

EMBODIED KNOWLEDGE, CONCEPTUAL CHANGE, AND THE *A PRIORI*; OR, JUSTIFICATION, REVISION, AND THE WAYS LIFE COULD GO

Robert D. Rupert

ABSTRACT

This essay defends a qualified version of Quine's thesis of universal revisability against David Chalmers's recent conditionalization-based criticisms of it. It is argued that an embodied view of cognitive processing undermines Chalmers's account of nonrevisable *a priori* justification, which presupposes that concepts prefigure the confirmation-relations into which they enter so as to make such relations rationally accessible to anyone who possesses those concepts. On the view developed here, bodily interaction with the world and the accompanying subconscious processing can change subjects' dispositions to apply their concepts in ways that are not rationally accessible to them, even given a complete description of that interaction, and do not constitute a change in the content of the concepts involved. Thus a subject who treats a proposition as indefeasibly justified *a priori* might nevertheless significantly lower her credence in that proposition, in ways that are not accessible to the subject on the basis of her grasp of the content of the relevant concepts. This discussion has further implications concerning the role of the *a priori* in the philosophical enterprise.

I. INTRODUCTION

W. V. Quine (1951) famously claimed that, in principle, any of our commitments, including commitment to seemingly fundamental principles of logic and mathematics, might reasonably be given up in response to new observations. But, as Grice and Strawson (1956) noted in an early response to Quine, this claim of universal revisability courts triviality unless Quine restricts it to cases in which the relevant meanings remain the same through the change in commitment. *Of course*, any sentence can be rationally abandoned, if the kind of abandonment in question tolerates revisions in credences subsequent to changes in meaning: If “two plus two equals

four” were to come to mean what is now meant by “London is located in China,” then it would be rational to revise downward our credence in the former sentence. Moreover, if a claim of universal revisability can be confirmed by revisions subsequent to changes in meaning, its truth grounds no challenge to *a priori* knowledge; the phenomenon of revisability subsequent to a change in meaning does not impugn the thought that, for instance, “two plus two equals four,” with its current meaning held fixed, is known *a priori* and our commitment to it unrevisable.

In some of David Chalmers's recent work (Chalmers 2011, 2012), he develops a seemingly plausible criterion of conceptual change, which he then wields in compelling

fashion against Quine's claim of universal revisability. By articulating a well-motivated account of conceptual change, Chalmers intends to refine and strengthen Grice and Strawson's anti-Quinean line—showing that universal revisability does not so clearly hold when restricted to cases in which meanings are held constant—and thereby to clear a path for a *a priori* justification of the strongest sort, the non-negotiable kind (Chalmers 2011, pp. 389–390).

In this essay, I defend a limited version of Quinean revisability, one aimed particularly at discussions of philosophical methodology. I argue that, for any claim of peculiarly philosophical interest, for any subject who accepts that claim on supposedly *a priori* grounds, there is a way that subject's life could go such that she would rationally abandon that claim without the contents of her commitments having changed. My development of this thesis takes a somewhat winding path. I begin with some views of Laura Schroeter (2004a, 2006), add a Williamsonian (2007) twist, and round things out with an appeal to work on embodied cognition, the nature of mental representations, and a naturalized notion of rationality. Here is the gist of the picture, in preview: Human subjects' ongoing participation in the world—their daily interactions with objects, properties, and people in their environment—alters their tendencies to apply their concepts and more generally to activate mental representations. Human cognition consists partly in the activation of mental representations, which produce behavior, which behavior often eventuates in corrective feedback, which feedback typically changes the response profiles of the subject's mental representations, which alters the subject's dispositions to engage with the world, and so on and on, day in and day out. Moreover, across a wide range of cases, there's no reason to think this process alters either the identity of the relevant representing units, considered nonsemantically, or their semantics. And

there is mounting evidence that these kinds of subconscious, cognition-shaping processes stand at the center of scientific reasoning and thus, from the Quinean perspective, at the core of rationality. I propose that this embodiment-driven evolution in cognitive dispositions applies to philosophical concepts and commitments, about, for example, causation, responsibility, knowledge, and reference. I take this view to be of interest in its own right; in addition, if correct, it undermines Chalmers's argument against Quinean revisability (at least in its limited form that pertains to distinctively philosophical claims) and, indirectly, supports skepticism about *aprioristic* philosophical method.

II. PRELIMINARIES

Three preliminary points are in order. First, I am not out to prove the thesis of *universal* revisability. It will suffice for my purposes to cast doubt on our ability to acquire relatively secure, armchair knowledge about, for instance, causation, mental content, moral responsibility, reference, time, free action, and knowledge itself. Whether some claims of mathematics or logic are genuinely unrevisable can be left open, for my primary interest is meta-philosophical; on the assumption that a significant amount of philosophical investigation falls outside the scope of pure mathematics and logic, my conclusions bear importantly on our understanding of philosophical method, even if some claims of pure mathematics and logic are known *a priori*.

Second, although Chalmers acknowledges a fallibilist conception of a *a priori* justification, according to which a *a priori* justification can be defeated, he aims to defend the possibility of the strongest sort of a *a priori* knowledge, which rests on a form of justification that does not admit of significant downward revision in the associated credences (2011, p. 389). He takes this tack partly to avoid too easy a response to Quine, a response that simply concedes to Quine that, because of the

fallible nature of *a priori* reasoning, our commitment to any claim is revisable (even with the claim's meaning held fixed) and that thus we never possess entirely secure knowledge justified *a priori*.

I applaud Chalmers's willingness to take up the strong foundationalist mantle. Nevertheless, Chalmers's narrow focus on the infallibilist view of *a priori* justification threatens to limit artificially our exploration of the relations between the epistemic dimensions of revisability and theses about the nature of *a priori* justification. It's one thing to accept the thesis of universal revisability, as a fallibilist about the *a priori* seemingly must do, and it's quite another to accept a particular account of the dynamics of revision. Here, a fallibilist about the *a priori* and a Quinean about revisability inevitably butt heads, as a result of their conflicting accounts of the reasons that universal revisability holds.

Laurence Bonjour, for example, accepts the fallibility of *a priori* justification (although he also thinks some authors exaggerate the extent to which *a priori* justified claims have, historically, been revised—1998, p. 111), while arguing, *contra* Quine, that *a priori* justified beliefs play an indispensable role in the revision process itself. In response to new observations, a subject may well abandon a claim that she had held on *a priori* grounds, but according to Bonjour, that process must itself rest on at least one *a priori* justified belief. The subject must rationally grasp the conflict between the new observations and the *a priori* claim being given up, or else the subject would not be rationally abandoning the claim in question; moreover, the rational appreciation of the conflict between observation and prior commitment itself constitutes *a priori* insight. Thus Quine's thesis of universal revisability hardly undermines an epistemology based on *a priori* justification, according to Bonjour. To the contrary, the only acceptable account of the process of revision itself precludes Quinean skepticism

about the *a priori*. Bonjour rightly takes this to mark a deep distinction between his view and Quine's (1998, pp. 122–123). Central aspects of the rationalist picture rejected by Quine remain in play for Bonjour, aspects to do with the grasping of meanings and the role such grasping plays in justification.

Therefore, although someone who holds a fallibilist conception of the *a priori* can accommodate the thesis of universal revisability, she may do so for reasons inconsistent with a Quinean conception of justification. Thus an evaluation of the connection between revisability and the *a priori*—even one focused in the first instance on an infallibilist conception of the *a priori*—may well inform our evaluation of fallibilist conceptions of the *a priori*. For such an evaluation might be built on views about meaning, conceptual content, and rationality—and their role in revision—that hit at both the infallibilist conception of the *a priori* and simultaneously the picture of revision offered by the fallibilist about the *a priori*. Put more directly, the views on which my criticism of Chalmers's approach rests also cut against Bonjour-style indispensability arguments; if I'm right about the reasons why universal revisability holds, the fallibilist loses her most compelling defense of *a priori* justification against the skepticism about the *a priori* often thought to be engendered by Quine's revisability claims.

Think of the dialectic in this way. When we revise strongly held commitments—even those we previously took to be *a priori* justified—it does seem to us that we are revising our beliefs correctly and rationally (cf. Bonjour 1998, pp. 91–96). To put a fine point on the disagreement between *apriorists* and Quineans, then, we should distinguish between the claims, on the one hand, that such seemings have explicitly modal content and that such content, assuming it's present, plays an indispensable role in the justification of belief-revision, and the claims, on the other hand, that such seemings don't have modally

rich content or, if they do, such content need play no role in the justification of belief-revision. On the Quinean approach I favor, we can replace all revision-related statements of the form “It now seems to me that it *has to be* this way” with statements of the form “This now seems, in fact, right, and very much so,” without undermining the rationality of revision. Thus there is a coherent and plausible Quinean tale of belief-revision—one which I intend to flesh out below—that neither fallibilists nor infallibilists about *a priori* justification can accept.

Third, I’m not out to defend Quine’s meaning of holism, as it might be extended to thought content (see Fodor and Lepore 1992, chap. 2). Quine’s semantic holism rests, to a significant extent, on his behaviorism about matters psychological and a confirmation-based attitude toward semantics. And, to be fair, the best science—and philosophy of science—of Quine’s day might reasonably have been taken to support these behaviorist and verificationist commitments; thus his commitments may appear sensible, viewed in context and in the light of Quine’s naturalistic orientation. Nonetheless, I have no truck with behaviorism or verificationism and, partly for that reason, no intention of defending the holism about meaning to which these views, in Quine’s hands, led.

