Why Can't Armchair Philosophers Naturalize the Mind?

Sinan Dogramaci

Abstract My topic is aposteriori naturalism, roughly the view that mental facts are determined by non-mental facts but philosophers cannot discover the details of the determination from the armchair. Section 1 gives aposteriori naturalism a more precise definition and some motivation. Section 2 turns to critical examination, raising a challenge for the view. In section 3, I show that aposteriori naturalists can answer the challenge I raise, but it requires allying their view with certain other substantive positions concerning the epistemology of mental states, positions specifically concerning the nature of self-knowledge and knowledge of other minds. These other positions are controversial, but they are independently defended and accepted by many. My aim is not to offer a novel defense of these other views, but rather to make it clear that aposteriori naturalism must be held in combination with other substantive epistemological views.

1 Aposteriori Naturalism Introduced

1.1 Definitions

Let **naturalism** be the thesis that every mental fact is metaphysically determined by some non-mental fact. One fact **determines** (or necessitates or metaphysically entails) another when it could not have been that the former but not the latter fact holds. Among various formulations of naturalism, this "supervenience" thesis is perhaps the weakest. I focus on it because I will develop a challenge to it, and a challenge to it is thus also a challenge to all the stronger theses that entail it. While the supervenience thesis may not be the most intrinsically interesting formulation of naturalism, it is the most interesting one to raise a challenge for.

By a **fact**, I just mean a true proposition, a truth. I understand naturalism to apply not only to the actual facts, but to the facts in any possible world. So, some actually false mental proposition that is true in another world is a fact in that world, and naturalism says this mental proposition is determined by some non-mental proposition, one that's also true there; it could not have been that the mental proposition is true while the non-mental proposition is false.

Call a fact or proposition **mental** if it is one we express using any mental vocabulary, and call it **non-mental** otherwise. **Mental vocabulary** includes both **intentional vocabulary**, such as "believes", "fears", and "desires", and **phenomenal vocabulary**, such as "pain", "tickle", "perceptual experience", and "experience as of

a red thing". I can't give an exhaustive list of intentional or phenomenal vocabulary. Some vocabulary, like "perceptual experience", might be counted as both intentional and phenomenal. Being mental/non-mental is an extrinsic feature of a fact, so, on a coarse individuation of facts, a single fact may be both mental and non-mental.

Naturalism entails that any mental fact, say, that I believe birds fly, is determined by the metaphysically strongest non-mental fact (picture a very long conjunction settling all non-mental details of the world). That I believe birds fly is plausibly also determined by some much weaker, more local non-mental fact. Naturalism entails that for any mental fact M, there is at least one non-mental fact N that determines M and M is not also determined by any non-mental fact distinct from and itself determined by N: call such a non-mental fact N a **minimal determiner** of the mental fact M. We are just making precise one understanding of minimal sufficient conditions.

Let **aposteriori naturalism** be defined as the conjunction of a metaphysical claim together with an epistemological claim, as follows. The metaphysical half is just naturalism. The epistemological half says that, for any given pair of non-mental and mental facts, it is knowable aposteriori, not apriori, whether the non-mental fact is a minimal determiner of the mental fact. Let **apriori naturalism** be the view that says such claims are knowable apriori. Note that aposteriori naturalism is consistent with naturalism being apriori; it only claims the minimal determiners cannot be discovered apriori.

The **apriori/aposteriori distinction** is notoriously vague and controversial. I will understand the distinction in the way best suited to my dialectical purposes in this paper: let the apriori truths just be those that it is possible for an ideally rational thinker to know, given only the evidence available to armchair philosophers, whatever that evidence may be.

1.2 Motivations

What motivates interest in aposteriori naturalism? Two inspirations for aposteriori naturalism trace to Kripke. First, in *Naming and Necessity*, Kripke (1980) taught us that it is often aposteriori that one fact determines another: that Hesperus is inhabited determines that Phosphorus is too, but this is an aposteriori truth; that water is in my cup determines that H_2O is too, but this is again aposteriori. Kripke opened up for consideration the view that it is likewise aposteriori that a given non-mental fact is a minimal determiner of a given phenomenal mental fact. Kripke rejected that view, but it has since been endorsed by Loar (1990/1997), Hill (1997), Balog (1999), Block and Stalnaker (1999), Tye (2000, 2009), Perry (2001) and Papineau (2002).

Second, in *Wittgenstein on Rules and Private Language*, Kripke (1982) drew from Wittgenstein (1953) a skeptical argument that no non-intentional fact deter-

mines a given intentional mental fact, which then prompted Horwich (1995, 1998, 2005) and Soames (1998) to endorse aposteriori naturalism as a view that Kripke's skeptical argument fails to refute. For, Kripke's skeptical strategy was to consider all the plausible candidate non-intentional minimal determiners of a given intentional fact, and for each candidate to claim from the armchair that he failed to see how it could be a determiner of that intentional fact; of course, if aposteriori naturalism is true, it cannot be seen from the armchair which candidate is a minimal determiner.

Indeed, it is arguable that the diagnosis Horwich and Soames give of Kripke's failure can also serve to explain why so many smart and creative armchair philosophers have, over many years and despite their best efforts, failed to propose any widely accepted minimal determiners of intentional facts. Non-mental determiners of intentional facts were proposed by Lewis (1970, 1972, 1974), Loar (1981), Stampe (1977), Stalnaker (1984), Dretske (1981, 1988, 1995), Fodor (1987, 1990), Millikan (1984, 1989) and Papineau (1984, 1987). But, as Loewer (1997) concludes a survey of the major proposals, "None of the naturalization proposals currently on offer are successful." If aposteriori naturalism is true, perhaps this should not be surprising.

2 A Challenge for Aposteriori Naturalism: Knowledge of Other Minds

2.1 A Preliminary Point: Non-mental Facts Afford Access to Mental Facts

The challenge I want to raise arises because of a crucial discrepancy in the analogy between the classic Kripkean examples of aposteriori necessitation, and the relationship between the non-mental and the mental.[1] The discrepancy I want to exploit is easy to see, at least once pointed out. It isn't generally possible to infer *from the armchair* anything non-trivial about Hesperus, or about water, on the basis of premises wholly about Phosphorus, or about H_2O. Before I can infer, say, that Hesperus is inhabited, or that water is in my cup, just from the premise that Phosphorus is inhabited, or that H_2O is in my cup, I require the aposteriori premise that Hesperus is Phosphorus, and that H_2O is (or constitutes) water. In stark contrast, it would appear to be widely agreed that any armchair philosopher *can* infer a wealth of non-trivial mental conclusions just on the basis of wholly non-mental premises. To illustrate, just consider one of the most famous thought experiments in contemporary philosophy, Putnam (1975)'s Twin Earth case. Suppose Toscar lives on Twin Earth, a duplicate of Earth with the exception that H_2O is replaced by a different, superficially indistinguishable chemical: Putnam invites us to draw an intentional,

1 Others have raised challenges to aposteriori naturalism by emphasizing discrepancies in the analogy other than the one I will be focusing on. In particular, I will not be re-playing the argument made by Kripke, Chalmers and Jackson against the *apriori or aposteriori* determination of the mental by the non-mental; see Kripke (1980), Chalmers and Jackson (2001), Chalmers (2010).

mental conclusion, namely that Toscar doesn't believe his cup holds water, just on the basis of wholly non-intentional, non-mental premises, in particular the premise that there's no water, no H_2O, in his environment.

This discrepancy by itself doesn't pose any challenge to aposteriori naturalists since examples like the Twin Earth case don't, by themselves, make it plausible that we can apriori discover what any *minimal determiners* of mental facts might be. They only illustrate the preliminary point that non-mental facts afford armchair philosophers a kind of epistemic access to mental facts, whereas no analogous relationship holds in the classic Kripkean cases of aposteriori necessitation. But, we can exploit this feature of the epistemic relationship between the non-mental and the mental to develop a genuine challenge for aposteriori naturalism.

2.2 The Assumption behind the Challenge

The challenge I'm going to develop rests on the following intuitively true assumption: *ordinary people, armchair philosophers included, have an ability to know a wide range of mental facts about other minds, and to acquire this knowledge by an inference solely from non-mental facts*. I don't assume there is no "problem of other minds"; I allow there may be a good skeptical paradox going by that name, but a paradox is not a good argument that we lack ordinary knowledge of other minds. I don't assume our knowledge of other minds is *self-consciously* inferred from *explicitly believed* non-mental premises; we may only tacitly believe the non-mental premises from which our knowledge of other minds is inferred, or—we might only say—based. But we do have an intuitive awareness of this basing; we are aware, for example, that if the right non-mental facts came to be disputed, then our knowledge of the mental facts would then be jeopardized. (An analogy: if I ask you what continent you are in right now, your conclusion will be inferred from, based on, other things you know, though the inference might not be self-consciously made, and its premises may only have been, and may even remain, merely tacitly believed.)

Now, it's of course true that much ordinary knowledge of mental facts about others is inferred from further mental facts, for example when we would justify some claim about another person by saying "She believes p and it's obvious to her that if-p-then-q, so ...", or "He believes anything reported by Fox News, and Fox New reports that p, so ...". I don't deny that. What I assume is only that these further mental facts are themselves inferred from yet further facts. I'm assuming that ordinary knowledge about other minds generally bottoms-out in non-mental facts. The mental facts we know about others are epistemically ultimately founded on non-mental facts we have justification to believe.

We can make the plausibility of the assumption vivid. Imagine a case in which you initially don't even know that some object is a living creature, must less one

with beliefs, desires, pains and perceptual experiences. Imagine watching an alien rock monster slowly begin to stir, and as it becomes animated and interacts with the environment, you come to know that it sees, believes, desires and fears this and that, and feels pain when this and that happens to it. The suggestion is that this sort of case exemplifies a prevalent ordinary ability, one possessed by any armchair philosopher, to draw an inference to mental facts on the basis of nothing but non-mental facts. (See Jackson and Pettit (2004 [1993]: pp.75-78) and Jackson (1996: pp.382-4) for some more support for the claim, our assumption, that we have an ordinary ability to attribute mental facts (Jackson and Pettit only discuss intentional facts) to others purely on the basis of non-mental cues.[2])

As I'm emphasized, ordinary people need not explicitly know which non-mental facts serve as bases for the mental facts we infer. Our ordinary ability to know mental facts only requires that we are aware of our non-mental bases in a tacit way. However, the armchair philosopher can, in principle, reflectively engage in plenty of suppositional reasoning and reasoning by conditional proof in order to infer, and thereby come to explicitly know, a wealth of conditionals. (This is the kind of reasoning modeled by $\supset$-introduction in natural deduction logic.) If you can know q by an inference from a tacitly known premise p, then you can use suppositional reasoning and conditional proof to know the conditional if p then q, and, furthermore, you can know that conditional even if you never *know* but only *suppose* p, inferring q only under the supposition. So, there exist lots of conditionals with non-mental antecedents and mental consequents that can, in principle by armchair reflection alone, become explicitly known, and so the conditionals are apriori in our sense.

(These conditionals are similar to, but not the same as the *application conditionals* defended by Chalmers and Jackson (2001). Chalmers and Jackson believe that, for any concept with an extension, there exist apriori conditionals whose consequents apply the concept, and whose antecedents richly describe some possibility *considered as actual*, that is, a possibility that may be expressed in the form "If it turns out that, in the actual world . . . ". I am claiming, on the other hand, that there are apriori conditionals whose non-mental antecedents are not restricted in that way; my conditionals' antecedents may describe *counterfactual possibilities*, ways things could have been even if they actually are not that way, and their consequents then

2 Jackson and Pettit say that it is "the commonsensical view that raw behavior [they also call it "brute physical movements" a page earlier] is our data for projecting behavior in circumstances from past to future" (p.75), and, that we capture these projections in terms of holistic belief-desire psychology (pp.76-8). While Jackson and Pettit credit Davidson and Dennett for first articulating the broader idea they mean to advert to (see Davidson (1984) and Dennett (1987)), the claim essential to the challenge we're presently developing against aposteriori naturalism can be separated out from any of the more controversial claims Davidson and Dennett made, such as claims about whether interpretation requires attribution of generally true or generally rational beliefs (a topic that is, however, relevant to the view recommended later in the present paper, in section 3).

apply some mental vocabulary. I'll say more later, in sub-section 2.4, about the comparison to Chalmers and Jackson's work.)

2.3 Executing the Challenge: Indefeasible Inferences to the Mental

Now, the existence of a wealth of such apriori conditionals does not quite take us far enough to pose any serious problem for aposteriori naturalism. So far, we only said there are many apriori true conditionals $N \supset M$, for non-mental N and mental M. What the critic of aposteriori naturalism must argue is that we can discover *minimal determiners* of mental facts from the armchair.

The aposteriori naturalist thus might try to defend her position by saying that our ordinary ability to infer mental facts about others on the basis of observed non-mental cues is an ability to rationally draw *fallible* inferences; the non-mental facts that serve as our bases do not *determine* the truth of the mental conclusions we infer. In other words, the aposteriori naturalist might try to defend her view by conceding that we are in a position to know apriori that $N \supset M$, but it is an item of deeply contingent apriori knowledge,[3] thus we are not in a position to know apriori (indeed it is not even true) that $\Box(N \supset M)$, and in particular we are not in a position to know apriori what any *minimal* determiners of M are.

But this isn't enough to halt the challenge to aposteriori naturalism. The critic of aposteriori naturalism can execute her challenge as follows.

Suppose you say our ordinary ability to infer mental fact M from non-mental basis N is, like any typical inductive inference, an ability to make fallible inferences. If fallible inductive inference is the appropriate model here, then you will also agree these inferences of ours are typically *defeasible* in the sense that there are some further non-mental facts, defeaters, such that if you added them to your basis, you would be rationally required to retract your belief in M.[4] Inductive inferences from observed evidence to scientific theories are generally understood to be like that: I rationally infer theory T on the basis of observation O, but I don't think O determines

3 See Hawthorne (2002) for a defense of the existence of deeply contingent apriori knowledge. A contingent truth is defined as deeply contingent when its truth is not semantically guaranteed; truth is semantically guaranteed in the standard meter-stick and inventor-of-the-zip examples of contingent apriori truths from Kripke (1980) and Evans (1979). Hawthorne gives an example, "The Explainer", which is similar to $N \supset M$. The Explainer supposes a large set of scientific observations O, infers the best explanatory theory T under the supposition, and then uses conditional proof to know apriori the material conditional $O \supset T$.

4 I'm using "defeaters" to describe a kind of evidence, and I'm assuming evidence is something represented by subjects. Some epistemologists say there can be defeaters that don't have to be represented by the subject, for example the mere prevalence of fake barns might defeat your true belief that this is a real barn, even if you're clueless about the prevalence of fake barns. I'm not using "defeaters" in a way that includes that.

T, rather I think the inference is fallible, and if I were to learn $O \wedge O'$ I'd no longer rationally believe T. Thus, while the conditional $O \supset T$ is apriori, $(O \wedge O') \supset T$ is not. However, notice that $(O \wedge \neg O') \supset T$ will then be apriori. And if we represent the defeaters for the inference from N to M as N', N'', etc., we can then say these conditionals are apriori: $(N \wedge \neg N') \supset M$, $(N \wedge \neg N' \wedge \neg N'') \supset M$, etc.

Let's pause to illustrate a bit more concretely. First take a simple, stock case of inferring the unobserved from the observed. Suppose you know an urn contains 100 marbles, each either white or black. The urn is opaque, but you are allowed to randomly select and observe marbles as often as you like, as long as you replace each selected marble before selecting again. Many times you observe and replace a marble, always seeing white, never black. This eventually makes it rational to infer all the marbles inside the urn are white, but it does not determine it, and there is an obvious defeater, namely observation of a black marble. Ordinary inferences concerning mental facts about other minds are plausibly similar. The observation of someone stepping out of the way of a campfire makes it rational to infer he believes there is a fire there and desires to avoid pain. Further observation could defeat this attribution, however. For example, if he turned away from a campfire only to run headlong into a forest fire (and thus would seem either to not desire avoiding pain, or to not experience pain); if he turns out to be a marionette (and thus would seem to have no mental states); or, if he is running in a haphazard zigzag that just happens to have him dodging the campfire but tripping up over other obstacles (and thus would seem to have few beliefs about what's in front of him).

Continuing the critical challenge to aposteriori naturalism now, the next claim is that, after a certain point, one can suppose enough negated defeaters to result in an *indefeasible* inference from a non-mental basis to the mental conclusion M.[5] Supposing all the observations we ever make of a randomly selected and replaced marble are observations of a white marble, there are no remaining potential defeaters for the inferred conclusion that all the marbles in the urn are white. (A subtlety: maybe you can't *know* every observation will be of a white marble unless you already know all the marbles in the urn are white, but the only claim here is that you can simply *suppose* it. Suppose also that there are no oracles to inform you there's a black marble that never gets selected, or any other such sources of information about what's in the urn.) Analogously, there is some complete list of potential defeaters for a mental fact inferred from a non-mental basis, though of course I don't explicitly know what goes into the list. But if the negations of all those defeaters were

5 Some epistemologists think no inference is defeasible because there is always some defeater in the form of a (justified) belief that one's reliable guru says such-and-such. But, here we are only discussing inferences from exclusively non-mental, and so non-intentional, bases, so we can ignore the guru case. Our concern with indefeasibility is concern over whether the non-mental basis of the inference can be expanded so as to make the mental conclusion M no longer inferable.

supposed, then the inference from them and N to M would become indefeasible. (See Schiffer (1993: p.96) for an endorsement of the claim of the present paragraph.[6])

The final step of the challenge to aposteriori naturalism is to pose the question: is the *indefeasible* inference plausibly still *fallible*, or does enlarging the basis by including the negations of all defeaters eventually produce a *determiner*? At this stage, the typical inductive case, such as the urn illustration, becomes crucially different from the mental case. No number of observations of replaced marbles determines the color of the next marble, much less the contents of the urn. We can all agree that, in general, the facts about the observed do not *determine* the facts about the unobserved; indeed, you could even take the *totality* of facts about the past, and they won't determine any contingent facts about the immediate future. The crucial difference, though, with the naturalist about the mental, to whom our discussion is addressed, is that she *does* think there exist non-mental determiners for each mental fact. It is the aposteriori naturalist's naturalism that dialectically allows us to raise a special challenge for her. *Given* that she believes there exist non-mental determiners for mental facts, how can the naturalist deny that *this* is one? Indeed, how can she deny that as soon as the inclusion of negated defeaters turns the inference into an indefeasible one, the antecedent describes a *minimal* determiner of the intentional conclusion? The aposteriori naturalist has no apparent grounds for denying that there exists an *apriori*, *necessary* conditional of the form $\Box((N \wedge \neg N' \wedge \neg N'' \wedge \ldots) \supset M)$, whose antecedent will be a *minimal* determiner as long as it only includes N and negations of defeaters.

It is crucial to appreciate that our challenge here does not proceed by simply arguing in favor of apriori naturalism. A non-naturalist will be unbothered by what we've said here. We are raising a challenge that arises for someone who accepts naturalism, someone who believes there are minimal determiners out there, but who denies that they can be discovered from the armchair. In effect, our challenge asks: *if* you are a naturalist, then how can you not also be an apriori naturalist?

6 Here is a partial quote: "[S]uppose Regina is what she appears to be: a paradigm human being with a well-functioning brain and nervous system, the offspring of paradigm human beings and genetically similar to them and all other paradigm human beings. And suppose that Regaina's sense organs are in top-notch condition, and that the cited behavior coheres in expected ways with all her past, present and future actual and counterfactual behavior. Then I submit that nothing could show that Regina didn't believe that there was a dog before her: not anything that might be discovered about her brain; not the realization that no nonintentional fact is explained by propositional attitudes that isn't also explained by physical facts; and not the realization that the belief relation can't be identified with any physical or topic-neutral relation. . . . To be sure, this is no *argument* that I have just sketched. But it is the intuitive view, I dare say, whose status seems secure in the absence of a compelling argument against it. And how could there be a truly compelling argument? What abstruse philosophical premise could have a plausibility greater than that of the assumption that, come what may, human beings like you and I have beliefs and desires?"

So, our challenge to the aposteriori naturalist can be understood as a demand for an explanation why a certain very obvious candidate for being a minimal determiner of a mental fact is not one. Why aren't the non-mental bases of indefeasible inferences to mental conclusions minimal determiners? That is the fundamental hard question that the aposteriori naturalist has to give a satisfying reply to if the view is to be believable. It is not strictly inconsistent or incoherent to resist the challenge by digging in and insisting that mental inferences are always fallible, or even insisting that they are always defeasible, or perhaps even insisting that we cannot make our ordinary inferences to mental conclusions when we merely *suppose* the non-mental premises (rather than *know* those same premises). But these responses cannot stand on their own; they need to be offered as part of a well-supported view. On their own, these responses fail to address the challenge as a demand for explanation: these *ad hoc* responses just call out for more explanation. Without some explanation or some good reasons in support of such apparently *ad hoc* claims, aposteriori naturalism cannot be justified.

2.4 Unattractive Responses to the Challenge

Before turning to the response to the challenge that I want to recommend to the aposteriori naturalist, I want to mention a few other ways of responding, and say why I don't recommend them.

(1) *McDowell's View.* McDowell (1982) offers one way to reject my challenge's assumption, and thus evade the challenge. He says our knowledge of phenomenal facts about other minds is typically non-inferential, and thus, in particular, not based on any non-mental facts. Following an interpretation of Wittgenstein (1953), McDowell suggests we know that, say, Jones is in pain, not on the basis of any external behavioral criteria or any premise at all; rather, we know it directly.

McDowell's view does not offer an attractive way for aposteriori naturalists to evade my challenge. I won't bother to make any elaborate argument here. The view is simply highly implausible, and there are other lines of response that are much more attractive.

(2) *Horgan's View.* Should the aposteriori naturalist respond by saying that a minimal determiner, constructed out of many negated defeaters in the way described above, will be too intractable to qualify as knowable, much less knowable apriori?

This way of responding to my challenge may take inspiration from Horgan's view. Horgan claims both that "mental properties and facts are supervenient on physical properties and facts [i.e. naturalism]", and that "this supervenience thesis could well be true even if there is no way to tractably specify the non-mental conditions that suffice for mental phenomena." (Horgan (1994: p.477); see also Horgan (1993)). (The view is similar to, and perhaps inspired by, McGinn's famous mysterianism;

see McGinn (1993, 1999).) The primary support for Horgan's view, the reason for expecting minimal determiners to be so unwieldy, is that counterexamples were repeatedly raised refuting past attempts to state them.[7]

I see two reasons why it is not the most attractive response to my challenge to invoke Horgan's view and to say that minimal determiners are too intractable to qualify as knowable, much less knowable apriori. The first reason is that in order to respond to the challenge by invoking Horgan's view, we would need to adopt a restricted interpretation of the apriori, one that excludes a fact from being apriori knowable just because it is too complex for any ordinary, unaided human mind to comprehend, and this seems to me a non-standard and *ad hoc* restriction on the apriori. Are arithmetical truths no longer apriori when they cross a threshold of complexity putting them beyond ordinary comprehensibility? (Horgan himself did not claim anything about apriority or knowability. We are considering a response to my challenge that is inspired by, but says slightly more than, Horgan's view.)

The second reason turns on some subtle features of the dynamics of rational belief revision. One motivation for aposteriori naturalism is that it offers us a way we can retain our high confidence in naturalism even in the face of the historical parade of failed attempts to locate minimal determiners from the armchair. However, to the extent that our confidence in naturalism was not already confidence specifically in aposteriori naturalism *antecedently* to learning of the historical failures, we must to some extent decrease our confidence in naturalism. If any of your antecedent confidence in naturalism was confidence in apriori naturalism, that is, if you antecedently had some confidence (as Lewis, Stalnaker, Fodor, and others presumably did) that philosophers could locate minimal determiners of the mental from the armchair, then evidence against this is, for you, evidence against naturalism, and you cannot restore any such lost confidence in naturalism by reapportioning your confidence from apriori naturalism to aposteriori naturalism. (This is a subtle, but, I take it, intrinsically plausible point about the dynamics of rational belief revision, at least upon reflection. It could also be proved, for example, in a Bayesian model.[8] Let

7 Kriegel (2009: p.461) states the associated point concerning McGinn's mysterianism clearly: "The primary motivation for mysterianism may be captured by an inductive inference from the evident and flagrant inadequacy of all known theories of consciousness, coupled with the aforementioned sentiment that the inadequacy is unusually profound." Horgan says just the same thing about his own view: "A number of philosophers (e.g. Baker (1987), Stich (1992), Tye (1992)) have recently argued that projects for 'naturalizing' mentality probably cannot succeed. For one thing, counterexamples keep surfacing, and the accumulating inductive evidence suggests that they always will." (Horgan (1994: p.477); also see Horgan (1993: pp.579-80).)

8 Consider a three-way partition of logical possibilities: aposteriori naturalism, apriori naturalism, and non-naturalism. It's a feature of Bayesian Confirmation Theory that if you had non-zero confidence in apriori naturalism and in non-naturalism, and you lose all confidence in apriori naturalism (i.e. you conditionalize on its negation), then your confidence in naturalism itself (i.e. the

me also note here that Horgan's view is not plausibly supported *just* on the premise that the determination of mental facts by non-mental facts proceeds *holistically*, i.e. many mental facts must get determined for even one to be. Lewis, Stalnaker, Fodor and others were aware of and highly sensitive to the issue of holism—Fodor, of course, rejecting it—even while confident of apriori naturalism: holism alone does not support intractability.) Horgan's view enjoys little support antecedently to, and much support in *response* to, the historical failure. If the aposteriori naturalist co-opts Horgan's view, she is thus left with a less interesting, less motivated view. If we wish to retain as much as we can of our high confidence in naturalism even in the face of the historical failures to naturalize the mental from the armchair, Horgan's view is less attractive than a form of aposteriori naturalism that can be supported antecedently to the historical failure. If there is an alternative way for the aposteriori naturalist to respond to my challenge, other things equal, it will be more attractive.

(3) *Chalmers and Jackson's View.* Another potential way to respond to the challenge I've raised is to say that it is effectively just the same as a version of a challenge that Chalmers and Jackson (2001) raised (against the possibility of "aposteriori reductive explanation"), and the aposteriori naturalist should resist my challenge in the same way, whatever that way may be, that one should resist the view of Chalmers and Jackson. Let me give the relevant claims from the view of Chalmers and Jackson (2001), and then explain why I do not think my challenge is best resisted just as a by-product of resistance to their challenge.

As I briefly, and parenthetically, mentioned earlier (at the end of sub-section 2.2), what Chalmers and Jackson claim is similar to, but crucially different from, what I've claimed. They make several claims about facts that concern natural kinds, such as water, that are similar to what I've claimed about mental facts. They think we can apriori know certain conditionals linking "non-water" facts to "water facts", as long as the antecedent describes a possibility considered as actual, also called an "epistemic possibility", a possibility we may express in the form "If it turns out that, in the actual world . . . ". For example, they argue we can apriori know a conditional roughly like: if it turns out that, in the actual world, my cup is filled with the clear, odorless substance found in local lakes and oceans, then water is in my cup. Such an apriori conditional, if it exists, would also plausibly be necessary, and thus its antecedent would determine its consequent. (Although the conditional's antecedent

disjunction of the apriori and aposteriori sub-varieties) must go down. The result it easiest to see if you use van Fraassen's Muddy Venn Diagram model of Bayesianism; see van Fraassen (1989). Confidence is the fraction of the Venn diagram's total mud sitting on the various spaces representing various propositions. To conditionalize on the negation of apriori naturalism, just wipe off the mud (confidence) that was on the space for apriori naturalism. Obviously, the mud remaining on naturalism, which is now all on the space for aposteriori naturalism, will be a smaller fraction of the total remaining mud.

does not describe a literally counterfactual possibility, it's still true that *it could not have been* that the antecedent holds while the consequent fails, which is just how I defined "determines".)

Since their view is slightly similar to mine, one might initially think that a good strategy for resisting Chalmers and Jackson's view will also be a good strategy for resisting my challenge to the aposteriori naturalist. I don't think this is right, though.

To begin with, my challenge aims to support certain claims about mental facts that are much stronger, and (as I'll explain) enjoy a different source of supporting intuitions, than the corresponding claims Chalmers and Jackson aim to support about natural kinds: my challenge asks why armchair philosophers can't discover, for the mental facts, non-mental determiners which may be counterfactual possibilities, not epistemic possibilities. My challenge involves the claim that you can draw an indefeasible inference to a mental conclusion about someone just from a rich supposition about their behavior, environment, and other non-mental facts, but there was no claim that your supposition must the sort that can be expressed using "it turns out, in the actual world, that the non-mental facts [are this way or that]". So, my challenge concerned why we cannot, from the armchair, discover determiners of a sort much more interesting than the sort that Chalmers and Jackson claim is discoverable from the armchair.

(Counterfactual possibilities have potential to provide far more interesting determiners than epistemic possibilities because, plausibly, only they can provide possible non-mental facts that not merely determine, but, in some more substantive way, ground or reduce the mental, that is, provide non-mental facts in virtue of which the mental facts hold. Again, my challenge targeted the less intrinsically interesting determination, or "supervenience", thesis because all the more interesting views entail it, and thus are hostage to it.)

If, now, we just consider why Chalmers and Jackson rightly refrain from claiming that there exist apriori, necessary conditionals linking a *counterfactual* non-water fact to water's presence, this will illustrate why my challenge draws support from intuitions quite different from anything supporting their view. Even supposing for the moment, with Chalmers and Jackson, there is an armchair ability to infer the presence of water on the basis of certain (known or supposed) non-water facts (about being clear, odorless, etc.), these non-water facts only concern superficial marks which, even Chalmers and Jackson agree, do not attach to water throughout the whole space of counterfactual possibilities. This is just one of the intuitive points made by *Naming and Necessity* and Putnam's Twin Earth example. The only non-trivial properties that attach to water throughout all counterfactual possibilities are properties that we could only know water to have aposteriori. In other words, it is intuitive that, in some counterfactual possibilities, the superficial marks from which the armchair reasoner can infer the presence of water are false evidence (for

example, if it is "twin water"). By contrast, we can ask: is it similarly intuitive that, in certain counterfactual possibilities, those non-mental marks that allow an armchair reasoner to infer, even to indefeasibly infer, a mental conclusion are false evidence? Is it intuitive that mental phenomena are only contingently linked to the non-mental properties from which armchair reasoners are prepared to indefeasibly infer their presence? My challenge draws its strength from the fact that it is not intuitive that the mental, in this way, resembles natural kinds like water or any other standard Kripkean cases of aposteriori necessitation. There is, I claim, a genuine question, a question that my challenge demands the aposteriori naturalist give a well-supported answer to, why the non-mental fact from which an armchair reasoner can indefeasibly infer a mental fact is not also a determiner of that mental fact.

So, Chalmers and Jackson only claim, and rightly only claim, there are apriori, necessary conditionals formulated in terms of epistemic possibilities. Even here, though, there are good reasons to doubt their claim, reasons which are not equally compelling against any claim involved in my challenge. Many philosophers, among them aposteriori naturalists like Block and Stalnaker (1999), do wish to resist Chalmers and Jackson's claims that we can apriori know conditionals whose antecedents describe epistemic possibilities without using "water", and whose consequents states that water is present here or there. (Also see Byrne and Pryor (2006) and Schroeter (2006) for opposition to Chalmers and Jackson.) No matter which side of the debate is ultimately right, the opponents of Chalmers and Jackson can claim a good deal of initial plausibility to their line of response. We can illustrate the plausibility of the idea behind the response more easily if we use a less commonplace natural kind. Suppose Joe read about a natural kind, grog, in a newspaper report that mentions grog is being tested as having possible medical benefits. Suppose Joe forgets where he heard about grog, only recalling it's a substance being tested for medical benefits. One of the common morals of *Naming and Necessity* is that Joe might be able to talk about, and think about, grog while lacking any uniquely identifying description (in terms omitting "grog"). It seems fairly common that natural kinds—and the same goes for artifact kinds, and named individuals—are typically talked about and thought about by ordinary people who cannot, in any other terms, uniquely describe what they are talking or thinking about, and thus cannot apriori know the conditionals Chalmers and Jackson say they can know.[9]

But, I think it is clear, we cannot generate a parallel, equally plausible line

9 One strategy sometimes explored for defending the view that we do possess uniquely identifying descriptions of natural kinds and named individuals is to consider metalinguistic descriptions, such as "the thing that whoever taught me the word 'grog' had in mind". Aside from direct objections to this strategy (see Byrne and Pryor (2006), for example), it would obviously not allow us to know conditionals linking the non-mental to the mental, since the description uses mental, specifically intentional, vocabulary.

of response to my challenge. It is simply not equally plausible that we can talk about and think about mental phenomena (conceived as such) without being able to uniquely identify the mental on the basis of non-mental premises. Our talk and thought about mental phenomena (conceived as such) is not like our talk and thought about natural kinds, artifact kinds, or named individuals (conceived as such), and our grasp of mental concepts is not plausibly epistemically divorced from their non-mental application conditions in the way we plausibly can divorce concepts of natural kinds, artifact kinds, and named individuals (conceived as such) from their application conditions.

(There are various likely explanations of this difference between mental concepts and concepts of natural kinds and other things. One tempting line of explanation appeals to the idea that mental concepts are among our innate concepts, while concepts of natural kinds, artifacts and many named individuals are acquired. Support for this draws from recent cognitive science; see, for example, Carey (2009: especially chapter 5). But I won't speculate any further on this, since my purposes in this paper don't require it.)

So, the aposteriori naturalist view about the mental has a burden to answer our explanatory challenge, a burden not equally shared by Chalmers and Jackson's opponents, who advocate the view that it is entirely aposteriori which non-water facts determine water facts, or which non-grog facts determine the grog facts. If the aposteriori naturalist wants to defend the analogy between, on the one hand, the epistemic status of the determination of mental facts by their non-mental minimal determiners, and, on the other hand, the aposteriori determination of natural kinds by their own minimal determiners, then the aposteriori naturalist will still owe us some further explanation of how the former can be aposteriori.

3 The Recommended Response: Knowledge of Other Minds Based on Self-Knowledge; Self-Knowledge Not Based on Determiners

In fact, I think the prospects are good that the aposteriori naturalist can give such an explanation, and in this final section I elaborate the way I recommend she give such an explanation. I'll recommend that the aposteriori naturalist adopt certain substantive views of the epistemology of mental phenomena, views that are not simply imported from the case of natural kinds, artifact kinds or named individuals.

I'll claim that if aposteriori naturalism is supplemented with these further views, then the demand for explanation posed at the end of our challenge has a good answer. What I'll say won't suffice on its own to prove that aposteriori naturalism is true; it will show that if two major philosophical and psychological debates resolve in the likely ways, then aposteriori naturalism not only has its answer to the challenge, but

enjoys the independent, positive support that its motivations require (in particular, support that is antecedent to, not dependent on, learning of philosophers' historical difficulty at locating minimal determiners). Thus, my final conclusion will be conditional: aposteriori naturalism's success is contingent on favorable, but also likely, outcomes of certain larger debates.

The challenge of the last section concerned our epistemic access to mental facts about other minds. The first of the two large debates I have in mind concerns our epistemic access to mental facts about our *own* minds. The second debate concerns the *role* of knowledge of our own minds in our ordinary ability to know other minds. I'll discuss them in turn.

3.1 Recent Views on Self-Knowledge

The current debate over how we have knowledge of our own minds is one of the liveliest in contemporary philosophy and psychology. It would naturally be unwise to suggest the aposteriori naturalist place a bet on any one of the many competing views. But my suggestion is not that the aposteriori naturalist bet on a view, but *against* one particular view. This is because nearly all of the current views of how we have knowledge of our own mental states entail a common consequence, which is this: knowledge of a mental fact about oneself, even when it is indefeasible, *is not based* on non-mental facts that *determine* the known mental fact. That consequence, common to nearly all the competing views, is all the aposteriori naturalist will need.

To illustrate, let me point out how the consequence follows on most mainstream views of self-knowledge.

Self-scanning views have the consequence trivially. These views are traditionally associated with Locke, and contemporary advocates include Armstrong (1963), Lycan (1996), Nichols and Stich (2003), and Goldman (2006). According to such views, we non-inferentially detect our mental states via a mechanism similar to our perceptual mechanisms for learning about our external environments. If self-knowledge is never based on other beliefs, it is obviously never based on beliefs in non-mental facts that determine the known mental fact about oneself.

The self-scanning view may seems to better account for self-knowledge of phenomenal facts than of intentional facts. At the same time, other views, such as the one I'll mention next, may seem to better account for self-knowledge of intentional facts than of phenomenal facts. So, we should keep in mind that the views could be held in various sorts of mixed combination.

Transparency views of self-knowledge, in particular as an account of our knowledge of our intentional states, are more popular than self-scanning views these days. Generally credited to an insight from Evans (1982: pp.225-7), its contemporary advocates include Dretske (1995: chapter 2), Gordon (1995, 2007), Gallois (1996),

Peacocke (1998), Tye (2000: chapter 3), Byrne (2005, 2011), Setiya (2011) and Silins (2012). Let me just describe how a transparency view aims to explain self-knowledge of our own beliefs. According to a transparency view, we may believe something of the form *I believe that* ***p*** on the basis of the corresponding belief or judgment with content of the form ***p***.[10] This basic idea enjoys an intuitive appeal that Evans brought out with his famous example: if someone asks me whether I believe there will be a third world war, I can, and perhaps can only, answer this question by attending to the question whether there will be a third world war.

While the view has some intuitive appeal, a host of apparent obstacles arise once we ask how the basic idea could be generalized to account for our knowledge of non-attitudinal mental states (such a pains), our knowledge of intentional attitudes other than belief (such as desires, suppositions, intentions, fears, even disbeliefs), and our knowledge of the absence of belief (as when you know you are positively uncertain about something, or when you know you simply haven't bothered to take any attitude yet). The primary aim of a number of the previously cited works is to show how the transparency view can handle, indeed is best suited to handle, each of these apparently harder cases of self-knowledge.

My purpose here isn't to say anything new in defense of transparency views or any other views of self-knowledge. My purpose is only to point out that this popular approach to explaining self-knowledge has the consequence that will be of use to the aposteriori naturalist: on a transparency view, your knowledge of an intentional fact about yourself, *I believe that* ***p***, is based (plausibly even indefeasibly based) on a believed or judged premise, ***p***, and that premise, which is typically non-mental, *obviously does not determine the known intentional conclusion*. Indeed, a transparency view even has the consequence that, although knowledge of our own mental states is based on certain non-mental premises, we cannot come to know the corresponding conditionals for this inference. That is, while advocates of a transparency view say that we may infer *I believe that* ***p*** from ***p***, their view requires

10 The cited authors disagree over whether or not such an introspective belief, although it is *based* on the belief that ***p***, is appropriately described as *inferred* from the belief that ***p***. Gallois (1996: pp.46-7) and Byrne (2011: p.203) consider it an inference; Gordon doesn't say one way or another; Dretske (1995: pp.60-1), Peacocke (1998: pp.73, 90), Tye (2000: pp.52-3), Setiya (2011: p.184) and Silins (2012: introduction) each say or strongly suggest they think this transition is not a case, or at least not an ordinary case, of "inference" or "reasoning". One source of hesitation to call it an inference is that self-knowledge is, in some intuitive sense, especially "direct" or "immediate". Another point, which Dretske notes, is that the belief of the form *I believe that* ***p*** is indefeasible and justified regardless of the truth of the belief that ***p***. ("[I]f this is inferential knowledge, it is a very unusual form", Dretske (1995: p.61).) Nonetheless, the cited authors agree there is a mental process here whose input is a belief (or judgment, as Silins insists) about the external world and whose output, formed on the *basis* of that input, is a belief that one is representing the world as such. That was my general characterization of the transparency view.

the peculiar proviso that we may not apply conditional proof to similar suppositional reasoning: we may not infer *I believe that* $\boldsymbol{p}$ under the supposition that $\boldsymbol{p}$, and so we may not apply conditional proof to arrive at knowledge of what is typically a false conditional, $\boldsymbol{p} \supset$ *I believe that* $\boldsymbol{p}$. (Recall that, at the end of sub-section 2.3, I said it would be *ad hoc* to just insist that ordinary reasoning from a known non-mental premise to a mental conclusion cannot take place when the premise is merely a supposition: here we are putting that claim, *ah hoc* in isolation, into an overall view that has received independent defense.)

Finally, *constitutivist* views, just like self-scanning and transparency views, also entail the consequence that we have knowledge of mental facts about ourselves that isn't based on knowledge of non-mental facts that determine those mental facts. Advocates of these views include Shoemaker (1996, 2009), Boyle (2011), and Schwitzgebel (2011). (Moran (2001) endorses either a transparency view or a constitutivist view. I am inclined to interpret him as endorsing a transparency view, but Boyle (2011) argues that he is better interpreted as a constitutivist. Shoemaker (2009) suggests Peacocke (1998) might be interpretated as a constitutivist, though the matter is unclear.)

The general idea behind such views is that being in a mental state constitutes self-knowledge of that mental state, though this is usually qualified in various ways. For example, the way constitutivist views explain our self-knowledge of beliefs is by claiming that it is constitutive (or at least partly constitutive) of believing (or at least of rationally believing) that $\boldsymbol{p}$ that you also believe (or at least are disposed to believe) that you believe that $\boldsymbol{p}$. Boyle (2011: p.228) offers a succinct statement of his own view, "in the normal and basic case, believing P and knowing oneself to believe P are not two cognitive states; they are two aspects of one cognitive state—the state, as we might put it, of knowingly believing P." (There's a clear affinity here with the transparency account: where the transparency theorist posits a *basing* relation, the constitutivist replaces it with a (partial) *constitution* relation.) Other authors, especially Shoemaker, have defended versions of the view that extend it to cover a larger range of mental states, including phenomenal states such as pain. As with transparency views, one might think constitutivist views are better suited to explain our self-knowledge of certain states than others; so, again, one might choose to hold some mixture of a constitutivist view with other views of self-knowledge. My aim is not to offer any novel defense of this or that particular constitutivist view, only to point out that, on such a view, the consequence we're after clearly follows: your knowledge of a mental fact about yourself, even when it is indefeasible knowledge, will not be based on a determining non-mental fact.

There is only one contemporary view of introspection that lacks that consequence: *skepticism*. Specifically, the view in question is skepticism about our having *peculiar* or *special* access to our own mental states, where such access is a form of

access that does not (by itself) afford access to others' mental states.[11] It's extremely intuitive that we have such access. The contemporary skeptic of such views is Peter Carruthers.[12] But even he restricts his skepticism to just a few types of intentional states, specifically judgments and decisions, allowing that we have special access to other intentional states (in particular, so-called inner speech) as well as to our phenomenal states. Furthermore, even with regard to judgments and decisions, Carruthers admits there is a strong intuition that special access is real, admitting that philosophers are "virtually united in thinking there is introspection for judgments and decisions, just as there is for perception and imagistic states."[13] (He uses "introspection" in a technical sense to mean special access.[14])

In a similar vein to Carruthers's skepticism, some psychologists have argued in favor of various claims concerning the parity between our access to other minds and our access to our own minds. Gopnik (1993) is a famous example, and one that specifically focuses on our knowledge of the contents of our intentional states.[15] See Schwitzgebel (2010: section 2.1.3) for a review of this and other psychological literature, concluding that "it is probably impossible to sustain a view on which there is complete parity between first- and third-person mental state attributions. There must be some sort of introspective, or at least uniquely first-person, process."

As I said, the bet I recommend to the aposteriori naturalist is a likely winning one: the bet is against skepticism about special access to our own minds. As long as one or another of the special access views wins, the aposteriori naturalist can help herself to the consequence that we can have some knowledge, even indefeasible knowledge, of mental facts that is not based on knowledge of non-mental facts that determine those known mental facts.

Now, special access to knowledge of one's own mental states obviously doesn't quite get us the final result we're after. We are seeking to answer the challenge of the last section, which asked: how can knowledge of non-mental facts serve as an indefeasible basis for knowledge of a mental conclusion about someone else, and yet those non-mental facts do not determine (and thus do not minimally determine) that mental fact? What we have so far is only a story (one story or another) about

11 See Shoemaker (1993) and Byrne (2011) for elaboration of the notion of such special access.

12 See Carruthers (2009, 2010, 2011).

13 See Carruthers (2010: p.82).

14 As he defines it earlier, "introspective access to our own mental states is epistemically quite different—in kind, and not just in degree—from the access that we have to the thoughts and perceptions of other people". Carruthers (2010: p.76-7).

15 See Nisbett and Wilson (1977) for an even more famous example, though one of less obvious direct relevance. They claim people often confabulate the *reasons* that caused them to choose one pair of socks rather than another. It would take some argument to say how such a claim about reasons relates to our present discussion, which is concerned with intentional states like belief and desire. See Carruthers (2009, 2010) for an attempt.

how you could have indefeasible knowledge of a mental fact about yourself without it being based on knowledge of determining non-mental facts. What we still need to see, then, is how to extend this to the third-personal case.

3.2 Recent Views on Access to Other Minds

The second bet I thus recommend to the aposteriori naturalist is, naturally, that there is an ineliminable role played by our special access to our own mental states in our access to all mental facts, including mental facts about other minds.

The suggestion here doesn't require us to make the implausibly simplistic claim that a mental fact about another mind, say *Jones believes* $\boldsymbol{p}$, must be self-consciously based on explicitly held beliefs about oneself, such as *I believe* $\boldsymbol{p}$. Rather, the suggestion need only be that our mechanism for acquiring self-knowledge is deployed at some stage within a more complex inferential mechanism for acquiring knowledge of other minds; we almost certainly have no more than tacit knowledge of the involvement of self-knowledge in our inferences about other minds.

The suggested view of how we know facts about other minds will easily be recognized as a commitment of the so-called simulation (or better, simulation-plus-projection[16]) theory, a theory whose development can be traced back to Quine (1960)[17] (and also traced, though much more loosely, to the traditional suggestion that the skeptical problem of others minds is best solved by an analogical inference[18]). Grandy (1973) developed the suggestion in the form of his principle of humanity, which required that in interpreting others, "the imputed pattern of relations among beliefs, desires, and the world be as similar to our own as possible." (p.443) Grandy argued for his principle on the basis of a picture on which interpretation of others involves a process of simulation. "[W]e use ourselves in order to arrive at the prediction [concerning another's behavior]: we consider what we should do if we had the relevant beliefs and desires. Whether our simulation of the other person is successful will depend heavily on the similarity of his belief-and-desire network to our own." (p.443) Later Heal (1986), Gordon (1986), and Goldman (1989), the latter two explicitly following Grandy, inaugurated the contemporary discussion with their defenses of (slightly different versions of) simulation theory. The common

16 See Goldman (2006: p.40).

17 See Quine (1960: p.219), where, in a now famous passage, he says that "in indirect quotation we project ourselves into what, from his remarks and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned." Also see the following paragraphs.

18 See, for example, Russell (1948: chapter 6.8) and Hyslop and Jackson (1972). Defenders of the analogical inference are more concerned to argue that we can know that others are not zombies, thus staving off skepticism; their project is not to develop any more detailed view about exactly what conclusions may be inferred from what premises.

commitment of their views which I want to highlight here is only the claim that our mechanism for acquiring knowledge of other minds always draws from the outputs of psychological mechanisms or "routines" for acquiring self-knowledge.[19]

Now, if that's a commitment of simulation theory, does that mean I am recommending betting in favor of simulation theory and against its opponents, in effect recommending taking sides in a live debate? Actually, I'm not. In fact, many who were, historically, the main critics of simulation theory will today agree that central aspects of our ability to know the minds of others involve simulation. The best example of this (and it's hardly the only example[20]) is the collaborative work of Nichols and Stich, who have contributed numerous papers to the contemporary debate, beginning with Stich and Nichols (1992), a strongly worded critique of simulation theory, and eventually leading to their book, Nichols and Stich (2003), which offers an eclectic positive theory that explicitly incorporates fundamental aspects of simulation theory.[21]

We thus find, even when comparing the views of those like Nichols and Stich (2003) and Goldman (2006), who strongly disagree on a range of fundamental aspects of how we infer mental facts about other minds, a surprising point of agreement regarding the specific claim that I want to recommend to the aposteriori naturalist. The recommended claim, again, is that we make use of our mechanism for self-knowledge within our mechanism for knowing mental facts about other minds. The claim is endorsed by both Nichols and Stich (2003) and Goldman (2006). Both propose that interpretation of another mind begins with a stage in which the interpreter makes a default attribution to the target of some very broad swath of her own mental states, in particular her own beliefs.[22]

(What if I am attributing beliefs to someone in a counterfactual scenario, the sort of case that was featured in the challenge to aposteriori naturalism? Is there a worry about how will I know what I believe in the counterfactual case? No, the method

19 Many of the influential papers that advanced the debate between advocates and opponents of the simulation view can be found in Davies and Stone (1995*a*,*b*), and Carruthers and Smith (1996). For a recent (and largely sympathetic) discussion among psychologists, see Malle and Hodges (2005).

20 Another prominent example of a major critic of the simulation theory gradually evolving into a defender of some of the simulation theory's core claims is the psychologist Josef Perner. See Gordon (2009: section 2) on this and other recent history. Perner and Küberger (2005) ends on this telling sentence: "Many of the contributions to this volume [Malle and Hodges (2005)] bear witness to the fact that an increasing number of researchers from different areas find simulation a useful concept for explaining people's ability to understand other minds."

21 See Nichols and Stich (2003: pp.101, 132ff, and 212) for their own comments on the history of the debate and their own convergence toward an eclectic view. On p.134, they state their wish that the word "simulation" be retired from the literature for being too vague, but their motives may be suspect given the shift in their views over time. It doesn't seem unjustifiable to call their final view a simulation theory. Williamson (2007: p.148) and Carruthers (2013*b*: p.160) call it just that.

22 See Nichols and Stich (2003: pp.66-7, 85, 92, 106, 140-1) and Goldman (2006: sections 2.5 and 7.7).

of attribution involves a default attribution of the attributor's own beliefs, her own *actual* beliefs, not her beliefs in the counterfactual scenario. Otherwise, attributions to one's counterfactual self would become absurdly trivial.)

Of course, this much leaves most of the story about how we *correctly* determine what another person believes untold; interpreters are still left with the delicate and difficult task of then modifying this initial default stock of beliefs into a reliable representation of what someone else believes. Nichols and Stich (2003) and Goldman (2006) go on to develop things in divergent ways. But we don't need to adjudicate debates over any of these aspects of how we know other minds, intrinsically important and interesting as they may be.[23] What's important for our purposes here is only the remarkable fact that such historical rivals have agreed on this central, foundational role for self-knowledge in our complex overall mechanism for attributing mental states to others.[24]

3.3 Answering the Challenge to Aposteriori Naturalism

Now, finally, we can step back and consider the significance of these recommendations for how to supplement aposteriori naturalism. We now have a way of answering the demand for an explanation issued at the end of our presentation of the challenge in sub-section 2.3. The challenge began with the preliminary point that we have an armchair ability to infer substantive mental facts from non-mental facts, as illustrated by the Twin Earth inference, with no analogous ability to infer water facts or Hesperus facts from the associated determining H_2O facts or Phosphorus facts. I then introduced the intuitive assumption that we have an ordinary and general armchair ability to know mental facts about other minds solely on the basis of non-mental

23 We also don't need to debate what epistemological theory explains why our reasoning about other minds is justified. Our purposes here only require us to presuppose, with common sense, that it *is* justified, and ordinarily yields knowledge. Likewise for our earlier discussion of self-knowledge.

24 Carruthers (2011, 2013*a*) is sympathetic to much of the model the Nichols and Stich develop, and Carruthers himself observes that the mind reading literature has converged toward including an essential role for simulation in minding (Carruthers (2011: pp.225, 230)). However, Carruthers wishes to insist on a distinction between (a) views where an attribution a belief that p to another person merely by "drawing on" one's own belief that p, and (b) views where attributions to others are "based on" or "depend on" introspection of one's own belief that p. (Carruthers (2013*a*: pp.143-4)) He says Nichols and Stich (2003) only accept the former, and Carruthers endorses this view, and he says Goldman (2006) accepts the latter, wrongly in Carruthers' view. But, regardless of whatever Carruthers's intended distinction is, I believe my purposes here are fully met even by the view that Carruthers does wish to accept, as when he says "For the standard way of predicting what someone with a given belief will think or do is to assume that belief for oneself, and then to reason on one's own behalf (with suitable adjustments for the context, and for other differences from the target), attributing the result to the other person. (This is the core truth in simulationist models of mindreading; see Nichols and Stich (2003); Carruthers (2011, 2013*a*).)". Carruthers (2013*b*: p.160)

facts. The challenge continued by arguing that a suitably elaborated non-mental basis can, in principle, serve as an indefeasible basis for a mental conclusion. The challenge finally posed the question: how could a naturalist plausibly argue that such an indefeasible basis isn't a determiner of the mental fact we infer from it?

If the bets I've recommended are good, the aposteriori naturalist has her answer to that final question, an answer that concedes both the preliminary point about Twin Earth and the intuitive assumption that we can and do infer the mental solely from the non-mental. Knowledge of a mental fact about another mind is always epistemically dependent on access to mental facts about oneself; and mental facts about oneself, even if known indefeasibly, and even if known solely on the basis of non-mental facts, are not known on the basis of determining non-mental facts.

To illustrate, let's reconsider the example of Toscar, the duplicate of a normal earthling living in a duplicate of our environment, except with no H_2O. We know Toscar doesn't share our belief that there is H_2O in his cup, but he does share our belief that grass is green. And, indeed, we can infer such knowledge solely from non-mental facts. But, if the views of self-knowledge and knowledge of other minds that I've just been outlining are correct, then the character of this inference is very special. For a concrete illustration, let's presuppose the transparency theory of self-knowledge. How do I know that Toscar believes grass is green? I infer this from the non-mental premise that *grass is green*, which allows me, by transparency, to infer that *I believe grass is green*, and then I couple that sub-conclusion with a second non-mental premise that *Toscar and I are sufficiently similarly situated with respect to the proposition that grass is green*, which then allows me to finally conclude that Toscar believes grass is green. Actually, I'm not at all sure how to explicitly articulate and make precise that second premise concerning Toscar's and my similarity, but it doesn't matter; tacit knowledge is all I need. Perhaps philosophical reflection on cases would allow me to make that tacit knowledge explicit: for example, when I consider whether Toscar and I are sufficiently similar to share beliefs about water, I conclude we are not, and I decline to attribute to Toscar the belief that water is in his cup. Cases like this might allow me to gradually sharpen my explicit knowledge of the conditions in which I'm willing to attribute various mental states. But again, explicit knowledge, even the possibility of explicit knowledge, is not important to the aposteriori naturalist giving this response to our challenge. What's important is just that the non-mental bases of this sort of reasoning patently do not determine its mental conclusion. The non-mental (conjunctive) fact that grass is green and Toscar and I are similar does not determine the mental fact that Toscar believes grass is green. It perfectly well could have been that that non-mental fact is true while that mental conclusion is false.

The aposteriori naturalist who endorses this sort of account of our ordinary abilities to infer mental facts just from non-mental facts can happily believe there exist,

somewhere out there, minimal determiners of these mental facts, though armchair philosophers can't find them. What determines the fact that I believe grass is green? I don't know. I suspect it's something to do with my brain, and likely my causal relationship to my environment, and perhaps also my social relationship to others in my community, and perhaps also facts about my history, even the evolutionary history of my ancestors, and it's very likely there are numerous counterfactuals concerning these various areas included too. But what exactly would just one minimal determiner be? I don't know, and perhaps I can't know from the armchair.

Even if apriori knowledge can't be had, I might still gain aposteriori knowledge of what a minimal determiner of my belief that grass is green is. Consider how Kripkean aposteriori necessities are typically known, for example that water is H_2O, or that Hesperus is Phosphorus. (I take the following rough story to be commonplace; see Block and Stalnaker (1999: section 6) for one source.) I observe, and thus know aposteriori, that water and H_2O, and Hesperus and Phosphorus, are correlated in certain ways, for example they share certain properties. From the observed correlations, I infer, as the best explanation, an identity (or constitution) hypothesis: Hesperus is Phosphorus, and H_2O is (or perhaps constitutes) water. Perhaps I could likewise observe that believing grass is green is correlated with certain non-mental facts, facts other than the non-mental fact from which I infer that I believe grass is green (which the transparency theorist would say, again, is just the fact that grass is green). Maybe this mental fact correlates with a neurological, environmental, and/or historical fact. Of course, some famous attempts to naturalize the mental heavily drew from empirically known scientific facts. Dretske, Millikan and Papineau's attempts to naturalize the intentional drew on common empirical knowledge of our evolutionary history. But, again, these attempts, like others, did not uncover a widely accepted minimal determiner. An aposteriori naturalist might diagnose that failure by saying the job simply required more, maybe much more, empirical research.

This methodology for finding minimal determiners is exhibited in some recent approaches to studying phenomenal states. Many philosophers and scientists now think that by searching for a correlate of our phenomenal states in our neurology, we'll find a correlated non-mental, neural state. The discovery that some non-mental state is correlated with a mental state is evidence in favor of the hypothesis that the non-mental state is a determiner of the mental state (and, plausibly, is not merely a determiner, but also, perhaps, is identical to the mental state). (For reviews of and philosophical discussion of recent work in this area, see Block (2005) and Chalmers (1998, 2000). Chalmers is, of course, skeptical that a neural state correlated with a phenomenal state would be a determiner, but this is due to his sympathy for an independent argument—see Chalmers (2010)—against the existence of any non-mental determiner.) Furthermore, many philosophers now also

think that intentional facts are determined by (because they somehow reduce to) phenomenal facts (see Kriegel (2011, forthcoming)): therefore, the scientific search for the minimal determiners of the phenomenal facts may also yield the minimal determiners of the intentional facts. If aposteriori naturalism is right, this sort of approach is much likelier to succeed than the old approach of searching for the non-mental determiners of mental facts from the armchair.

4 Conclusion: A Historical Perspective

Philosophy, as a profession, has fashion trends. Topics go out of fashion, others come into fashion, often in a matter of a few years. It's a strange thing, since the philosophical problems themselves are very old, and progress on them isn't so rapid.

In the 70s and 80s, the naturalization of the mental, especially of intentionality, was all the rage. Things then died down over the 90s. Among the newer trends, starting with the leading papers of Heal (1986), Gordon (1986), and Goldman (1989), was the debate over the role of simulation in how we interpret other minds; the popularity of this trend grew over the 90s and into the early 2000s. And, at the present time, the topic of self-knowledge is enjoying an exceptional surge in attention. These subjective sociological observations are highly vague and, of course, could be argued with. Still, if they're worth something, this rough timeline might be taken to indicate two things.

First, there is a kind of intuitive, profession-wide awareness of when research on a topic has lost momentum, as it long has on armchair naturalization. The project of naturalizing the mental didn't die down just because of critiques by Kripke (1982) or Loewer (1997) or anyone else, but because there was a growing sense that the project wasn't going anywhere, at least while being pursued from the armchair. In place of the old armchair methodology, a new approach that seeks to empirically observe the non-mental correlates of mental states has gained momentum.

And second, perhaps it will often be the case that we are only in a position to satisfactorily explain why momentum fizzled on an out-of-fashion topic once we have the benefit of the insights provided by immersion in the new trends. In that spirit, one aim of this paper has been to draw from current work on self-knowledge and knowledge of other minds to shore up the aposteriori naturalist's diagnosis of why the old armchair naturalization project not only failed, but was bound to fail. Since it was bound to fail, though, we can at least feel reassured that the failure is no threat to our confidence in naturalism.[25]

25 For valuable comments on earlier versions of this material, I'd like to thank ...

References

Armstrong, David (1963). "Is Introspective Knowledge Incorrigible?" *The Philosophical Review* 72(4): 417–432.

Baker, Lynne Rudder (1987). *Saving Beliefs: A Critique of Physicalism*. Princeton, NJ: Princeton University Press.

Balog, Katalin (1999). "Conceivability, Possibility, and the Mind-Body Problem." *The Philosophical Review* 108(4): 497–528.

Block, Ned (2005). "Two Neural Correlates of Consciousness." *TRENDS in Cognitive Science* 9(2): 46–52.

Block, Ned and Robert Stalnaker (1999). "Conceptual Analysis, Dualism, and the Explanatory Gap." *The Philosophical Review* 108(1): 1–46.

Boyle, Matthew (2011). "Transparent Self-Knowledge." *Proceedings of the Aristotelian Society, Supplementary Volume* 85(1): 223–241.

Byrne, Alex (2005). "Introspection." *Philosophical Topics* 33(1): 79–104.

— (2011). "Transparency, