing some conceptual analysis of the notions used to denote them.[17]

Philosophers are sometimes challenged by scientists to explain why they routinely conjure up outlandish thought experiments that have nothing to do with what we know empirically of the way the world is—and what they hope to learn from it. Conceptual analysis makes sense of the philosopher's appeal to thought experiments (though there may be other ways to make sense of it). The "method of cases," as it is sometimes called, can invoke thought experiments to flesh out conceivable (read: conceptually possible) scenarios and then consider whether these scenarios intuitively qualify as cases of C, in the sense that the concept of C intuitively applies to them. Traditional conceptual analysts devised thought experiments in their armchair, consulted their own intuitions about those, and shared the results with the community of inquiry through publication. Experimental conceptual analysts devise thought experiments in lab meetings (where the details of presentation of those thought experiments in digestible vignettes are fixed), consult the intuitive reports of undergraduate students about them, and share their results as traditionalists have.

From this perspective, one philosophical role of x-phi might be in the context of serious metaphysics, in particular the part of serious metaphysics that involves bridging fundamental and derivative truths. Within this framework, a strong rationale for x-phi would show why the experimentalization of conceptual analysis is (at least sometimes) inevitable, or recommended, for pursuit of serious-metaphysical q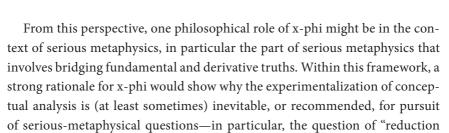uestions—in particular, the question of "reduction or elimination?" for various putative entities cited in non-fundamental truths.

In the next section, I take a first step toward showing that conceptual analysis may indeed sometimes call for experimentalization. As promised, my gambit is highly speculative. And in any case, it would provide only *one* possible rationale for x-phi—which may not be the *only* one. It is also a rationale that would work only for someone with sympathies for Jackson's project of serious metaphysics, which is not for everybody. My own belief is that the aforementioned cluster of philosophical practices cannot be made sense of unless one conceives of philosophy as centrally involving something like what Jackson calls serious metaphysics; conceptions of philosophy that do away with such a project will find that they cannot make sense of why philosophers' workaday looks the way it does. Obviously, I cannot argue for this here. Instead, I adopt Jackson's framework dogmatically, and try to show that it makes room for (occasional) experimentalization of the kind of inquiry revolving around thought experiments and intuitions.

## 3. Twenty-first-century conceptual analysis

The original technique of conceptual analysis consisted in producing severally necessary and conjointly sufficient conditions for an item falling under a concept. A paradigm of this is the analysis of knowledge as (i) justified (ii) true (iii) belief. For a variety of reasons, latter-day conceptual analysis often relies on a more powerful technique: Ramsification. This technique is more powerful in that it subsumes necessary-and-sufficient-conditions as a special case, as we will see below.

### 3.1 Straightforward Ramsey sentences

A simple Ramsey sentence for a concept C is produced by collecting platitudes about items that fall under C, making a long conjunction out of these platitudes,

replacing occurrences of "C" and cognates with a variable, and prefacing the conjunction with an existential quantifier. Consider for example our concept of the sun. Here are a dozen platitudes of relevance:

- The sun is extremely hot
- The sun looks yellow
- The sun rises in the east
- The sun sets in the west
- The sun shines longer in summer than in winter
- Basking in the sun makes people happy
- The sun is more or less spherical
- The sun is very far from Earth
- The Earth rotates around the sun
- Most or all life on Earth is impossible without the sun
- The sun is much bigger than the earth
- Too much exposure to the sun can cause cancer

These statements about the sun are "platitudes" in the sense that they are widely shared among non-experts. (Note that this does not require they be based on perception or "common sense," though often they are.) We construct a simple Ramsey sentence out of this dozen as follows:

> (Sun₁) There is a (single) $x$, such that $x$ is extremely hot, $x$ looks yellow, $x$ rises in the east, $x$ sets in the west, $x$ shines longer in summer than in winter, basking in $x$ makes people happy, $x$ is more or less spherical, $x$ is very far from Earth, the Earth rotates around $x$, most or all life on Earth is impossible without $x$, $x$ is much bigger than the earth, and too much exposure to $x$ can cause cancer.

The full Ramsey sentence for the concept Sun would involve many more platitudes than this, but the point is clear: Sun is the concept that picks out that worldly item that satisfies this kind of Ramsey sentence. It is this concept that is expressed by the public-language name "the sun."[18]

Conceptual analysis by Ramsification made an early appearance in the philosophy of mind, where Lewis (1966) used it to propose a functionalist analysis of PAIN in terms of pain platitudes. This sort of "analytic functionalism" was later contrasted with "psycho-functionalism" (Block 1980), which involved Ramsifying over scientific statements enjoying wide consensus among experts but not typically shared by the wider populace. Consider the following dozen statements about the sun:

- The sun's energy is produced by fusion reactions at the sun's core
- The temperature at the sun's core is at least 10,000,000°C
- The sun's core is mostly made of liquid hydrogen
- The sun's hydrogen has been fueling it for about 5 billion years
- Energy from the sun's core is transferred to the sun's surface, the photosphere, through radiation and convection
- The sun's photosphere is the part of the sun visible from Earth
- The sun's photosphere is about 5,000 km deep
- The sun's diameter is ~1,400,000 km
- The temperature on the sun's photosphere is ~5,500°C
- The sun's photosphere is mostly gaseous
- The temperature on the chromosphere, the sun's lower atmosphere, is ~60,000°C
- Temperatures on the corona, the sun's upper atmosphere, are often ~1,000,000°C

Collecting these and others like them, one may produce the following "scientific" Ramsey sentence:

(Sun$_2$) There is a (single) $x$, such that $x$'s energy is produced by fusion reactions at $x$'s core, the temperature at $x$'s core is at least 10,000,000°C, …

Late in his career, Lewis (1994) maintained that a proper Ramsey sentence should appeal to both platitudes and scientific consensus. A proper Ramsey sentence for Sun, for example, would use both dozens above.

This sort of two-band Ramsey sentence is very natural in analyzing concepts. In a way, the *Merriam-Webster* definition of the sun does just that:

The luminous celestial body around which the earth and other planets revolve, from which they receive heat and light, which is composed mainly of hydrogen and helium, and which has a mean distance from earth of about 93,000,000 miles (150,000,000 kilometers), a linear diameter of 864,000 miles (1,390,000 kilometers), and a mass 332,000 times greater than earth.

Here we are offered seven central statements about the sun, three broadly platitudinous and four broadly scientific.[19]

Although a two-band Ramsey sentence of this sort can become quite complicated, the basic structure of a Ramsey sentence can also remain rather simple. Thus, if we think that BACHELOR simply picks out the property of being both (i) unmarried and (ii) a man, we can formulate an exceptionally short Ramsey sentence with just two platitudes: something like "There is an (attribute of being

an) $x$, such that every $x$ is unmarried and every $x$ is a man." This would be a Ramsified version of the older necessary-and-sufficient-conditions analysis of BACHELOR. It is in this sense that Ramsification is more *powerful* a conceptual-analysis technique than definition by necessary and sufficient conditions: concepts that admit of traditional analysis by necessary and sufficient conditions always admit also of analysis by Ramsification, but some concepts that do not admit of traditional necessary-and-sufficient-conditions analysis may yet succumb to analysis by Ramsification. The hope is that the Ramsey technique is powerful enough to accurately capture the structure of *all* our folk concepts.

The philosophical program of conceptual analysis by Ramsification is often referred to as the Canberra Plan (Price & O'Leary-Hawthorne, 1996), due to its dominance at the Australian National University in Canberra. The approach raises a number of immediate difficulties, however. What happens if more than one entity satisfies a simple Ramsey sentence (think of jadeite and nephrite, both of which satisfy the platitudes about jade)? What if no entity satisfies absolutely all of them, but one does satisfy most (think of DOLPHIN before Linnaeus, whose Ramsey sentence presumably included the platitude "The dolphin is a fish")? How can different individuals share concepts if they associate different platitudes with them? How can concepts co-refer before and after the establishment of scientific consensus regarding central statements? These and other questions make clear that a more sophisticated brand of Ramsey sentence is needed.

## 3.2  Studies in advanced Ramsification

I start by mentioning four principled challenges to conceptual analysis by Ramsification. I call them the Quinean Challenge, the Kripkean Challenge, the Putnamian Challenge, and the Burgean Challenge. I then suggest technical modifications of Ramsey sentences intended to address them.

The Quinean Challenge concerns the rejection of the analytic/synthetic distinction (Quine, 1951). For Quine, when some item $a$ falls under a concept C, there is no dichotomy between truths about $a$ that are constitutive of its falling under C and truths that just happen to hold as well. Statements about things that qualify as C do not divide neatly into those that are nonnegotiable for something's status as a C and those that are negotiable. On the contrary, every conviction about the Cs is in principle revisable. Thus, for each platitude P about the sun, something can qualify as the sun even if it fails to exhibit the property P attributes to the sun. If this is right, then sentences SUN$_1$ and SUN$_2$ cannot be used to model our concept of the sun. For the falsehood of any platitude showing up in

SUN₁ OR SUN₂ would render the sentence false, that is, would mean that there is no thing that satisfies SUN₁ or SUN₂. Intuitively, however, the sun exists! Furthermore, in sentences such as SUN₁ or SUN₂ all platitudes have equal weight within the Ramsey sentence they make up—which conflicts with the more flexible Quinean picture of variable revisability.[20]

The Kripkean Challenge concerns rigid designation and direct reference (Kripke, 1972). When you and I look at the sun and converse on the weather, the reason we manage to communicate appears to have nothing to do with the platitudes we each happen to associate with the sun. It is perfectly conceivable that we should share *no* sun platitudes and still manage to communicate. Thus some concepts do not pick out their referent via a description of its properties. Instead, they pick it out *directly*. There may be properties of the object in virtue of which it is picked out by the concept, but these need not be properties the concept possessor herself associates with it. Accordingly, such concepts pick out the same thing(s) in (or rather *at*) every possible world, regardless of changes in the referent's properties across worlds. That is, these concepts are "rigid designators." Thus, what makes it the case that the sun is the referent of our concept SUN is simply the fact that the sun is the main cause of tokenings of SUN. It is because the same object causes tokenings of my sun concept and your sun concept that we manage to communicate, even if we share no beliefs about the sun's properties. Furthermore, our sun concepts pick out the sun at all other possible worlds in which it exists, even in worlds where the sun is so different from the way it is in the actual world that most platitudes concerning it no longer hold.

Similar themes animate the Putnamian Challenge, which has to do with an alleged *hidden indexical* in many concepts (Putnam, 1975). According to Putnam, concepts such as WATER involve a hidden indexical component: they pick out the watery stuff *around here*, that is, the watery stuff *of our acquaintance*. This too is crucial to communication. As noted, you and I can successfully communicate about the sun even if the platitudes we associate with it have nothing in common. Even if you were a scientist whose tacit beliefs about the sun conformed to SUN₂ and I a layperson in the grip of SUN₁, we would not be speaking past each other. Indeed, we can associate inconsistent platitudes with the sun and yet manage to communicate. This is because we would both be simply referring to the celestial object before our eyes—for Putnam, the object our perceptual experiences refer to indexically.

The Burgean Challenge concerns the division of linguistic labor (Burge [1979], though see already Putnam [1975]). Observe that many in our linguistic (and conceptual!) community are unaware of the statements collected in SUN₂.

Yet arguably some of them are essential to our shared concept of the sun, as indeed suggested by the *Merriam-Webster* entry. It is a structural feature of many of our concepts, claims Burge, that they involve deference to experts. In using "Sun," I implicitly consent to refer not to whatever satisfies *my* core beliefs about the sun but to what also satisfies certain core beliefs of helio-buffs. The same point can be made about *concepts*. The reference of my *own* C is governed not (only) by the platitudes *I* associate with Cs, but by the deeper truths that the aficionados associate with Cs. Interestingly, according to Burge there are also many concepts that involve deference not to experts but to the community at large—as when it is built into my concept SOFA that a sofa is in part the kind of thing that others in my community call a sofa (Burge, 1986).

Not everybody accepts the philosophical claims these challenges are based upon (regarding the analytic/synthetic distinction, direct reference, hidden indexicals, and the division of linguistic labor). But it is hard to deny that *some* insight underlies each. The question is whether a sufficiently sophisticated Ramsey sentence can accommodate these insights. I think it can.

*   *   *

Start with the Quinean Challenge. It seems intuitively clear that many concepts can successfully refer even if some statements in their associated Ramsey sentences are false. For example, it may turn out that the big celestial body you and I are looking at is *not* extremely hot. Instead, there is a tiny speck, always just next to the sun, that generates all the heat we ordinarily associate with the sun. Intuitively, this illusion would not undermine our concept's reference to the big celestial body we are looking at. It would not, for example, make it the case that our sun concept really picks out the speck, or has "indeterminate reference" (whatever that means). In other words, the "extremely hot" platitude is negotiable. Likewise for the "rotates" platitude: heliocentrism may turn out to be wrong, and SUN would still refer to the sun. The problem is that the above Ramsey sentences make these platitudes nonnegotiable: their falsity ensures reference failure.

Fortunately, Lewis (1972: 256) raises and solves this problem in half a sentence. His suggestion is that instead of forming a long conjunction of *all* platitudes about the sun, we form an even longer disjunction of all conjunctions of *most* platitudes (see also Jackson, 1998a: 35). Suppose our concept of the sun works in such a way that an object need only satisfy any eight of the dozen sun platitudes in $SUN_1$ in order to qualify as the sun. Then we could proceed by first forming every conjunction of (at least) eight among these platitudes, then forming the disjunction of all those conjunctions, and then replacing occurrences of

"the sun" with a bound variable. We would then obtain a Ramsey sentence agile enough to capture the way our concept of the sun works (in this scenario).

More generally, suppose in analyzing concept C we Ramsify over statements that ascribe properties $P_1, \ldots, P_{12}$, and any $x$ qualifies as a C just if it satisfies eight of these statements. Then our Ramsey sentence for C would take the form:

(Lws) There is a (single) $x$, such that $x$ is $P_1$ & $x$ $P_2$ & … $x$ is $P_8$, or $P_2$ & $x$ $P_3$ & … $x$ is $P_9$, or $P_1$ & $x$ $P_2$ & … $x$ is $P_7$ & $P_9$, or…

Call this a *Lewis sentence*. A Lewis sentence is a *species* of a Ramsey sentence, but one that can accommodate Quine's claims, including that negotiability comes in degrees. Within a Ramsification framework, this means that some platitudes are more central and some are more marginal. For example, "The sun is extremely hot" seems more central to our concept of the sun than "The sun is more or less spherical." Presumably, if a relatively marginal platitude turns out to be false, this should not affect the concept's tendency to refer successfully as much as if a central platitude does.[21] This can be reflected in the following Lewis-sentence feature. In lws above, each platitude appears in equally many disjuncts, and most disjuncts have the same number of conjuncts in it (eight).[22] But we can imagine a Lewis sentence in which some platitudes appear in more disjuncts than others. This would reflect a greater relative nonnegotiability, that is, a lesser relative revisability. Thus, we can model the fact that "The sun is extremely hot" is more central to our concept of the sun than "The sun is more or less spherical" by making the former show up in more disjuncts than the latter in our Lewis sentence for SUN.

Note that a statement that appeared in *every* disjunct would be *completely* nonnegotiable (irrevisable), in the sense that its turning out false would guarantee that the concept fails to refer. As noted, Quine's view was that *nothing* is entirely nonnegotiable in our concept of the sun. What this means is that there is no property ascribed by our concept of the sun such that if nothing instantiates it, then the concept is empty. This could be captured by ensuring that no statement appearing in the Lewis sentence for SUN showed up in every disjunct of the sentence. Of course, Quine may be simply wrong about this, perhaps not when it comes to SUN but with respect to some other concept(s). In fact, I would be surprised if none of our concepts involve a nonnegotiable descriptive component. But the Lewis sentence as such can be used to model both types of concept—those that involve nonnegotiable statements and those that do not—and the difference between them can in fact be captured by the formal properties of Lewis sentences.

It is possible to hold that while no individual platitude associated with a concept is nonnegotiable by itself, certain combinations of platitudes are *conjointly* nonnegotiable. For example, suppose we thought that the first four platitudes in sun₁ are such that if they *all* turned out to be false, then the concept would be empty after all. Then although no single platitude should appear as a conjunct in every disjunct of our Lewis sentence for the sun, the *disjunction* of these four platitudes should appear as a conjunct in every disjunct. The point is that Ramsey sentences can be agile enough to capture various potentially quite subtle aspects of the psychological reality of concepts.

(We might also think that some concepts are such that a certain platitude associated with them is so *central* to them that its truth guarantees that the concept refers. This could be captured by a Lewis sentence one of whose disjuncts featured a single statement.)

In conclusion, we may or may not be impressed by Quine's attack on the analytic/synthetic distinction, and accordingly may or may not *wish* to accommodate Quine's relevant ideas in our Ramsey sentences. What I have tried to show here is just that Ramsey sentences *have the resources* to accommodate the Quinean ideas. In other words, rejection of the analytic/synthetic distinction would not by itself undermine the project of conceptual analysis by Ramsification.

<div align="center">* * *</div>

What about the other challenges to conceptual analysis? Kripke, Putnam, and Burge have developed accounts that challenge the *analyzability* of concepts, by arguing that certain concepts' behavior is not fixed by their descriptive content (and/or that certain concepts do not *have* descriptive content). These raise momentous issues in the philosophy of language that cannot be settled here. What I want to do here is relatively modest: remind that *the phenomena* Kripke, Putnam, and Burge appeal to in the context of motivating their accounts of concepts *can* be accommodated within a rich conception of descriptive content, in a way that lends them to analysis by Ramsey sentence. The underlying philosophy-of-language question of whether these phenomena are *best* accounted for as (a) built into descriptive content, (b) involving non-descriptive content, or (c) somehow content-transcendent is something we cannot address here with any seriousness.

Starting with the Burgean challenge, it may well be that many folk concepts have built into them an element of deference to experts, or to others in one's community. But if so, a full Ramsey sentence could simply incorporate *deferential platitudes* such as:

- The sun has most of the properties the experts' consensus says it does
- The sun is the object referred to by other community members' use of "the sun"

Accordingly, Sun$_1$ for example could be reformulated as follows:

(Sun$_3$) There is a (single) $x$, such that $x$ is extremely hot, $x$ looks yellow, $x$ rises in the east, … $x$ is much bigger than the earth, too much exposure to $x$ can cause cancer *$x$ has most of the properties the experts' consensus days it does*, and *$x$ is the object referred to by community-members' use of "the sun."*

These kinds of sentence seem to accommodate the Burgean insight. Presumably, if some concepts pick out their referents in virtue of experts' beliefs, it is because that is how these concepts work—they "specify," so to speak, that this is how we are to home in on the right referent or extension. Thus unless I tacitly assented to the deferential platitudes when conversing about the weather with astronomers, we might very well be talking past each other.[23]

A similar approach has been applied to the Kripkean Challenge, by various defenders of descriptivist theories of meaning. Nothing in Kripke's arguments rules out, they have claimed, a sort of "causal descriptivism" or "rigid descriptivism" that would accommodate his insights within a descriptivist framework (see Kroon, 1987; Lewis, 1994). For example, we can accommodate Kripke's causal claim by adding to our Ramsey sentence for sun causal platitudes, such as:

- The sun is the main cause of Sun tokenings

As for the rigidity of some concepts, it can be accommodated simply by rigidifying their corresponding Ramsey sentences (Brody, 1977; Jackson, 1998b). Incorporating both rigidification and causal platitudes, we turn Sun$_3$ into something like this:

(Sun$_4$) There is a (single) $x$, such that, *in the actual world*, $x$ is extremely hot, $x$ looks yellow, $x$ rises in the east, … too much exposure to $x$ can cause cancer, $x$ has most of the properties the experts' consensus days it does, x is the object referred to by community-members' use of "the sun," and *$x$ is the main cause of tokenings of Sun*.

Admittedly, these moves by descriptivists have met with various further challenges (see Brock, 2004 for a particularly strong critique). As noted, we will not address here the underlying philosophy-of-language question of what the best theory of concepts is. The point is only that Ramsey sentences *can* readily

incorporate the kinds of information that direct-reference theorists have thought essential to concepts.

Similar remarks apply to Putnam's Challenge. If communication can be established with as little as joint perceptual awareness of the same object, as Putnam claims, then we should also add perceptual platitudes such as:

- The sun is that of which I seem to be perceptually aware when I have my sun experiences
- The sun is that of which others seem to be aware when they have perceptual experiences subjectively like my sun experiences

Sun$_4$ can be strengthened with platitudes of this sort to accommodate the insight underlying the Putnam Challenge.
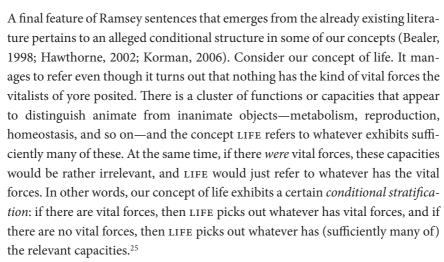
A word on the commitments incurred by the moves just sketched would not be out of place. On the one hand, I do not wish to commit to the truth of Kripke's, Putnam's, and Burge's underlying claims about concepts' content. Perhaps (some or all) concepts involve no deference, no hidden indexical, no causal, and/or no rigidified content. I remain officially silent on this issue. My point is only that Ramsey sentences can accommodate the kinds of phenomenon Kripke, Putnam, and Burge appeal to in making the case for those claims.[24] Conversely, however, I do not wish to commit to a descriptivist theory of concepts' reference, which insists that our concepts' behavior is fixed *entirely* by their descriptive content. Rather, assuming that a concept's descriptive content is *what can be analyzed* in that concept, I want to highlight just how far the analysis of a concept can go. (More on this in section 4.2.)

It might be objected that it is psychologically unrealistic to expect the folk concept of the sun to incorporate such subtle deferential, causal, and perceptual platitudes. For example, small children with whom we routinely discuss the sun are unlikely to associate Kripke's causal platitude with their concept of the sun. In response, I would personally want to agree that the perceptual and community-deferential platitudes are much more central to our concept of the sun than any causal and expert-deferential ones. I take it that this is what is shown by the fact that we can communicate about the sun with children who are unlikely to incorporate the latter platitudes into their conception of the sun. The current point, to repeat, is that Ramsey sentence *can* capture the psychological reality of any concepts that do exhibit the kind of features suggested in the Kripkean, Putnamian, and Burgean challenges; it is a separate question which concepts exhibit which of these features.

\* \* \*

A final feature of Ramsey sentences that emerges from the already existing literature pertains to an alleged conditional structure in some of our concepts (Bealer, 1998; Hawthorne, 2002; Korman, 2006). Consider our concept of life. It manages to refer even though it turns out that nothing has the kind of vital forces the vitalists of yore posited. There is a cluster of functions or capacities that appear to distinguish animate from inanimate objects—metabolism, reproduction, homeostasis, and so on—and the concept LIFE refers to whatever exhibits sufficiently many of these. At the same time, if there *were* vital forces, these capacities would be rather irrelevant, and LIFE would just refer to whatever has the vital forces. In other words, our concept of life exhibits a certain *conditional stratification*: if there are vital forces, then LIFE picks out whatever has vital forces, and if there are no vital forces, then LIFE picks out whatever has (sufficiently many of) the relevant capacities.[25]

Within the Ramsification framework, this conditional structure can be captured as follows. In the first place, we formulate a Lewis sentence $L_1$ on the assumption that there *are* vital forces. Next, we formulate a Lewis sentence $L_2$ on the assumption that there are no vital forces. Finally, we construct a "mega-Lewis sentence" for LIFE, which specifies that for any $x$, if $x$ satisfies $L_1$, then $x$ is alive, and alive in virtue of satisfying $L_1$, and if $x$ does not satisfy $L_1$ but does satisfy $L_2$, then $x$ is alive, and alive in virtue of satisfying $L_2$.

Note that every Lewis sentence can be turned into a mega-Lewis sentence by introducing priority relations among its disjuncts—by imposing, that is, a conditional stratification. This device has the potential to handle many cases where more than one entity satisfies the Lewis sentence. If in the Lewis sentence for some concept C, entity $E_1$ satisfies disjunct $D_1$ but $E_2$ satisfies $D_2$, it does not yet follow that C has "divided reference." It may be that C works in such a way that if anything satisfies $D_1$, than it is the unique referent of C, but if nothing does, then whatever satisfied $D_2$ is. (This can be reiterated indefinitely, so a whole hierarchy of potential referents is formed.) If so, we can turn C's Lewis sentence into a mega-Lewis sentence by "conditionally stratifying" it. At the same time, some concepts really do have divided reference, and those would be better modeled by a Lewis rather than a mega-Lewis sentence.[26]

The considerable flexibility of (mega-)Lewis sentences means that it is relatively hard for them to come out false—harder than it is for a simple Ramsey sentence, certainly. This will handle many cases in which folk concepts refer successfully even though the folk associate many false platitudes with them. For example, it explains why the pre-Linnaeus folk concept of a dolphin referred even though the folk associated with it the platitude "The dolphin is a fish." At

the same time, some concepts really are empty, and when they are, their corresponding (mega-)Lewis sentences should come out false. For example, whatever else is true of the Lewis sentence for GHOST, it should involve no disjunct which something in the world satisfies.

In conclusion, many of our concepts may be much more complex and flexible than traditional conceptual analysts have assumed, but this in itself does not show that concepts cannot be analyzed through Ramsey sentences. It is just that our Ramsey sentences need to be agile enough to capture the flexible psychological structure of our concepts. They acquire their agility through at least six devices: (1) adding deferential platitudes, (2) adding causal platitudes, (3) adding perceptual platitudes, (4) rigidifying, (5) forsaking simple conjunctions for disjunctions-of-conjunctions, and (6) conditional stratification.[27] In all probability, there are in fact other logical properties of Ramsey sentences that capture other psychological properties of the structure of concepts.
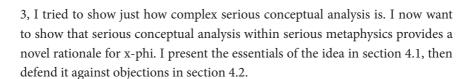
It might be objected that the envisaged Ramsey sentences are so complex that they could not possibly capture something psychologically real in us. Surely we are not aware of such immense subtlety in our concepts as is suggested by the above picture. However, this complaint seems to suppose that we have considerable first-person awareness of the structure of our concepts. But *if* this were the case, there would be no need for *analysis* of concepts. To analyze something is to identify its components and their interrelations. This kind of activity is only needed where the components of an item and their interrelations are not already known. From this perspective, discovering the structure of our concepts—analyzing them—is akin to discovering the grammar of one's native tongue. One can master that grammar perfectly without having any explicit knowledge of the rules of that grammar. It is the grammarian's job to *make explicit* the rules implicit in our grammatical knowledge-how, and it is similarly the conceptual analyst's job to *make explicit* the structure our concepts already implicitly have.

## 4.  Experimental philosophy as labor-intensive conceptual analysis

In section 2, I laid out Jackson's program of serious metaphysics, clarifying the role it assigns to conceptual analysis in linking fundamental and derivative truths and addressing the "reduction or elimination" question about the putative referents of notions that do not appear in the fundamental truths. In section

3, I tried to show just how complex serious conceptual analysis is. I now want to show that serious conceptual analysis within serious metaphysics provides a novel rationale for x-phi. I present the essentials of the idea in section 4.1, then defend it against objections in section 4.2.

## 4.1 The labor-intensity rationale for experimentalization

A rigidified mega-Lewis sentence with deferential, causal, and perceptual platitudes, as well as a great variety of other platitudes arranged in an extraordinarily complex logical structure—that is a far cry from "bachelor $=_{def}$ unmarried man." The precise structure of such a mega-Lewis sentence is not readily uncovered from the armchair. While the broad contours of a concept such as SUN, WHEEL, or GAME may be detected from the armchair by a sufficiently inventive and industrious philosopher, obtaining the exhaustive list of relevant platitudes, arranging those into the correct disjunction of conjunctions, and figuring out whether a conditional stratification and/or a rigidification are needed is an unmanageable task for a solitary armchair-bound conceptual analyst.

The great majority of concepts, I suspect, are such as to require extraordinarily complex Ramsey sentences to capture. Producing the kind of Ramsey sentence that would reflect correctly their structure is a sort of *labor-intensive conceptual analysis*. My suggestion is that the rationale for experimentalization is the need for this kind of labor-intensive conceptual analysis. Given the complexity of most of our concepts, capturing their psychological structure in full detail would require a multitude of teams of researchers working in parallel to (a) produce hypotheses about aspects of Ramsey sentences, (b) devise the thought experiments that could test those hypotheses, and (c) implement the tests through the familiar social-psychology-style questionnaires presented to the right kinds of subject. (It is an open question who the *right* subject is, given the performance/competence distinction harped on in section 1. To my mind, at least, it is doubtful that the average, relatively disengaged undergraduate student is it. More on this later.)

Producing hypotheses about aspects of Ramsey sentences that fully capture a concept's psychological structure—step (a) of the procedure—is already labor-intensive in the present framework. For every concept, the following decision points arise: (i) does the Ramsey sentence involve causal platitudes?; (ii) does it involve perceptual platitudes?; (iii) does it involve deferential platitudes?; (iv) is the Ramsey sentence rigidified?; (v) is it conditionally stratified? Each of these represents a potential confound that a different series of thought experiments

would have to address. But perhaps the hardest part of the task is (vi) to produce the exhaustive list of platitudes for a concept to begin with and (vii) to structure it into the right disjunction-of-conjunctions.

Take the aforementioned dozen platitudes regarding the sun, plus a half dozen perceptual, causal, and deferential platitudes. Ideally, to fully investigate what it takes for something to qualify as the sun, we would want to know for *each possible combination* of these eighteen platitudes whether satisfying *it* is sufficient for qualifying as the sun. For each combination, then, we would want to create a(n accessible) vignette that puts in concrete form a scenario in which just that combination of platitudes is satisfied. Note now two striking facts: (1) given 18 platitudes, the number of possible combinations of them is 262,143; (2) the real number of platitudes associated with the sun is surely much greater than 18, and probably goes well into the hundreds. And we have not yet addressed the issues of rigidification and conditional stratification! Thus generating the total questionnaire and presenting it just to oneself would already be an excessive amount of work, and in any case the resulting task would be too hard for a single subject (concentration will break down, exhaustion will set it, etc.). Different groups of subjects would probably have to be tested for different fragments of some approximation of the real Ramsey sentence for SUN, with the results synthesized by the theorist.

The point is that figuring out the actual, correct, *full* Ramsey sentence for SUN would probably swallow the entire career of a solitary armchair-bound conceptual analyst. What would be needed is rather a considerable *team* (with considerable resources) working together to produce questionnaires with some efficiency, then "delegating" the production of intuitions on concrete cases presented in the questionnaires (intuition-harvesting, we may call this) to experimental subjects. Efficiency in producing a questionnaire would be a matter of a priori elimination of certain combinations of platitudes not worth considering, having hunches (at "pre-trial," so to speak) on which platitudes are most central and whose interrelations are most in need of examination, and so on. Efficiency in the administration of the questionnaire would be a matter of finding the right groups of subjects, finding low-cost access to such subjects (as in SurveyMonkey), and so on. These are hallmarks of experimental research.

*　＊　＊*

On the line of thought I am sketching out, we did not in fact need the experimental part of Knobe's paper on the concept INTENTIONAL ACTION. The Knobe Effect is detectable from the armchair, as can be seen from the fact that most readers of Knobe's paper have had the majoritarian intuitions about Knobe's

vignettes. In contrast, which disjuncts of the Ramsey sentence for sun feature "The sun rises in the East" as a conjunct is something that would be tremendously hard to detect from the armchair. The point is that while there may be concepts for which armchair-bound conceptual analysis can produce the full Ramsey sentence, most concepts are unlike that and require more labor-intensive research.

At the same time, even for those a (potentially sometimes substantial) fragment of the Ramsey sentences may be obtainable from the armchair. Indeed, some core features of a Ramsey sentence—its least negotiable platitudes, its shortest disjuncts, whether it is rigidified—are arguably reproducible from the armchair. It is the finer details of the Ramsey sentence—its most marginal platitudes, its lengthier disjuncts, the nuances of its logical structure—that are most likely to be elusive from the armchair.[28] This may invite the question: why should we bother with experimental conceptual analysis at all?[29] We can see why by appreciating the likely relevance of experimental conceptual analysis for Jackson's serious metaphysics, in particular for the choice between reduction and elimination for various putative non-fundamental entities. On the reasonable assumption that suns are not fundamental entities, are there nonetheless suns in our world as described in terms of its fundamental entities? That is, are sun existentials entailed by the fundamental truths? We cannot decide this without a clear grasp of what is involved in there being a sun. Even if we have a full theory of what exists expressed in fundamental notions, we still need to know whether some of what exists qualifies as a sun. And for this we need an analysis of the concept of a sun. Now, it is perfectly possible that in many cases the question of reduction or elimination can be answered by considering a core fragment of a concept's full Ramsey sentence, one potentially obtainable from the armchair. But presumably, for many concepts answering the question of reduction or elimination will require the full Ramsey sentence, or at least a gigantic fragment of it—one unlikely to be obtainable from the armchair.

For example, the fact that we are rather confident, without leaving the armchair, that some of what there is qualifies as a sun, suggests that the armchair-obtainable fragment of sun suffices to come down on reduction. The fact that we are likewise confident that none of what there is qualifies as miasma suggests that the armchair-obtainable fragment of miasma suffices to come down on elimination. But some cases appear to elude resolution from the armchair. Given how the world is at the fundamental level, is there such a thing as free will? Here reduction and elimination are both live options, and one antecedently plausible diagnosis—not the only one, by any means, but quite a promising one all the

same—is that the concept of freedom (free action as well as free decision) is so complex that nothing we have uncovered so far suffices to resolve the reduction-or-elimination question.[30]

My suggestion, then, is that the need for experimentalization arises as follows. First of all, certain questions of serious metaphysics require conceptual analysis to answer: it is impossible to establish whether certain non-fundamental statements are true, notably existentials invoking notions absent in all fundamental truths, without conceptual analysis of the relevant notions. However, the psychological reality of our concepts is a lot messier than early conceptual analysts had realized. In particular, concepts' structure typically cannot be captured in a simple definition with a small number of severally necessary and conjointly sufficient conditions. Rather, an extraordinarily complex Ramsey sentence is required, requiring labor-intensive conceptual analysis to formulate. Crucially, the complexity of the Ramsey sentence often exceeds what can be reproduced from the armchair. Sometimes an armchair-reproducible fragment of the Ramsey sentence may suffice for the reduction/elimination adjudication, but plausibly, sometimes experimentalization becomes inevitable for settling the question of reduction or elimination.

## 4.2  Objections and replies

It may be objected that the labor-intensity rationale for experimentalization does not overcome the problem that arises around the performance/competence distinction. That problem attended other proposed rationales discussed in section 1, but labor-intensive conceptual analysis would face the same problem.

In response, I agree that labor-intensity by itself does not address the performance/competence gap. That issue must be addressed independently. One way to address it is to designate philosophers as the competent intuiters and run one's surveys with large groups of philosopher-subjects. This would already be much more efficient than writing a whole paper about some thought experiments, submitting it to journals, getting it published, and then seeing whether the philosophical community at large tends to converge on the expected intuition(s). Another way to address the performance/competence gap is through appeal to non-philosophers who are not entirely naïve either. Early introspectionists such as Wundt and Titchner required subjects to undergo a set number of trials before being considered competent introspectors; a similar device, perhaps more nuanced, could be used to "train up" competent intuiters, that is, subjects enjoying special facility with thought experiments, outlandish scenarios,

counterfactual reasoning about them, and so on. Perhaps simply raising the stakes for undergraduates, so that they are properly engaged with the question, say by promising them (misleadingly) a non-trivial payment if "they get the right answer," would enhance the quality of results. Coffee, relaxing atmosphere, the right time of day, a light intellectual warm-up before the real questionnaire—all these may enhance quality and do something to address the performance/competence gap. What matters for our purposes is that these ways of addressing the gap are *compatible* with the labor-intensity rationale for experimentalization. In contrast, the rationale that appealed to a potential discrepancy between philosophers' and folk intuitions to justify experimentalization, for example, seems only to aggravate the challenge presented by the performance/competence gap.

Another objection might be that many of our concepts are simply incoherent—or just include conflicting "streaks"—so that a Ramsey sentence that described them fully would have to be incoherent as well. In this circumstance, it is more natural to think of philosophy's job as not just to expose the structure of our concepts but also to "reform" or "stabilize" them—essentially, to replace them with similar but coherent concepts. If so, exposing incoherencies and internal tensions in our everyday concepts is also part of experimental philosophy's contribution.

My response is twofold. First, I am very skeptical of claims that folk concepts are often incoherent. There is a sense in which they may have different streaks in tension, but these more likely work together within a complex structure. For example, certain thought experiments may bring out a streak in our concept of life that privileges reference to vital forces, others reference to a cluster of capacities. This may superficially seem like an incoherence, but upon careful examination may turn out to fit within a perfectly coherent but conditionally stratified concept. In general, it may well be common that among the platitudes associated with a concept C, we might find both $p$ and $\sim p$; but presumably these will be insulated from each other within the overall logical structure of the Lewis sentence for C ($p$ will be a conjunct in some disjuncts of the Lewis sentence, while $\sim p$ will be a conjunct in other disjuncts). It is very hard to make sense of the idea of an incoherent yet serviceable concept. More plausibly, then, apparently conflicting streaks in a concept cohabit consistently within the concept's complex logical structure. To that extent, demonstrating the incoherence or instability of a concept requires much more than just bringing out apparently conflicting streaks; it also requires falsifying or disconfirming hypotheses about the concept's structure that would reconcile the apparently conflicting streaks in a complex (mega-)Lewis sentence.[31]

Secondly, although some x-phi research clearly attempts to demonstrate the incoherence or instability of folk concepts, this research is better thought of as part of x-phi's *negative* program. As noted in section 1, however, my goal in this paper is only to identify a rationale for x-phi's positive program. I readily confess that nothing about the project of serious metaphysics, and the role it assigns to conceptual analysis, rationalizes experimental research within x-phi's *negative* program.

A third objection is that while *some* concepts have the kind of extraordinarily complex structure that can only be fully tackled with experimentalist efficiency, this is not the case for the kinds of concept philosophers tend to be interested in. The concepts KNOWLEDGE, REFERENCE, or PROPOSITION, for example, are not going to require a rigidified conditionally stratified enormous disjunction of enormous conjunctions to capture.

However, it is actually very possible that a full and accurate capture of KNOWLEDGE would require quite a complicated Ramsey sentence. One possible diagnosis of the post-Gettier literature on the analysis of knowledge might be that it has been hampered by excessive allegiance to the hope that a compact number of severally necessary and conjointly sufficient conditions could fully capture the concept of knowledge. We just need to be more creative: add a fourth condition to JBT (Lehrer & Paxson, 1969), replace J with a reliability condition (Goldman, 1967), or some such. In a recent paper, Sosa (2009) appears to offer an eight-condition analysis of "S knows full well that *p*."[32] I suspect this more expansive account still grossly underestimates the internal complexity of KNOWLEDGE. One can only speculate here, but it may well be that the atmosphere of general dissatisfaction with the deliverances of the literature on the analysis of knowledge will not dissipate until a more labor-intensive approach to the concept is deliberately taken. It could *also* turn out, of course, that despite initial appearances, KNOWLEDGE succumbs to relatively straightforward analysis. Arguably, this turned out to be the case, contra Wittgenstein, with the concept GAME, which I take to have already been analyzed satisfactorily in terms of four independently necessary and conjointly sufficient conditions (Suits, 1978). In any case, as long as we do not actually possess a satisfactory analysis of some concept philosophers are interested in, as appears to be the case with KNOWLEDGE, it remains an open possibility that the reason is that we have underestimated the concept's complexity; that would already justify pursuing experimental conceptual analysis alongside armchair conceptual analysis. In any case, I fully recognize that the role I am assigning here to x-phi is relatively limited. To that extent, my position is intermediate between

the revolutionary zeal of some x-phi proponents and the dismissive disdain of some x-phi opponents.

Let me close with final comments on the potential charge, already discussed in section 3, that the various devices therein used to accommodate Kripkean, Putnamian, and Burgean insights fail to get us beyond the *descriptive content* of concepts—whereas the whole point of Kripke, Putnam, and Burge was that there is more to a concept's semantic profile than its descriptive content. Clearly, this is not the place to litigate issues of descriptivism versus direct reference, semantic internalism versus externalism, and so on. As noted, however, even if there is more to a concept than its descriptive content, it is the concept's descriptive component that lends itself to informative analysis. Accordingly, it is in virtue of its descriptive content that the concept plays a role in serious metaphysics. Importantly, if there are concepts that *have* no descriptive content, and simply refer to whatever causes their tokenings (say), then arguably the question of reduction or elimination *does not arise* for their putative referents. Since *something* certainly causes tokenings of C, it is trivial that C has a referent, and therefore that reductivism rather than eliminativism is true of C's referent.[33] Indeed, it is this reasoning that convinced Stich to forsake his original eliminativism about belief and desire (see Jackson, 2009; Stich, 1979). Now, if *all* concepts were of this sort, then conceptual analysis would indeed disappear from serious metaphysics, and reductivism would be the right view about *all* non-fundamentals. But we already know that this cannot be: some concepts *are* empty (PHLOGISTON, MIASMA, GHOST, ESCHER TRIANGLE, etc.). So we know that some concepts refer at least partly via a descriptive content, which unfortunately is not always satisfied by something.[34] For such concepts' putative referents, the question of reduction or elimination arises, and conceptual analysis—potentially with an experimental component—is inevitable for deciding the question.[35]

## Conclusion: A place under the sun

A final objection may be that my proposed rationale for experimental philosophy places the latter squarely within a very traditional, conservative conception of philosophy. This might not suit the self-conception of x-phi practitioners, who tend to paint experimental philosophy in more revolutionary colors. Some of them may well consider Jackson's project downright reactionary.

In response, I can only confess that my proposed rationale for x-phi embeds it in a very traditional conception of the philosophical project. This is by design.

For reasons I cannot go into here (but see Kriegel, 2016), I consider the project Jackson articulates as the core project of philosophy. Seeking a total theory of the world in terms of a privileged subset of fundamental notions, and then finding the place (if any) of putative non-fundamentals in it—that is the project we find from the Eleatics and Aristotle through Aquinas, Spinoza, and Brentano to Lewis, Sider, and Schaffer. Aristotle designated Thales as the first philosopher—Thales, who instructed that everything was water! What makes this a piece of philosophy, however disappointing? The answer is that Thales's thesis "is discriminatory, and it claims completeness." More specifically, it cites only water truths in its theory of everything, but claims that nothing in the world is left unaccounted for by these water truths. The central thread of the history of philosophy, as I see it, is the search for improvements on Thales's theory—theories both complete and discriminatory that are just *better* than Thales's. Obviously, I have not argued for this here. But from my perspective, the best way to ratify the status of x-phi as genuine philosophy (as opposed to cognitive science performed by academics whose tenure home is a philosophy department) is to identify a role for x-phi within this traditional project originating with Thales. And indeed, I have claimed that x-phi does have a role within this core project of philosophy, albeit a somewhat limited one: x-phi provides supplementation and intensification of methods for pursuing a component of this project (the component pertaining to analysis of concepts), which intensification may sometimes be needed to settle the status of non-fundamentals in the theory.

On this view, the methodological innovations of x-phi—at least as far as its positive program is concerned—pertain not to intellectual *ends*, but to the *means* for pursuing already entrenched ends. In this respect, armchair conceptual analysis and x-phi are far from rivals; on the contrary, they complement each other in pursuit of a common goal (Jackson, 2008, 2011).[36]

## Notes

1  This is certainly more plausible than the hypothesis that their concept of intentional action is different from yours and mine and *lacks* the normative aspect Knobe claims yours and mine have.

2  A concept's extension is the set of worldly items that fall under it; its intension, very roughly and to a very first approximation, is the "general rule" for its application. The concept BACHELOR's intension is something like "anything that is an unmarried man" and its extension is the collection of bachelors.

3  It might be objected that for Gettier, writing his paper was not the first step in conducting an empirical test. Whatever testing was involved in Gettier's process, it was over by the time Gettier wrote and published the article. The purpose of writing and publishing the article was to *let us know* that there are counterexamples to JBT—not to *check whether* there are such. However, from the perspective of the community of inquiry, nothing was established simply by Gettier thinking up scenarios and intuiting about them. It is only because Gettier's scenarios and intuitions were *so good*, and produced immediate and predictable assent in the community at large, that Gettier could write up his paper with the confident purpose of just *letting us know* something.

4  According to Google Scholar, Gettier's paper has been cited almost 2000 times. Oversimplifying a lot for the sake of drama, one could boldly assert on this basis that Gettier's pool of *relevant subjects* is in fact considerably larger than Knobe's.

5  This point is related to the so-called expertise defense of traditional armchair conceptual analysis (Kauppinen, 2007; Williamson, 2011).

6  Another is that philosophers usually consider both of Knobe's scenarios, whereas experimental subjects are provided with only one of them. In general, social-psychology-style experiments in which subjects are presented with contrasts yield different results than ones in which subjects are divided into groups each of which is presented with only one side of the theorist's intended contrast. (Thanks to Brent Strickland for pointing this out to me.)

7  More generally, although one can imagine an experimental study revealing to philosopher readers the error of their ways, it is harder to imagine the study doing this *solely* by laying out folk intuitions that manifestly differ from philosophers' intuitions. In saying "solely," I want to rule out here cases in which a study involves an ingenious new thought experiment that philosophers had not previously considered, or raises substantive considerations that philosophers had previously failed to appreciate. For those elements do not go to the study's status as experimental and could be integrated into any piece of traditional armchair philosophy.

8  I have personally engaged in translations from French and German to English and from English to Hebrew and can attest that the phenomenology of the non-identity of translating and translated words is very stark.

9  The reasons to adopt a principle of charity in this area are the same as the reasons to adopt the principle of charity in interpretation in general: (a) there are transcendental rationality constraints on the understanding of others (Davidson, 1974) and (b) there are evolutionary pressures to weed out incoherent concepts, irrational beliefs, and so on (Dennett, 1987).

10  Of course, an alternative reaction to the truth-insensitivity findings is to dismiss *instinctual* verdicts on thought-experimental scenarios and replace them with

*considered* verdicts in the study of concepts—verdicts formed after suitable reflection, including philosophical reflection.

11  It may still have something *indirectly* to do with philosophy, insofar as philosophers can and often do appeal to empirical science, including social psychology (of which x-phi is a branch), to inform their treatment of traditional philosophical problems.

12  I am assuming here that there is some substantive essence to philosophy. Just as the "institutional theory of art" claims that what makes an object a piece of art is simply the fact that it is in a museum, one could paddle an "institutional theory of philosophy" according to which what makes a bit of inquiry philosophy is that it is performed by people salaried by philosophy departments. I am assuming here that this is false. One counterexample to it is Plato, who was not salaried by any philosophy department, but whose work is philosophical; obviously, there are many others.

13  I develop my take on the matter in Kriegel (2016).

14  I use "statement" as conveniently ambiguous between proposition and sentence, and "notion" as conveniently ambiguous between concept and word. In either reading, a notion is a constituent of a statement.

15  Essentially the same project is developed in greater detail in David Chalmers's *Constructing the World* (2012). Chalmers goes beyond Jackson in *arguing* for the central presupposition of serious metaphysics, namely, that *there exists* a small group of truths from which all others can be derived. In addition, for Chalmers the relationship between fundamental and derivative truths is that of "scrutability," where "$p$ is scrutable from $q$" means that the conjunction of $q$ and all a priori truths deductively entail $p$.

16  Tautologies using derivative notions, such as "Every policeman is a policeman" may still be entailed by the fundamental truths, but non-tautological statements with derivative notions will not.

17  Existentials are in some respects the most basic non-fundamental statements requiring conceptual analysis to assess the truth of (given the fundamental truths). For other statements invoking a non-fundamental notion N will typically entail existentials in which the existence of N's referent or extension will be claimed.

18  Observe that the example, which will follow us through much of the paper, concerns *the* sun and not *a* sun. Accordingly, I label the concept SUN (upper-case "s") rather than SUN (lower-case "s").

19  In other words, the *Merriam-Webster* Ramsey sentence is as follows: (SUN$_{MW}$) There is a (single) $x$, such that $x$ is the luminous celestial body, around $x$ revolve the earth and other planets, from $x$ the planets receive heat and light, $x$ is composed mainly of hydrogen and helium, $x$ has a mean distance from earth of about 93,000,000 miles (150,000,000 km), $x$ has a linear diameter of 864,000 miles (1,390,000 km), and $x$ has a mass 332,000 times greater than earth.

20 Quine's picture of how concepts work thus differs from the traditional descriptivist picture along two dimensions. First, in lieu of the dichotomy between fully nonnegotiable and fully negotiable statements about Cs, it offers a gradient model in which statements about Cs vary in their degree of negotiability (or revisability) along a spectrum. Secondly, in Quine's model there are no fully nonnegotiable statements about Cs—no convictions whose degree of revisability is strictly zero. (One might have proposed that concepts work like dimmer lights: once they are on, they have degrees of brightness, but they can also be completely off, with the difference between "on" and "off" being binary and determinate. But in Quine's model there is no "off" mode for negotiability: there are only degrees of brightness in the "on" mode.)

21 At the same time, reference failure may result from sufficiently many relatively marginal platitudes turning out to be false. Still, it is a measure of a platitude's centrality that its weight is such as to induce reference failure in conspiracy with relatively few other falsified platitudes. The falsification of "The sun is more or less spherical" would have to be combined with more similar instances than the falsification of "The sun is extremely hot" to results in SUN failing to refer (i.e., turning out to be an empty concept).

22 I say "most" because in the case imagined *at least* eight platitudes needed to be true, which means that some disjuncts have more than eight conjuncts in them.

23 One might also imagine deferential platitudes that defer not to the current experts, but to something like the end-of-inquiry experts. If there is this kind of deferential dimension in some concepts, it would explain why those concepts can co-refer before and after scientific revolutions.

24 The moral I would draw from the discussion of these three challenges is that although different individuals' concepts of the sun may be modeled by different Ramsey sentences, communication between these individuals is nonetheless possible because, and to the extent that, there is sufficient *overlap* among the Ramsey sentences that capture their respective concepts. Crucially, different individuals probably share perceptual, causal, and deferential platitudes, and typically also a number of other convictions about the sun. Furthermore, for many concepts, the Kripkean causal platitude and the Putnamian perceptual platitudes are highly central, perhaps even nonnegotiable: they appear in most or all disjuncts of the relevant Lewis sentence. SUN is a case in point. If "The sun is that of which I seem to be perceptually aware when I have my sun-experiences" turns out to be false, my concept of sun may well be empty after all. If so, this platitude should show up in every disjunct of the Lewis sentence for my sun concept. Plausibly, now, the analogous platitude will appear in every disjunct of the Lewis sentence for my interlocutor's sun concept. This kind of core overlap undergirds communication.

25 Hawthorne and Braddon-Mitchell apply this to the concept CONSCIOUSNESS, with the thought that it works a bit like the concept LIFE: *if* there is a primitive,

irreducible property of consciousness, then CONSCIOUSNESS picks out it, but if there is none, then CONSCIOUSNESS picks out a cluster of functional feature associated with conscious states. This is supposed to show that the conceivability of zombies is perfectly compatible with physicalism about consciousness.

26  As a general rule, it is important to keep in mind that it is not the Ramsification's job to tell us what is intuitive—only to capture whatever psychological structure emerges from what we deem intuitive. I will return to this point in section 4.

27  It does not follow, however, that every concept's structure is correctly described by a rigidified mega-Lewis sentence with deferential, causal, and perceptual platitudes. Rigidification, for example, is relevant only to Twin-Earthable concepts; concepts such as FRIEND are not Twin-Earthable, and therefore do not call for rigidification. Meanwhile, concepts such as BACHELOR are rather flat and structureless, and so do not call for the prioritization characterizing mega-Lewis sentences. Nor do FRIEND, BACHELOR, or CUP call for deferential or perceptual platitudes (there seem to be no recognized friendship experts to whom the lay implicitly defer and no perceptual acquaintance needed to be competent with these concepts).

28  I leave it open whether every fragment of the Ramsey sentence for SUN could *in principle* be detected from the armchair, though they could not *all* be, or whether there are fragments so intricate that experimentalization is necessary to uncover them.

29  A traditionalist armchair philosopher might argue that my proposed rationale for experimentalization only underscores the insignificance of x-phi, as it calls on the latter only for the finer, less central details of concepts. There are a priori reasons to suspect that this cannot be quite right. Philosophy of science used to be in the business of telling scientists what they do wrong. Modern philosophy of science has wisely restricted itself to trying to understand why science works the way it does, rather than trying to reform it. One suspects that the same humble approach would suit the "philosophy of philosophy." Clearly, x-phi is a flourishing research program. So our first priority should be to understand what contributions it makes to research that sustain its flourishing. (One could alternatively go for sociological explanations in terms of changing fashions, of course, but this seems rightly thought of as a last-resort explanation.)

30  Another diagnosis is that our empirical knowledge about the world at the fundamental level is insufficient to come down on reduction or elimination of free will. This second diagnosis does not seem more antecedently plausible. Indeed, it seems to me antecedently less plausible.

31  Note that however many such hypotheses we would falsify, charity would exhort us to seek others before adopting the thesis that the concept is incoherent.

32  Sosa's analysis, if I have understood it correctly, is this: S knows full well that $p$ iff (i) S believes that $p$, (ii) S's belief that $p$ is true, (iii) S's belief that $p$ is competently

held, (iv) it is the case that (ii) because it is the case that (iii) (read: S's belief that *p* is apt), (v) S's decision to make a judgment on whether *p* is correct (i.e., is the right decision), (vi) S's decision to make a judgment on whether *p* is competent, (vii) it is the case that (v) because it is the case that (vi) (read: S's belief that *p* is meta-apt), and (viii) it is the case that (iv) because it is the case that (vi). This is complicated by the standard of armchair conceptual analysis, but is very "clean" as compared to the kind of meta-Lewis sentence I suspect our concept of the sun would require.

33 It is a separate question whether the referent is a homogeneous kind or property, on the one hand, or a "wildly disjunctive" one on the other; but either way there will be a cause.

34 When I say a concept refers *partly* via a description, I mean that the description at least puts constraints on the kinds of thing that are eligible referents of the concept. That thesis would always be a last-resort view.

35 There is also the view, of course, that *all* concepts involve such a descriptive component, or at least that there is *a* notion of reference that is determined in part by descriptive content (Jackson, 2009).

36 This work was supported by the French National Research Agency's ANR-11-0001-02 PSL* and ANR-10-LABX-0087. For comments on a previous draft, I am very grateful to Chad Carmichael, Dan Korman, and David Rose. I have benefited from presenting an early version of the paper at ENS Lyon and a later one at the Jean Nicod Institute; I am grateful to the audiences there, in particular Filippo Contesi, Lilia Gurova, Michele Odisseas Impagnatiello, Max Kölbel, Stephen Laurence, Michael Murez, Nico Orlandi, Jean-Michel Roy, and Brent Strickland.

# References

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

Bealer, G. (1998). A theory of concepts and concept-possession. *Philosophical Issues*, 9, 261–301.

Block, N. J. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9, 261–325.

Brock, S. (2004). The ubiquitous problem of empty names. *Journal of Philosophy*, 101, 277–298.

Brody, B. A. (1977). Kripke on proper names. *Midwest Studies in Philosophy*, 2, 64–69.

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73–122.

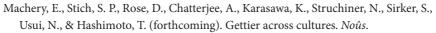Burge, T. (1986). Intellectual norms and the foundations of mind. *Journal of Philosophy*, 83, 697–720.

Chalmers, D. J. (2012). *Constructing the world*. Oxford and New York: Oxford University Press.

Davidson, D. (1974). On the very idea of a conceptual scheme. *Proceedings and Address of the American Philosophical Association*, 47, 5–20.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dunaway, B., Edmonds, A., & Manley, D. (2013). The folk probably do think what you think they think. *Australasian Journal of Philosophy*, 91, 421–441.

Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, 23, 121–123.

Goldman, A. I. (1967). A causal theory of knowing. *Journal of Philosophy*, 64, 357–372.

Hawthorne, J. (2002). Advice for physicalists. *Philosophical Studies*, 109, 17–52.

Jackson, F. C. (1998a). *From metaphysics to ethics: A defence of conceptual analysis*. Cambridge: Cambridge University Press.

Jackson, F. C. (1998b). Reference and description revisited. *Philosophical Perspectives*, 12, 201–218.

Jackson, F. C. (2008). Review of J. Knobe and S. Nichols (eds), *Experimental Philosophy*, Oxford: Oxford University Press. *Notre Dame Philosophical Reviews*.

Jackson, F. C. 2009. Eliminativism and the theory of reference. In D. Murphy, & M. Bishop (eds), *Stich and his critics*, 62–73. Oxford: Wiley-Blackwell.

Jackson, F. C. (2011). On Gettier holdouts. *Mind & Language*, 26, 468–481.

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

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190–193.

Knobe, J. (forthcoming). Experimental philosophy is cognitive science. In J. Sytsma & W. Buckwalter (eds), *A companion to experimental philosophy*. London and New York: Blackwell.

Korman, D. (2006). What externalists should say about dry earth. *Journal of Philosophy*, 103, 503–520.

Kriegel, U. (2016). Philosophy as total axiomatics. *Journal of the American Philosophical Association*, 2, 272–290.

Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (eds), *Semantics of natural language*, 253–355, 763–769. Dordrecht: Reidel.

Kroon, F. (1987). Causal descriptivism. *Australasian Journal of Philosophy*, 65, 1–17.

Lehrer, K., & Paxson, T. (1969). Knowledge: Undefeated justified true belief. *Journal of Philosophy*, 66, 225–237.

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy*, 63, 17–25.

Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249–258.

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, 92, 1–12.

Machery, E., Stich, S. P., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (forthcoming). Gettier across cultures. *Noûs*.

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay denial of knowledge for justified true belief. *Cognition*, 129, 652–661.

Price, H., & O'Leary-Hawthorne, J. (1996). How to stand up for non-cognitivists. *Australasian Journal of Philosophy*, 75, 274–292.

Putnam, H. (1975). The meaning of "meaning." In *Mind, language, and reality*, 215–271. Cambridge: Cambridge University Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60, 20–43.

Rose, D., & Danks, D. (2013). In defense of a broad conception of experimental philosophy. *Metaphilosophy*, 44, 512–532.

Sosa, E. (2009). Knowing full well. *Philosophical Studies*, 142, 5–15.

Starmans, C., & Friedman, O. (2012). The folk conception of knowledge. *Cognition*, 124, 272–283.

Stich, S. P. (1978). Autonomous psychology and the belief-desire thesis. *The Monist*, 61, 571–591.

Suits, B. *The grasshopper: Games, life and utopia*. Toronto: Toronto University Press.

Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research*, 76, 138–155.

Systma, J., & Livengood, J. (2011). A new perspective on experiments concerning semantic intuitions. *Australasian Journal of Philosophy*, 89, 315–332.

Williamson, T. (2011). Philosophical intuitions and the burden of proof. *Metaphilosophy*, 42, 215–229.