III. THE KRIPKEAN THREAT TO A *PRIORI* KNOWLEDGE

In *Naming and Necessity*, Saul Kripke (1980) introduces the idea of necessary *a posteriori* claims, which are necessarily true, but which are known to be so (by humans) only by empirical investigation. Consider a claim such as “Water is H₂O,” which we now take to be true, and thus, *qua* identity statement, necessarily true, but which many reflective speakers, at many times, who grasped the relevant meanings, did not take to be true (where the grasping of meanings amounts to such things as having command of the

language sufficient to achieve reference, to think and talk about the things referred to, and to communicate effectively with other speakers about the individuals, properties, and kinds represented in thought and language). Similarly, many speakers who, by every standard measure, grasped the meanings of “temperature” and “mean molecular kinetic energy” were not in a position, absent further empirical data, to justify the true identity statement “Temperature just *is* mean molecular kinetic energy,” even though that statement is necessarily true (according to philosophical lore—cf. Wilson 1985).¹

Chalmers appreciates this threat to *a priori* knowledge (2004, pp. 153–158). He characterizes the problem as a break in the golden triangle connecting reason, meaning, and modality. In an intact golden realm, reason contemplates meaning to yield *a priori* knowledge of what is necessary and possible. The existence of *a posteriori* necessities, however, alienates reason from meaning; reason cannot reliably deliver *a priori* knowledge—knowledge of necessities—by the contemplation of meanings alone. One might equally characterize this as the alienation of meaning from modality. The kind of meanings to which reason has access are not the kind that alone yield reliable access to modal facts. Relative to my concern about philosophical method, it matters not whether, when characterizing the Kripkean quandary, we (a) leave intact reason’s access to meanings but disconnect those meanings from modality, or (b) preserve the connection between meaning and modality and put the meanings in question (sometimes) beyond the grasp of competent speakers or thinkers (Putnam offers his readers an analogous choice—1975, pp. 219–222). Either way, manipulation of the things to which we have immediate access in our reasoning (the meaning-like entities that directly guide our reasoning) may bring us no closer to knowledge of the essential natures of philosophers’ target properties or relations. In

fact, the items to which we have direct access in our reasoning may lead us badly astray; intellectual manipulation of these items might cause us to *deny* necessary truths about the target properties and relations—that is, about causation, reference, free will, justification, and the rest.

To render this concern vivid, consider the possibility that the kinds, properties, or relations of interest to philosophers are natural kinds, the essences or natures of which can be discovered (by humans) only by empirical investigation. In which case, for any given kind, property, or relation of special interest to philosophers—causality, reference, free action, justification—we should investigate it in the way we investigate electricity or dark matter, by careful empirical observation, controlled experiment, quantitative analysis of data, and theory construction and revision. On this view, if there is such thing as justice at all, it is a property or relation out there in the world with a nature waiting to be discovered in the way the nature of electricity was discovered. If this is correct, *a priori* reasoning does not deliver knowledge of the nature of the kinds of interest to philosophers, any more than it delivers of the nature of water or temperature.²

In an attempt to reunite reason, meaning, and modality, Chalmers recruits the resources of two-dimensional semantics (Chalmers 2004). The resulting framework is rich and complex, but for immediate purposes, certain essentials will do. Of greatest importance in the present context is the dimension of meaning—the primary intension—that functions something like a Fregean sense, as distinct from the referential dimension of meaning. Kripke might have convinced the philosophical world that water is H₂O necessarily. But this way of thinking about the meaning of “water” treats worlds counterfactually; it is to focus only on secondary intensions—the actual referents or extensions—of “water” and “H₂O,” asking about the presence or

location of those actually-referred-to things as those very things might appear in other possible worlds. Alternatively, we might consider various possible worlds as actual, without relativizing our judgments about them to what we take to be the reference of our terms (or concepts) in the actual world. Imagine, for instance, that “water” had had its reference fixed in an XYZ-world (Putnam 1975); if we assume that to be the actual world, then “water” refers to XYZ; that is, if we index our discourse itself to that world, our use of “water” refers to XYZ. This is a bit misleading, because we want to be able to treat as actual worlds with no speakers or thinkers in them and thus no language. More precisely, then, we can ask what, supposing only a canonical description of a given world (which Chalmers calls a “scenario”) we would judge “water” to apply to in that world.³ The pattern of such judgments across all complete, coherent scenarios constitutes (or reflects our grasp of) a word’s primary intension. And, in some of those worlds treated as actual, we will not judge that “water” and “H₂O” apply to the same collections of things, which shows that, in one clear sense, “Water is H₂O” is not necessarily true, and that the kind of contingency in question can be known *a priori* solely by consideration of the relevant terms’ primary intensions.

It is not clear, from what’s been said, how the two-dimensional framework might insulate *aprioristic* philosophy from the Kripkean threat. Perhaps the two-dimensionalist has explained our intuition of contingency—the intuition that water didn’t have to be H₂O—but how does two-dimensional semantics secure reliable *a priori* knowledge of necessities? Chalmers characterizes *a priori* justification as the sort of justification one has for a sentence (or claim) *S*, when, for any scenario *D*, the conditional “if *D* [that is, take *D* as actual], then *S*” is judged to be true. For instance, no matter which world one takes to be actual, one will judge that “bachelor,” if it

has a referent at all, refers to things that are a subset of the set of things to which “male” refers. “If *D*, then if something’s a bachelor, then it’s male” is judged true relative to all *D*s, and thus “All bachelors are male” can be justified *a priori*.

Some readers might be inclined to dismiss primary intensions as cooked-up. Chalmers argues otherwise (and, note well, in a way that does not presuppose our ability to *define* interesting philosophical terms—2012, chap. 1). Kripke—and all of his convinced readers—must rely on some kind of conceptual knowledge concerning reference when evaluating the thought experiments Kripke uses to argue that, for instance, names refer rigidly. Similarly, the widespread acceptance of decisive counterexamples to philosophical theories, such as the Gettier counterexamples to the justified-true-belief theory of knowledge, presupposes that we can recognize whether a sentence including a term of interest accurately describes a hypothetical situation treated as actual. It’s plausible to think that primary intensions drive this kind of reasoning and thus that they are fundamental to philosophical inquiry.

I remain unconvinced by such observations.⁴ It is not difficult, however, to see Chalmers’s point. *Some* sort of concept-related dispositions drive our philosophical reasoning, and our theorizing more generally. Concepts of reference, justice, and the rest would seem to be operative in the background, providing a standard of success, at the very least. After all, how can Kripke know he’s identified the nature of reference, unless he had, from the outset, some fairly definite idea of what he was looking for?

Thus far, I’ve described primary intensions and the way in which Chalmers exploits them to characterize *a priori* justification, and I’ve described a reason for positing primary intensions. What of their connection to Quine’s thesis of universal revisability? Chalmers contends that if we accept a claim and, via

our grasp of a primary intension, can see, upon (ideal) rational reflection, that no matter which world we treat as actual, we would accept that claim, then that claim cannot be rationally revised. Our understanding of the primary intension of the claim encompasses the conditions under which we would assign “false” to it; and if there are no such conditions, then that claim is unrevisable, as well as being justified *a priori*. If there are such philosophical claims, Quine’s revisability thesis (even in its limited form) is false.

IV. SCHROETER’S IMPROV-BASED RESPONSE

Laura Schroeter (2004a, 2004b, 2006) criticizes various aspects of this new *apriorist* program, calling into question the fruitfulness of the sort of conceptual analysis described in the preceding section. Schroeter argues that Chalmers presupposes a faulty view of concepts and that, more to the point, the correct view of concepts speaks against Chalmers’s approach. According to Schroeter, conceptual knowledge does not consist in a fixed set of criteria for the correct application of the concepts in question; instead, conceptual knowledge is more akin to an ability to interpret representational practices that themselves remain open to improvisation in response to real-world eventualities. “According to the improv model, there is no stable pattern of assumptions and cognitive dispositions that you treat as your ultimate epistemic criterion for determining what counts as [for example] water come what may” (Schroeter 2006, p. 570). Rather, “[t]he holistic improviser holds himself accountable to the upshot of holistic interpretation of his historical practice with the term ‘water’” (Schroeter 2006, p. 573). If Schroeter’s improv model is correct, the subject does not, by virtue of possessing a given concept, have a fixed recipe for the application of that concept, a recipe that informs the subject’s reflective judgment about how that concept would apply no matter which

world is considered as actual. Instead, it is often only after the fact, that is, after having had actual-world experiences, that the subject can look back and decide what it was, even what fundamental kind of thing it was, that she was thinking about all along.

According to Schroeter, which world one actually is in and which experiences one actually has partly determine one's pattern of judgments concerning what one's concepts apply to in various *other* worlds taken as actual. For instance, the details of our actual observations can affect what sort of thing we take water to be and thus (a) what we go looking for when we ask after the referent of "water," (b) what we accept as an answer to questions about its reference, and (c) what answers we give to object-level questions about what is water in various hypothetical situations. In which case, there is no stable primary intension; which primary intension gets associated with a word or concept depends on which world is being taken as actual, in a distinctive sense: which world is taken to be the world in which one exists and has certain experiences. Consider a scenario in which the prevailing clear potable liquid is a chemical hodgepodge (Schroeter 2006, p. 578). In some such worlds, it turns out that water is a culinary kind, rather than a natural kind. When such a world is considered as actual (or more properly, as experienced), our judgments about what we're looking for in other scenarios (other *Ds*), when we look for water, differs from the pattern we exhibit in the actual world. When we take the culinary-world scenario as actual (or experienced), then, relative to any other *D*, when considering what "water" applies to in that world treated as actual, we'll look for the appropriate culinary kind. As a result, judgments about the conditionals "If *D*, then *S*" can differ depending on which world one actually appears in or, better yet, takes oneself actually to be in.

Schroeter also argues from this world's actual history. Aristotle conceived of water as a "basic configuration of prime matter"

(2004a, p. 437). Nevertheless, he was clearly talking about the same stuff we talk about when we use "water"; thus his overarching theory of water does not provide a static list of requirements that anything must meet if "water" (or the Ancient Greek equivalent) is to apply correctly to it. Schroeter again:

Depending on what his environment actually turned out to be like, Aristotle's concept might have represented a fundamental physical substance, a heterogeneous phenomenal kind, or even a process. To dismiss these different possibilities as a priori impossible would be to fly in the face of our best interpretive and epistemic norms. Aristotle's tacit beliefs about what sort of thing [water] is are no more immune to rational revision in the light of empirical evidence than his tacit beliefs about its essential nature. (2004a, p. 441)

And one might reasonably interpret this as a point about our current situation. Our future could go in such a way that we rationally revise our fundamental characterization of some kinds, without thereby changing the actual-world referents of the associated kind terms (and thus without changing the concepts involved); and this might change our judgments concerning what would count, in other worlds taken as actual, as instances of those kinds.

In response, Chalmers stick to his guns. Our judgments about conditionals of the form "if *D*, then *S*" might vary from world to world considered *as an experienced world*, but this simply reflects conceptual change; it doesn't support a strong revisability claim, and it doesn't preclude interesting *a priori* knowledge arrived at by investigating the primary intension associated with concepts held constant (Chalmers 2012, pp. 227–228). When we consider a world as actual (even one with no thinkers in it at all), we apply the primary intension we actually have. It's another thing to imagine that we, ourselves, are in another possible world, say, a world in which our experiences confer upon us

a concept of water as a culinary kind; and it's another thing still to consider how, if the nature of our experiences in this world change significantly, we might make different judgments about various conditionals when taking other worlds as actual. In such cases, we may simply acquire a different concept of, for instance, water, and associate a different primary intension with "water." Thus the possibility of such cases provides no reason to resist Chalmers's two-dimensionalist route to *a priori* knowledge via the application of the concepts we currently possess and primary senses we currently grasp. Chalmers realizes, however, that this sort of response is likely to be more effective⁵ when accompanied by a plausible characterization of conceptual change (2011, p. 391; 2012, pp. 203–204).

V. CHALMERS ON CONDITIONALIZATION

A concern about Chalmers's conceptual-change-based rejoinder now arises: if Chalmers deploys his two-dimensional framework, adverting to our grasp of primary intensions, in order to characterize concepts and conceptual change, he risks begging the question against skeptics about the *a priori* (among whom Quine may be the best known); it might sound as if he's saying that because we have an *a priori* grasp of primary intensions, we know that *a priori* knowledge can be achieved. Partly for this reason, Chalmers instead characterizes conceptual change in a Bayesian fashion.

Here, Chalmers enlists the rule of conditionalization, according to which a subject's credence in a conditional probability $P|E$ (where P is a proposition and E is a description of total evidence), at t_1 , should be equal to her credence in P at any later time at which she is in possession of total evidence E . In other words, if a subject thinks that the probability of P given E is n , then once she finds out that E , her credence in P should be n . In Chalmers's hands, this becomes a

characterization of conceptual change: "If a subject is fully rational, and if the subject acquires total evidence specified by E between t_1 and t_2 , and if the content of sentence S does not change between t_1 and t_2 , then $cr_2(S) = cr_1(S|E)$ " (2011, p. 401), where the subscript attached to the credence operator corresponds to the time index (cr_1 being the subject's credence in the proposition in question at t_1). This principle entails that, when a rational agent appears not to obey conditionalization, some of the subject's relevant concepts have changed.

Chalmers builds his criticism of Quine's revisability thesis on this foundation, partly by distinguishing between changes in commitments that are prefigured in a concept (those that are reflected by patterns of conditionalization) and those that are, instead, postfigured (Chalmers 2011, pp. 394–395). Changes in commitments that are not prefigured by the concepts involved, and reflected in the credences assigned to the relevant conditional probabilities, are either irrational or result from conceptual change. After all, it sounds crazy (at least from Chalmers's perspective) to insist that one is being rational and using one's concepts in a consistent way while saying: "I used to believe that, of the cases in which E holds, the proportion of those in which P holds is n . You know what? E holds! But, I think the probability of P is actually m , not equal to n ."

Of course, even if Chalmers has accurately characterized conceptual change, that doesn't, by itself, establish that any propositions are in fact known *a priori* or that Quine's revisability thesis is false; perhaps there is no proposition P such that we currently have a high credence in it and our conditional credences $P|E$ are high, for all E . This framework does, however, provide the *apriorist* with a clear and plausible way of accommodating cases that appear to involve the rejection of commitments that had seemed to be justified *a priori*. Many people have been

inclined to think that parallel lines simply cannot intersect, no matter what. Later, they become familiar with evidence in favor of General Relativity as well as the analytic tools (non-Euclidean geometry) for articulating its theoretical claims and constituent models. As a result, they reject the claim that parallel lines cannot intersect. On the assumption that these subjects are not being irrational, the conditionalization test entails that their concept of parallel lines has changed. This would seem to be the right conclusion; at least by informal poll, subjects felt that their concept of a parallel line changed when they were first introduced to General Relativity and non-Euclidean geometry, or at least that they were being introduced to an alternative concept of parallel lines.

At this point, one might object that, despite the intuitive appeal of Chalmers's analytic framework, humans are not likely to be able to exploit it to acquire secure *a priori* knowledge. Humans simply lack the appropriate intellectual powers; failures of imagination abound, and as a result, for any given proposition *P* that we think we know *a priori*, we are in no position to say with confidence whether it seems *a priori* (a) because our concepts prefigure no change in commitments or (b) because our concepts do prefigure change, but it's prefigured relative to *E*s we haven't the imagination to consider or the ability to grasp. We would be in a position to differentiate between these possibilities only if we were ideally rational (and were to possess computational powers far beyond actual human ones) and could scrutinize the full range of compact world-descriptions and issue confident judgments about them.

A second objection, or family of objections, adapts extant criticisms of conceptual- or inferential-role accounts of meaning (Block 1986) to the present case. Chalmers's conditionalization test for conceptual change seems to presuppose a way of individuating concepts of a piece with inferential-

conceptual-role semantics: his appeal to conditionalization seems to entail that patterns of confirmation relations partly individuate concepts; otherwise, one could rationally change—for at least some concepts, with regard to some *E*—one's pattern of confirmation relations, thereby allowing for violations of conditionalization, without thereby changing one's concepts.⁶ This, however, leads to an unacceptable holism about concept individuation. Given the likelihood that no two people, relative to a given term, are committed to precisely the same set of confirmation relations, no two people possess the same concept; and if this is overstating the worry, it does not do so by much.

We might wring three more specific problems from this general observation about holism. First, confirmational holism about concept individuation robs concepts of their role as word meanings for a shared language.⁷ Second, it deprives concepts of their explanatory role in confirmation-related reasoning. If I connect only contingently (that is, not as part of the nature of the concept) a concept's correct application with certain kinds of evidence, then the fact that I reach a given conclusion can be explained by the contingent fact that I connected such-and-such concept with those bits of evidence. No such explanatory work (or work in the modeling of cognitive processing) can be done, however, by concepts that are constitutively grounded in the confirmation-related inferences one draws from them. One can't appeal to content that is constituted by the inferences in which a concept participates in order to explain those inferences (Fodor 1998, chap. 1); similarly, if one wants to explain how concepts enter into the cognitive process of confirming or disconfirming beliefs, it's of no use to appeal to the differential presence of concepts individuated in a way that rests constitutively on confirmation relations. Third, it raises a problem about the import of philosophical reasoning. If concepts are sliced as thinly

as confirmation relations, then there's an infinity of concepts. Given that credences are real-valued, relative to a specific *SIE*, the credence-menu is infinitely long. And what might determine that we should commit to one menu item rather than another? What determines, on Chalmers's view, which concepts we *should* be using, the concepts the application of which track the properties and relations we care about? When I talk about moral responsibility, I want to be talking about the sort of responsibility that Hitler *actually had* for Nazi atrocities. I don't mean to be thinking about any of an infinite number of alternative conceptions that we could associate with "moral responsibility" (and there's nothing special about the example given; the problem arises just as forcefully with regard to reference, causality, knowledge, etc.). If this doesn't exactly trivialize *a priori* justification, it makes it seem that, at any given time when we're reasoning *a priori*, we're likely to be reasoning about something other than what we mean to be reasoning about (cf. Dennett 2006). Part of the point here is that *a priori* knowledge is too cheap on Chalmers's view. A subject can simply commit to a certain range of confirmation relations (in the form of conditional probabilities), then claim to know things *a priori* about the concept that has just been individuated by those commitments. But, in that case, it seems unjustified to think the *a priori* knowledge so acquired accurately reflects the nature of the properties or relations of interest when we point to, for example, an instance of moral responsibility (or reference, or justification, and the others) in the actual world.

Think of the third holism-related objection in this way (and perhaps this is better treated as a further objection). The typical subject, relative to any given time, has made a limited number of explicit commitments to conditional probabilities and, explicit commitments aside, has a limited range of determinate dispositions to issue judgments

about conditionals of the form "if *D*, then *S*." When pressed, the typical subject wouldn't be sure which is the right way to add to that list. In other words, even her rational dispositions don't determine a single correct way to go on (cf. Kripke 1982). Thus, for the typical subject at a given time, her explicit commitments and cognitive dispositions are consistent with an infinite number of differing ways to "complete" her set of explicit commitments or to round out her cognitive dispositions. There are simply too many possible sets of total evidence and too many possible ways to respond to them (recall that credences are real-valued); life is short and human interests (the pressure that drives commitments or the formation of determinate dispositions) too narrow. If, when asked about the conditional probability of a given *P* relative to a set of total evidence she's never considered, she "chooses" one possible value, she then commits herself to responding in a certain way to that evidence, if it ever comes into her possession; if she chooses a different possibility, she commits herself to responding to that evidence in a different way; either way, she might then be able to arrive at *a priori* knowledge about the concept individuated by the pattern of future responses to evidence that chooses to endorse, amongst all of the possibilities. But, what reason is there to think that the pattern of confirmation relations chosen was built into the concept she already possessed or, more to the point, individuates a concept that accurately captures the nature of the properties in the world we're interested in, rather than constituting one aspect of one of an infinity of gerrymandered patterns of response to evidence. When I first heard puzzles about personal identity over time, I felt as if I was faced with a choice concerning how to apply the relevant concepts. I also wondered why I should commit to one of these choices over the others. And, on the assumption that I make a choice of one pattern over the others, why think my choice reflects reality (the real

relation *being-the-same-person-as*) or reflects a better choice of conditionalization patterns than the pattern my neighbor commits to—and which is then reflected, for instance, in my neighbor’s responding differently to thought experiments from the way in which I do?⁸

As I’ll argue below, our confirmation-related dispositions change frequently. Moreover, the degree and frequency of such change should make us wonder why the concepts we have at any given time, by Chalmers’s criterion of concept individuation, are of any particular philosophical interest. Chalmers’s framework may offer us a way to acquire unrevisable *a priori* knowledge at a given time, yet that knowledge may be about merely possible phenomena and properties. Thus Chalmers’s approach offers us no way of knowing, without scientific enquiry, whether the *a priori* knowledge that might result from the application of his framework concerns the properties, kinds, and relations that populate our actual world. Perhaps, then, Quine was, technically, wrong. It’s not that all of our (distinctively philosophical) commitments are rationally revisable; rather, it’s that any concept is rationally “abandonable.” Perhaps a subject can have an enormous amount of *a priori* knowledge about her concepts (about what they would apply to and how they relate to each other) at a given time, but because these Chalmersian concepts are not widely shared and most likely won’t be possessed for long, it is trivial knowledge. Given the way in which the human mind-brain is built, we change Chalmersian concepts naturally, frequently, and without reflection. Only when scientific investigation guides this process of conceptual change will we come to possess concepts that accurately characterize the kinds and properties in our actual world. In contrast, *a priori* knowledge delivered via Chalmers’s framework will be abundant but apparently unimportant. For it is constituted by an understanding of some, but not all, of

the infinite number of modal profiles related to confirmation and evidence and there will be nothing privileged about the profiles on which we choose to focus.

In the remainder, I focus on the question of revisability, arguing that, for any given philosophical claim to which one is committed, there is a way one’s life could go such that it would not be irrational to give up that claim in a way that violates conditionalization and to do so without changing the content of one’s concepts. But these arguments could just as well be cast in terms of concept abandonment: for any subject and any pattern of confirmation that individuates (in the Chalmersian way) a concept of philosophical importance, there is a way that subject’s life could go such that it’s not irrational for the subject to abandon (not necessarily consciously) that concept for another, perhaps closely related one. This is a concession of sorts to Chalmers, but a hollow one. The picture described above in connection with concept abandonment seems to offer no consolation to the rationalist in search of secure *a priori* philosophical knowledge: at any given moment, a subject can have *a priori* knowledge about the application of her concepts (although perhaps only upon extensive rational reflection), but a subject’s concepts shift and change constantly as she interacts with the world, in ways that are inaccessible to consciousness. Moreover, although these conceptual shifts may tend to move the subject, over time, closer to concepts that track actual kinds and properties in the world, the subject doesn’t, at most times, have concepts from which she can extract *a priori* knowledge about actually instantiated properties and relations of interest. The concepts from which the most useful *a priori* knowledge might be derived are, on this view, likely to be the ones she doesn’t yet have and will acquire only by participating in—or at least attending carefully to the fruits of—the scientific project.

VI. WILLIAMSON,
EMBODIED COGNITION,
AND WAYS LIFE CAN GO

In this section and the next, I defend widespread revisability. I do not claim that, for any sentence in which a subject has a high credence, there is some *E*, such that the subject's current credence for *S*, conditional on *E*, is low. I do not even claim this relative to the domain of interest to me—distinctively philosophical sentences or claims, which might exclude logical and mathematical claims. Rather, I argue that for any subject, for any *S* (of distinctively philosophical interest), which is such that, for any *E*, the subject's credence in (*S*|*E*) is high, there is a way her life could go, *L*, such that the subject will, subsequent to *L*, have a low credence in *S*; and relative to such *L*s, it's not the case that a subject can typically currently recognize, from a description of *L*, that her credence in *S* will be low subsequent to *L*; moreover, this reduction in her credence in *S* involves neither conceptual change nor irrationality.⁹

I say more presently about *L*s, but first a bit about inspiration. As part of a discussion of philosophical method and of thought experiments relevant to questions in epistemology, Timothy Williamson remarks: "Why should not subtle differences between two courses of experience, each of which sufficed for coming to understand 'know' and 'believe,' make for differences in how test cases are processed, just large enough to tip honest judgments in opposite directions?" (2007, p. 168). And also: "In a similar way, past experience of spatial and temporal properties may play a role in skilful mathematical 'intuition' that is not directly evidential but far exceeds what is needed to acquire the relevant mathematical concepts" (2007, pp. 168–169). To be clear, I do not impute to Williamson the points made below,¹⁰ but one aspect of his vision strikes me as essentially correct and of a piece with what follows. He claims that there's an unfilled gap between

the collection of sensory experiences that could plausibly be treated as partly constitutive of a concept (or sensory experiences that confirm the application of a concept), on the one hand, and general knowledge about the application of the concept that is justified *a priori*, on the other; and Williamson fills this gap—this realm between the *a posteriori* and the *a priori*—with subconscious simulation and other forms of skill that frequently play a central role in embodied approaches to cognition.

Schroeter, too, sometimes emphasizes skill (2004a, p. 441; 2006, p. 573). In contrast to the way in which I would treat such skills, however, Schroeter consistently privileges a commonsense, first-person epistemic perspective and a personal-level reflective ability to make sense of one's representational practices (e.g., 2006, p. 583). I opt for a more naturalistic approach, drawing on work in cognitive science to try to show that concepts should be understood in such a way that our possession of a given one does not prefigure its confirmation-related behavior in the way Chalmers demands. As I see things, the confirmation-related behavior of our concepts is substantially a function of bodily interaction with the world and the changes such interaction brings. More in keeping with Williamson's view, I take these processes to be, in the typical case, subconscious, in the sense that they are not directly accessible to verbal report.¹¹

What is an *L*, a way life could go, and what is its connection to cognitive science? I have in mind the rich, ongoing set of interactions a human organism has with the world that shape the processes in the organism that produce intelligent forms of behavior (that is, that produce the *explananda* of cognitive science). Although most philosophers tend to focus on what they describe as conscious, reflective, deliberative reasoning, it is a commonplace of cognitive science that much, likely the lion's share, of our behavior is

generated largely by subconscious processes or, to put things more accurately, by processes the variations in which are not reliably tracked by the contents of the subject's verbal reports (such processes might affect verbal reports in regular ways, but the *contents* of those verbal reports don't accurately describe the processes that produce them, causally affect them, or reliably co-occur with them in the cognitive system). Even the processes that philosophers are inclined to describe as "personal-level" proceed in steps executed in ways that subjects don't reliably introspect, by mechanisms the workings of which subjects don't reliably introspect, affected by factors subjects don't reliably introspect. The literature here is enormous and continues to grow (a small sampling: Nisbett and Wilson 1977; Lau, Rogers, and Passingham 2007; Kahneman 2011; Huang and Bargh 2014; for philosophical glosses of some of this work, see Gendler 2008; Rupert 2011; and Alfano 2012).

What is happening with us cognitively, then, when we solve problems? Consider one kind of case, reasoning about the physical world by the use of simulations. Mary Hegarty describes the mechanisms at work in such reasoning thusly: "This research suggests that mental simulation is based on internal spatial representations of mechanical systems . . . involves analog imagery, and can be dissociated from reasoning based on descriptive representations or explicit knowledge" (Hegarty 2004, p. 281). Many embodiment-oriented theories overstate their case—claiming that all concepts reduce fully to sensorimotor processing, for example—and I have no interest in pressing such sweeping theoretical claims. But, the basic evidence of the contribution of subconscious motor- and bodily-related processing to the production of intelligent behavior appears to be robust. Consider, for example, Daniel Schwartz's (1999) work on the use of dynamical modeling to solve such problems as the glass-tilting problem:

Two glasses of equal height, one wide and one narrow, are filled to the same level. To what angle must each be tilted in order for the water to begin pouring? Is it the same angle? It's well demonstrated that subjects can construct mental images from descriptions. It turns out, though, that subjects do not seem to use mental images constructed from spatially oriented descriptions to solve the glass-tilting problem; in fact, such images interfere with their ability to solve the problem. Rather, subjects' responses seem to be driven by subconscious expectations about such things as the relevant forces involved and how they relate to movement. Moreover, this skillful understanding of mechanical processes emerges from interaction with the world. This kind of experimental work is fascinating because it illustrates the role of so-called dark processing in problem-solving, but also—and of special importance in the present context—because it helps us to see why an intellectually sophisticated subject might not be able to generate accurate predictions about how she would apply her concepts in world *D* (considered as actual) in which she has undergone *L*, or if she were in possession of total evidence *E*, as a result of having undergone *L*, given only representations (say, linguistic descriptions) of the relevant *D*, *E*, and *L*. Comprehension of the description of an *L* confers upon the subject neither the concept-applying dispositions she would have subsequent to *L* nor an understanding of those dispositions sufficient to allow her to predict what someone with those dispositions will apply her terms or concepts to.

To return to some of the questions raised in the immediately preceding section, we might ask whether the honing and changing of such dispositions create new concepts. The answer depends partly on what we want a theory of concept individuation for. And, to ask a perhaps prior question, "What is the purpose of talking about concepts at all?" One obvious answer to the latter question is that concepts

are (or correspond to) the components of the contents of the very claims to knowledge of interest in the debate over revisability and the *a priori*. If that's so, however, then individuating them in Chalmers's (implied) way will frustrate our purposes. We would like concepts to be the sort of thing that can compose claims subject to collective confirmation, about which we can share insights, and about which we can disagree meaningfully, even if our confirmational patterns don't match up perfectly with each other's (cf. Williamson's remarks about competent users of "know" and "believe"; 2007, p. 168).

Having taken Chalmers's way of individuating concepts off the table, where do we turn for a theory of concept-individuation? Given my naturalistic leanings, I recommend that we answer questions about concept individuation by asking about the causal-explanatory role of mental representations¹² in contemporary scientific work, in particular, about the role they play in models of the data relevant to cognitive science (i.e., instances of intelligent behavior).¹³ I make no attempt here to present a detailed theory of mental representations, but notice that two posited aspects of them, of great use in cognitive science, remain the same across many changes in embodied skill: syntax and representational content (cf. Fodor 1990, chap. 6). To be clear, further, less-stable features of mental representations interest cognitive scientists—for instance, their inferential roles and the substantive knowledge structures associated with individual concepts or mental representation—but the most coherent interpretation of causal-explanatory work in cognitive science takes the former, more stable features of mental representations to individuate the representations and to provide the sort of stability required to ground and flesh out appeals to actual inferential roles and knowledge structures. The central distinction operative here is sometimes described as a distinction between concepts and conceptions or between knowledge structures

and the atoms out of which those structures are built (for more on this, see Rupert 2008, 2013). It's gratuitous to add something further, to try to identify the concept with the structure in which it appears, particularly in light of the objections raised above to Chalmers's way of individuating concepts. Thus, although we might (or might not) change beliefs—which encode inferential patterns or knowledge structures—as the result of embodied experience, independently motivated accounts of mental representations will not categorize such changes as a change in concept, but rather a change in beliefs of which the unchanged concept is a component or a change in the knowledge structure that governs our reasoning about, or interaction with, a single property or kind of thing that was represented throughout changes in the knowledge structure in question.

In sum, then, cognitive science seems to provide an account of concept individuation—as the continued contribution of the same nonsemantically individuated units with the same externalist content—that entails unchanged concepts in many cases in which embodied experience alters dispositions to apply the relevant terms or concepts. It's plausible then that a subject can violate conditionalization—even in situations in which a description of the relevant *E* includes a description of the way the subject's life goes—without its being the case that her concepts changed over the relevant period of time or that she was irrational.

VII. RATIONALITY, NORMATIVITY, AND EPISTEMOLOGY NATURALIZED

Why, though, aren't changes in embodied skill—or, to be more neutral, changes in patterns of concept application that result from bodily experience in the world—nothing more than common forms of irrationality? It's easy enough to imagine Chalmers taking on board nearly all of what's been said here, but

writing it off as lacking in normative import (cf. 2011, p. 410). At this juncture, naturalist rubber hits the road.

Philosophers often criticize Quinean naturalized epistemology because, they claim, it lacks the appropriate normative force (Kim 1988). Quine wants to replace epistemology with psychology, but that's merely to change of subject, the criticism goes. Epistemologists care about how we *should* reason or how we *ought to* form our beliefs, but psychology describes only how we do, in fact, reason, or how we do, in fact, form our beliefs. I contend that this criticism rests on a fundamental misunderstanding of Quine's naturalized epistemology. On his view, normativity enters the picture at the very first step, when we categorize certain belief-forming enterprises (contemporary physics, for instance) as successful. Ask Quine where the epistemically good stuff is, and he points to what appear to him to be the most successful knowledge-gathering enterprises. Psychology (or nowadays, cognitive science) can tell us about the processes leading from stimulation at the sensory receptors to theory-fixation among successful contemporary natural scientists, and it thereby tells us how to get the epistemically valuable stuff; the initial intuition, the one pointing at the relevant scientific enterprises, had the form "those are cases in which people got the epistemically valuable stuff," and thus provides a source of normativity. ("What's the status of these intuitions?" one might ask. They are revisable, subject to holistic considerations of what seems right, all things considered, just as the naturalist is inclined to treat all intuitions—at least until systematic dogmatism begins bearing fruit in the sciences.)

More directly connected to matters at hand, however, are results pertaining to physical reasoning, which explains my choice of Hegarty's and Schwarz's work above as illustrations of embodied concept application.

Admittedly, the work done by such researchers typically involves lay subjects, not working scientists. Nevertheless, much of the work on skilled interaction with the world focuses on processes that seem to be of a piece with what's done in the lab (and this point sits well with Quinean views about the continuity of everyday problem-solving skills and scientific activity). In addition, though, some such work examines scientific reasoning in particular; for instance, Nancy Nersessian (1999) connects simulative—not propositionally encoded and not-necessarily consciously accessible—problem solving with what is one of the most fundamental aspects of scientific inquiry: the use and application of models. It is by now old hat that experimental design is akin to a skill or craft, something developed by trial and error, apprenticeship, the emulation of skilled mentors, and so on. If model application falls into the same category, much of what counts as scientific rationality—paradigmatic rationality on the Quinean view—appears to be the sort of thing that can be shaped by the undergoing of an *L* but the effects of which may not be revealed to the pre-*L* subject by a description of *L*. And, I would add that various other aspects of scientific reasoning fall into the same category; from thought experiments to interpretation of data (say, judging the probability that a particular factor confounds the interpretation of the data), this is the order of the day: the practice of good science depends heavily on subconscious, embodiment-related factors shaped by experience in ways that do not reveal themselves in the discursive descriptions of those experiences given in a statement of total evidence *E* or specification of an *L*. As remarked above, one might be able to predict the effects of one's having undergone a given *L* were one to have in hand a completed cognitive science; but it can hardly be a demand of rationality that one have in hand a completed cognitive science!

To summarize, instances of successful scientific inquiry carry normative force as paradigmatic cases of rationality. The activities of scientists in such cases depend for their success on the application of skills the ongoing development of which can easily lead to violations of conditionalization without conceptual change. Therefore, not all violations of conditionalization involve irrationality or changes in meaning. Therefore, there are normatively rational cases of the violation of conditionalization that involve neither irrationality nor conceptual change. This undercuts Chalmers's way of arguing against Quinean revisability as well as the corresponding defense of *a priori* methodology in philosophy. Moreover, to the extent that I'm correct in claiming that this sort of process might obtain with regard to any philosophical statement in which a subject has, at one time, uniformly high conditional credences, I offer an empirically oriented version of the Quinean revisability argument against the *a priori* justification of philosophical knowledge. And note that the argument, if successful, undermines not only our confidence in the strong form of *a priori* justification that Chalmers means to defend; it also calls into question the indispensability reasoning used by advocates for fallibilist conceptions of the *a priori*, such as BonJour. For if I have presented an accurate picture of the source of violations of conditionalization without irrationality or meaning change, then one can, *contra* BonJour, abandon a claim one had thought was justified *a priori* without having to rely on another *a priori* justified claim; instead, one might have undergone an *L* that changed one's concept-applying dispositions so as to lead to a violation of conditionalization that is neither irrational nor involves meaning change but which issues in a significantly lowered credence in the claim one had previously thought was justified *a priori*.

VIII. CONCLUDING REMARKS

In closing, I remark on what I haven't done here. To make the central argument of this paper stick, as a criticism of *aprioristic* philosophical method, it would be necessary to show that cognitive-scientific results on embodied concept application do in fact extend to philosophically interesting concepts. That such empirical results are forthcoming seems more likely in the case of some philosophical concepts than others. Causality provides a plausible target, given the great extent to which our subconsciously represented strategy for detecting causal relations emerges from our ongoing embodied interaction with the world, from the experiences of detecting contingencies, acting subconsciously on the basis of the representation of the detected contingencies, and receiving corrective feedback from the world—in some cases, enough corrective feedback that the bodily strategies for identifying causal connections may themselves change. Free will and intentional action may provide manageable targets as well, given that our application of the relevant concepts is likely to be guided, in the first instance, by experientially conditioned dispositions to track (or attempt to track) contingencies holding between environmental events and, for example, motor commands as represented in the pre-supplementary motor area. Perhaps justice will fall in line too. For the application of justice-related concepts would seem to be shaped, to a great extent, by our emotional dispositions and reactions, which themselves have significant embodied components shaped by the ways our lives go. The act of referring would also seem, in many contexts, to be deeply tied to our embodied interaction with the world (Rowlands 2006), in which case one might think that *any* philosophical concept that refers (including the concept of reference itself) has its application shaped by interactions with the world; if the bodily

movements that help us to achieve reference are themselves expressions of strategies and dispositions, then if those bodily guided acts of referring tend to be frustrated, our body-based strategies of referring that guide the use of just any concept might change, leading to changes in the dispositions to apply those concepts across cases, and this would include

the strategies for applying the concept of reference itself.¹⁴ Obviously, however, these matters remain open and partly subject to investigation by the flourishing embodied movements in cognitive science.

*University of Colorado at Boulder and
University of Edinburgh*

NOTES

An earlier version of this paper was presented at the University of Edinburgh. Thanks to audience members there, particularly Andy Clark and Brian Raeburn, for their constructive feedback. Thanks also to two anonymous referees for this journal for their suggestions, to David Chalmers for e-mail discussion, and to Bryan Pickel and especially Laura Schroeter for helpful comments on an earlier draft of the manuscript.

1. It's likely that some readers will resist this way of thinking about the issues. For, if the necessarily true propositions in question are of the form $\langle a = a \rangle$, then, of course, any subject who grasps these propositions could justify them *a priori*. Therefore, the claims in question are not *a posteriori* after all, or so it might be thought.

A correct modal point lies in this vicinity, but it's a red herring in the current context. Kripke is interested in the sort of justification that actual subjects possess (1980, pp. 34–35); thus, to say that P is *a priori* justified for S , in the sense relevant here, is to make a claim about S herself, that she in fact justified P on the basis of meanings or concepts to which she has cognitive access. In the germane sense of "*a priori*," then, the mere fact that a proposition P has the form $\langle a = a \rangle$ and thus *could* be justified purely *a priori* (this is the modal point referred to above) doesn't show that all persons who can express or think that P have, or are in a position to have, an *a priori* justified belief that P . To hold otherwise seems to presuppose that being able to think or express that P puts one in a position to inspect directly the proposition so thought or expressed, so directly as to allow the subject to see immediately the form of what's expressed, namely $\langle a = a \rangle$. That presupposition, however, begs the question against Kripke. For, in effect, Kripke argues against just that presupposition, via his various examples of cases in which we successfully refer to properties and individuals and successfully express propositions of the form $\langle a = a \rangle$ despite ignorance and error, that is, despite an inability (the result of having too little or erroneous information) to demonstrate, by our linguistic and other forms of behavior, knowledge of all aspects of the necessarily true propositions being expressed, including their logical forms, and despite the absence of a first-person awareness of a grasp of the relevant aspects of the propositions in question.

Thus, although it may never be an empirical discovery that that $a = a$, it can be a largely empirical discovery that a necessary truth is expressed by a linguistic or cognitive structure that expresses P , where P has the structure $\langle a = a \rangle$ (or, put differently, it can be an empirical discovery that the structure in question, which we've been using competently for some time, expresses P —assuming that my reference to P , right here and now, presents the logical form of P transparently). Thus subjects don't have immediate cognitive access to all modally relevant structural features of a proposition expressed—in particular, features to do with the patterns of modal covariation among the components of the proposition—simply in virtue of being in a position to express it. This suffices to generate the methodological challenge to *aprioristic* philosophizing discussed in the main text.

2. It's fairly clear that the Kripke of *Naming and Necessity* (1980) didn't hold this view. Otherwise, he would have argued for the thesis that proper names rigidly designate by trying to show that reference does the most effective or impressive causal-explanatory work in the relevant empirical sciences when it is supposed that a name gets fixed to the same object across all possible worlds (in which the name has any referent at all), rather than by appeal to intuitions about possible cases involving Nixon, Gödel, and the like.

3. Or what we would judge "water" to apply to upon ideal rational reflection. This qualification is central to Chalmers's framework, but it also exposes the view to a host of problems, specifically to do with a lack of application to actual humans. We have little reason to think that philosophical activity materially approximates ideal rationality, particularly if being ideally rational requires one to possess enormous computing power, of the sort one would think is required to represent consciously, and derive commitments from, world descriptions, compacted though they may be.

4. Here's why. Assume that Fodorian atomism (Fodor 1981, 1998, 2008) holds of philosophical concepts and thus that they don't have the mental analogues of definitions, that is, mental representations of necessary and sufficient conditions for the application of those concepts. Assume, too, that they have secondary intensions and that they also hold positions in nondefining diagnostic or inferential networks, that is, in causally connected networks of mental representations some of which mediate the activation of the others. To use one of Fodor's examples, PROTONs track protons regardless of whether users of "proton," even scientific users, can define the term, and this tracking is effected by the activation of other mental representations that (a) are causally connected to the activation of PROTON (Fodor 1987, p. 121) and (b) are themselves causally sensitive to (and represent) properties correlated with the presence of protons.

Consider how this framework might apply to a philosophical concept, for instance, JUSTICE. Given the assumption of atomism, JUSTICE—as a psychological entity—is not constituted from psychologically real components. Assume, too, that although it has no primary intension, it has, as its secondary intension, a genuine property in the world, *being just* (bracket the question whether properties in the world represented by atomic concepts might themselves be metaphysically compound). Within a given subject, JUSTICE, in all of its atomic glory, bears nondefinitional causal relations to other mental representations (e.g., EQUAL DISTRIBUTION, DESERVED, FAIR), causal relations that help to activate (or inhibit activation of) JUSTICE; in the absence of definitional relations, it can still be the case that, for instance, the activation of FAIRNESS is, for example, 70 percent likely to cause the activation of JUSTICE in a given subject, and that such causal relations as this, holding among the mental representations, help JUSTICE to track—in the sense of "track" commonly used in discussions of causal-informational theories of mental content—*justice* (and similar remarks apply to REFERENCE, CAUSATION, RESPONSIBILITY, etc.). If such subject-specific causal networks are sufficiently similar, subjects might exhibit widely shared reactions to at least some thought experiments (but see Weinberg, Nichols, and Stich 2001), but without its being the case that philosophical terms or concepts have determinate, or even consistent, primary intensions, in Chalmers's sense. Perhaps the collection of other representations to which JUSTICE is causally connected—think of these as the diagnostic concepts—yields messy, context-dependent, even sometimes conflicting results when a wide variety of scenarios are considered, even within a single subject (think of the history of philosophical reasoning about personal identity over time, and think about order effects and framing effects). To be clear, these diagnostic networks might serve us well in our everyday lives, allowing us to track the property *being just* well enough to refer to it, think about it, and recognize its instances in our world. But the associative, probabilistic structure of such networks will produce patterns of responses to scenarios that may miss the mark radically (and may produce a muddled mess when repeatedly applied to a wide range of scenarios). The contingencies represented by their associative structure may well facilitate the tracking of the property in question given the contingencies of our world, but this is consistent with their leading us badly astray in response to descriptions of other worlds taken as actual; the diagnostic

tells that allow us to track justice in this world might throw us off the trail in other worlds, for instance, because the structure of contingencies differs significantly from those in our world—or those diagnostic procedures may have their normal operation confounded by descriptions of other worlds, delivering inconsistent results depending, say, on which aspects of the specifications of worlds we attend to. This kind of approach is psychologically plausible and can accommodate the data—that most philosophers have been convinced by Gettier thought experiments—without any appeal to primary intensions that can ground substantive *a priori* philosophical knowledge.

5. As a response to Schroeter, but also as a way of bolstering Grice and Strawson's (1956) response to Quine's arguments for universal revisability (Chalmers 2011, p. 390).

6. I say "partly individuate concepts" only because Chalmers's view is consistent with there being more than one concept that plays precisely the same role for a subject in her subjective pattern of confirmation relations. Perhaps, for example, the intrinsic qualitative characters of the acts of entertaining two concepts distinguish those concepts from each other even though they play precisely the same role in the subject's confirmational economy. Allowing for this possible complication does nothing to blunt the objections raised in the main text.

7. It's open to Chalmers to assert an analytic-synthetic distinction (or its analogue in a system of mental representations), thereby marking off the shared portion of a concept, but that would beg the question against Quine. Moreover, it would seem to allow for a direct characterization of the *a priori*, rendering otiose any mention of conditionalization and credences: we have *a priori* knowledge of truths guaranteed by the entailment relations into which a concept enters analytically. In contrast, if, in an attempt to preserve a role for conditionalization in the characterization of the *a priori*, Chalmers were to appeal to a cognitively *inaccessible* analytic-synthetic distinction, the resulting qualified criterion of conditionalization would not seem to be rationally binding (otherwise, we would be rationally required to obey conditionalization as it's relativized to an analytic-synthetic distinction to which we do not have access).

8. Chalmers might appeal to ideal rational dispositions here; perhaps various subjects' dispositions will converge by a process of ideal rational reflection that includes philosophical exchange, and perhaps the promise of this eventuality neutralizes concerns to do with indeterminacy of commitments and concept proliferation raised in the main text. This move does not strike me as promising, however, as a defense of *aprioristic* philosophical method. To those interested in philosophical method, in what's accessible to philosophers doing their daily intellectual work, such an appeal to idealized rational dispositions seems either beside the point, by dint of the extreme nature of the idealization, or, if the dispositions are characterized in a way that brings them closer to the human condition, amounts to an admission of defeat on the part of the *apriorist*. If, as would appear to be the case historically, the path to convergence runs, at least partly, through empirical investigation—by the doing of scientific work that shapes new dispositions (in ways that weren't prefigured by our concepts prior to doing that empirical work)—then naturalistic methodology emerges as Chalmers's own unintended recommendation for progress in philosophy. That is to say, if ideal rational reflection is required to put one in a position to gain *a priori* philosophical knowledge, and ideal rational reflection requires access to as-yet-undeveloped dispositions, and if the development of those dispositions requires doing empirical work, then philosophy should proceed in a way that looks a lot more like Quinean naturalism—with all dispositions open to revision—than the sort of armchair philosophizing favored by those who self-identify as rationalists or who accord a fundamental philosophical role to *a priori* justification.

Part of what's at issue can be put a bit differently. It's plausible that, with regard to the vast majority of philosophical claims, our current dispositions (even if assumed to be transparent to reason) don't determine extensions relative to some *Ds*; thus, assuming Chalmers's characterization of *a priori* justification, humans have little or no *a priori* philosophical knowledge. For, there are few, if any, relevant *Ss* such that for any *D*, "If *D*, then *S*" is judged true. Rather, for almost any *S*, there will be at

least some D , such that “If D , then S ” is judged “I don’t know” (and not just because S contains vague terms).

9. A referee for this journal has asked whether the meaning of an E might change upon the subject’s undergoing of L , and whether, as a result, the S in which the post- L subject has low credence might have changed its meaning over the course of L (because the change in meaning in components of E might change the confirmation relations into which S enters), undermining my argument for revisability.

A few comments are in order. First, in the previous section, I criticized a verificationist-style view according to which changes in the confirmation relations into which a concept or representation enters change that concept’s or representation’s meaning. Those criticisms apply in the present context as well.

Second, note that I have universally quantified over Es . The idea is this: begin with a subject and an S such that the subject has a high credence in SIE , for every E whatever; I claim in the main text there is some L or other such that, upon having L , the subject’s credence in S is low (without the subject having been irrational or S having changed its meaning). Generally speaking, a subject’s having undergone L will, as a matter of course, provide her with some E or other she wasn’t in possession of prior to having undergone L ; but whatever that E is, it is a member of the set of those conditional upon which the subject had a high credence in S prior to having undergone L , because that set includes all Es . This holds even if, for some reason, the subject would have used different words to express the E in question prior to having undergone L than she uses to express the now-believed-to-obtain E after she has undergone L .

One might object at this point that the general claim I’ve made in the main text is true but only vacuously so. For, it might reasonably be claimed, there are no subjects who have dispositions relative to all Es , for any philosophical S whatever, and thus that there are no subjects who, relative to a given philosophical S , have a high credence in SIE , for all E . Rather, it seems plausible that (a) for any subject, for any philosophical S , S bears confirmation-relevant relations to at least some events that the subject doesn’t have a means of representing, (b) a subject does not comprehend what she has no means of representing, and (c) one doesn’t have current credences in claims components of which she does not currently comprehend. To be clear, this objection, if it stands, does not vindicate Chalmers’s view, but instead would make all the more pointed some of the objections raised above—for example, that no subject has *a priori* justification for any philosophical S , by Chalmers’s own account, because no subject has a full range of sufficiently determinate credences, and that only by doing science will we come to have the full range of the relevant credences (and thus to be in a position to acquire *a priori* justified philosophical knowledge, by Chalmers’s standard). Moreover, we should ask what happens when the subject comes to understand components of some E (upon undergoing some L) that she currently does not understand. If, as Chalmers’s view would seem to imply, this amounts to a change of meaning in the S of interest (by dint of its now, from the subject’s standpoint, entering into new confirmation-relations), then it would seem to follow that, given the objection on the table and the argument in support of it, no one yet understands the philosophical claims she might, in the limit, be *a priori* justified in believing. Whether this counts as a point against Chalmers’s view of philosophical method depends partly on whether one thinks the process of coming to understand a philosophical S consists in philosophical give-and-take or is, instead, consists in any significant way in the pursuit of science, broadly speaking.

A qualification of my primary claim in the main text is in order, then, at least tentatively. I take it to be a plausible psychological hypothesis that, for any subject, relative to any philosophical S , there exists some Es such that the subject comes to grasp, appreciate, or understand E only as the result of having undergone future L ; and I accept that it would be irrational of (or impossible for) such a subject to have a high credence in SIE , for all E . I maintain, however, that this is all beside the point for present purposes. Take the claim in the main text to be the following: For any given S such that the subject has a high credence in SIE for all Es that she understands, there is an L such that, upon undergoing L , the subject has a low credence in S , without this change in credence being due to irrationality or a change in the meaning of S (or its component concepts), and where it was not a consequence of having under-

gone L that the subject came to possess or grasp a new concept, the understanding of which was the difference-maker with regard to her having lowered her credence in S . We must distinguish questions about the viability of *aprioristic* philosophical method from questions about revisability. Ultimately, I'm more interested in the former, which can be prosecuted independent of questions about revisability. Nevertheless, the proposed qualification secures a substantive revisability claim, the truth of which bears on the evaluation of Chalmers's account of *a priori* justification.

An alternative approach may appeal to some readers and should be kept in mind. Consider a subject's current disposition to have high credence at some later time, t_2 , in SE_1 , where E_1 can be described or understood only by using terms or concepts the subject doesn't currently possess. This provides a way to make sense of a subject's currently having a credence in S conditional upon on some total evidence E , where she doesn't currently understand or have the resources to grasp a representation of E . And, to be fair, this might vindicate a certain aspect of Chalmers's approach, in that it makes room for the possibility of an actual subject's currently having determinate credences for SE , for a given philosophical S relative to all E . But, it also leaves open a possibility that seems empirically well-supported: that the manifestation of the kind of disposition in question, as the result of the subject's life later going a certain way, cannot be predicted by the subject from a current description of the way her life will go. Moreover, for reasons that become clear below, I see no reason to think that a subject violates canons of rationality by failing to have first-person access to the profiles of the dispositions in question. And with regard to changes in meaning, either the obtaining of conditions for the manifestation of the disposition in question changes the meaning of the relevant S , in which case various concerns raised above recur (Why think the components of S , with their current meanings, represent properties instantiated in the actual world or are interesting for any other substantive reason? Why think the process leading to the adoption of new concepts, one more suited to the actual world, will result from philosophical analysis and exchange, rather than from engagement with scientific processes?); or, the obtaining of the conditions for the manifestation of the dispositions in question does not change the meaning of S , in which case, given that the subject doesn't have first-person access to profile of the disposition in question, she can't currently predict the confirmation-related behavior of S and thus rationality cannot demand that she obey the rule of conditionalization on pains of having changed the meanings of her concepts.

10. It's worth noting, for example, that Williamson takes a more optimistic view about the role of thought experiments in the pursuit of philosophical knowledge than is suggested by the picture I develop. I leave diagnosis of this divergence in views for another day.

11. An important related point resides in this vicinity: even if we sometimes report accurately on the kind of cognitive processing in question, we typically can't tell, from a description of an L , how having L will affect our cognitive processing, particularly our patterns of concept application (or the patterns of the activation of our mental representations). Now, as Andy Clark suggested to me, at the end of cognitive-scientific inquiry, one might be able to derive the changes to one's cognitive profile, which will result from a given L , from our true, complete cognitive scientific theory. Fair enough, but such promise is of no use to Chalmers, partly because we don't have such a theory in hand, but also because the correct application of that theory may not be prefigured in the relevant concepts we possess now. If I am correct about the role of subconscious processing, the subject's application of the final theory of cognition will itself be driven by cognitive processes honed by her subconscious interaction with the world, including her practice applying that very theory of cognition (of manipulating formal models and the like). Thus, even if we were to have in hand now a discursive expression of the final theory of cognition and a description of a given L , we might not be in a position to apply the theory skillfully to predict accurately what happens to a subject who undergoes L , at least not until we ourselves have had practice working with the theory and applying it to real cases (and getting corrective feedback that affects our cognitive processing at a subconscious level)—that is, until our lives have gone the way in which lives of the relevant kind of practicing scientists (or sophisticated consumers of the work of the relevant practicing scientists) go.

12. Cognitive scientists tend not to emphasize the personal-subpersonal distinction or the robust conception of first-person rationality that holds sway over many philosophers. Partly as a result of this, they tend not to assign a special status to concepts, as a privileged category of mental representation. This is not to say that cognitive scientists don't use the word "concept"; it's rather that they tend not to pack into it many of the assumptions that many philosophers do, about, say, conscious access, or the rational constraints on the structures in which concepts appear (that, e.g., they must form a consistent set), or free recombination with other concepts, or their essential normativity. The result is that, in the cognitive scientific literature, "concept" is often used interchangeably with "mental representation" in a way that's likely to mislead many philosophers. In the main text, I use "mental representation" both to steer clear of this potential confusion and because I'm pursuing a naturalistic angle, which lends no particular credence to the philosopher's special conception of concepts (or to the attendant idea of a personal-subpersonal distinction).

13. What is intelligent behavior? One thing that seems central to it is its flexibility. That being said, a naturalist doesn't owe the reader much here, over and above methodological remarks. There is, in fact, a range of forms of behavior (writing books, engaging in conversation, translating complex drawings into massive buildings, coordinating electoral processes, and so on) that strike us as in need of explanation—they differ, it would seem, from the regular change of color in leaves and knee-jerk responses at the hands of physicians' hammers—and that's reason enough to give them a label ("intelligent" or whatever) and to try to find a theoretically unified account of them (just as we would do with observable phenomena that strike us as of a piece with regard to, say, electricity or disease). As the project of trying to understand those phenomena proceeds, the phenomena may be split into subgroups or some may be excluded from the domain of relevant *explananda*. We might, for instance, try to model all of the kinds of data initially taken to be relevant to cognitive scientific theorizing, with the result that 90 percent of them admit of a unified explanation and 10 percent are accounted for quite differently. In that case, we'd say that not all of the data were of a piece after all and aren't all instances of intelligent behavior. No problem here. The naturalist should embrace this as a standard aspect of the interplay between theory and data.

14. Changes in such dispositions will not necessarily change the externalist content of the representation involved. This may instead amount to a reduction in error rates in the application of the concept in question to the kind, property, or individual it referred to all along. For a theory of referential content that allows for the fixation of determinate externalist content even while error rates are high, see Rupert (1999).

REFERENCES

- Alfano, Mark. 2012. "Expanding the Situationist Challenge to Responsibilist Virtue Epistemology," *Philosophical Quarterly*, vol. 62, no. 247, pp. 223–249.
- Block, Ned. 1986. "Advertisement for a Semantics for Psychology," in *Midwest Studies in Philosophy*. Vol. 10: *Studies in the Philosophy of Mind*, ed. Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein (Minneapolis: University of Minnesota Press), pp. 615–678.
- BonJour, Laurence. 1998. *In Defense of Pure Reason* (Cambridge, UK: Cambridge University Press).
- Boyd, Richard. 1991. "Realism, Anti-Foundationalism, and the Enthusiasm for Natural Kinds," *Philosophical Studies*, vol. 61, pp. 127–148.
- Chalmers, David. 2004. "Epistemic Two-Dimensional Semantics," *Philosophical Studies*, vol. 118, pp. 153–226.
- . 2011. "Revisability and Conceptual Change in 'Two Dogmas of Empiricism,'" *Journal of Philosophy*, vol. 108, no. 8, pp. 387–415.

- . 2012. *Constructing the World* (Oxford, UK: Oxford University Press).
- Chalmers, David, and Frank Jackson. 2001. "Conceptual Analysis and Reductive Explanation," *Philosophical Review*, vol. 110, pp. 315–361.
- Dennett, Daniel. 2006. "Higher-Order Truths about Schmess," *Topoi*, vol. 25, no. 1, pp. 39–41.
- Devitt, Michael. 1994. "The Methodology of Naturalistic Semantics," *Journal of Philosophy*, vol. 91, no. 10, pp. 545–572.
- Fodor, Jerry A. 1981. "The Present Status of the Innateness Controversy," in *RePresentations*, by Jerry A. Fodor (Cambridge, MA: MIT Press), pp. 257–316.
- . 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: MIT Press).
- . 1990. *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press).
- . 1998. *Concepts: Where Cognitive Science Went Wrong* (Oxford, UK: Oxford University Press).
- . 2008. *LOT 2: The Language of Thought Revisited* (Oxford, UK: Oxford University Press).
- Fodor, Jerry, and Ernest Lepore. 1992. *Holism: A Shopper's Guide* (Malden, MA: Blackwell).
- Gendler, Tamar Szabó. 2008. "Alief and Belief," *Journal of Philosophy*, vol. 105, no. 10, pp. 634–663.
- Grice, H. P., and P. F. Strawson. 1956. "In Defense of a Dogma," *Philosophical Review*, vol. 65, no. 2, pp. 141–158.
- Hegarty, Mary. 2004. "Mechanical Reasoning by Mental Simulation," *TRENDS in Cognitive Sciences*, vol. 8, no. 6, pp. 280–285.
- Huang, Julie, and John Bargh. 2014. "The Selfish Goal: Autonomously Operating Motivational Structures as the Proximate Cause of Human Judgment and Behavior," *Behavioral and Brain Sciences*, vol. 37, no. 2, pp. 121–175.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux).
- Kim, Jaegwon. 1988. "What Is 'Naturalized Epistemology'?" in *Philosophical Perspectives*. Vol. 2: *Epistemology*, ed. James E. Tomberlin (Atascadero, CA: Ridgeview), pp. 381–405.
- Kripke, Saul. 1980. *Naming and Necessity* (Cambridge, MA: Harvard University Press).
- . 1982. *Wittgenstein on Rules and Private Language* (Cambridge, MA: Harvard University Press).
- Lau, Hakwan, Robert Rogers, and Richard Passingham. 2007. "Manipulating the Experienced Onset of Intention after Action Execution," *Journal of Cognitive Neuroscience*, vol. 19, no. 1, pp. 81–90.
- Nersessian, Nancy J. 1999. "Model-Based Reasoning in Conceptual Change," in *Model-Based Reasoning in Scientific Discovery*, ed. Lorenzo Magnani, Nancy J. Nersessian, and Paul Thagard (New York: Kluwer/Plenum), pp. 5–22.
- Nisbett, Richard, and Timothy Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, vol. 84, no. 3, pp. 231–259.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning,'" in *Mind, Language, and Reality: Philosophical Papers*, Vol. 2, ed. Hilary Putnam (Cambridge, UK: Cambridge University Press), pp. 215–271.
- Quine, W. V. 1951. "Two Dogmas of Empiricism," *Philosophical Review*, vol. 60, pp. 20–43.
- . 1969. "Epistemology Naturalized," in *Ontological Relativity and Other Essays*, by W. V. Quine (New York: Columbia University Press), pp. 69–90.
- Rowlands, Mark. 2006. *Body Language: Representation in Action* (Cambridge, MA: MIT Press).
- Rupert, Robert D. 1999. "The Best Test Theory of Extension: First Principle(s)," *Mind & Language*, vol. 14, no. 3, pp. 321–355.
- . 2006. Review of *Embodiment and Cognitive Science*, by Raymond W. Gibbs, Jr., *Notre Dame Philosophical Reviews*, vol. 2006, no. 8, p. 20.
- . 2008. "Frege's Puzzle and Frege Cases: Defending a Quasi-Syntactic Solution," *Cognitive Systems Research*, vol. 9, nos. 1–2, pp. 76–91.
- . 2011. "Embodiment, Consciousness, and the Massively Representational Mind," *Philosophical Topics*, vol. 39, no. 1, pp. 99–120.

- . 2013. “On the Sufficiency of Objective Representation,” in *Current Controversies in Philosophy of Mind*, ed. Uriah Kriegel (New York: Routledge), pp. 180–196.
- Schroeter, Laura. 2004a. “The Limits of Conceptual Analysis,” *Pacific Philosophical Quarterly*, vol. 85, no. 4, pp. 425–453.
- . 2004b. “The Rationalist Foundations of Chalmers’s 2-D Semantics,” *Philosophical Studies*, vol. 118, pp. 227–255.
- . 2006. “Against *A Priori* Reductions,” *Philosophical Quarterly*, vol. 56, no. 225, pp. 562–586.
- Schwartz, Daniel L. 1999. “Physical Imagery: Kinematic versus Dynamic Models,” *Cognitive Psychology*, vol. 38, no. 3, pp. 433–464.
- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich. 2001. “Normativity and Epistemic Intuitions,” *Philosophical Topics*, vol. 29, pp. 429–460.
- Williamson, Timothy. 2007. *The Philosophy of Philosophy* (Malden, MA: Blackwell).
- Wilson, Mark. 1985. “What Is This Thing Called ‘Pain’? The Philosophy of Science behind the Contemporary Debate,” *Pacific Philosophical Quarterly*, vol. 66, nos. 3–4, pp. 227–267.
- . 2008. *Wandering Significance: An Essay on Conceptual Behavior* (Oxford, UK: Oxford University Press).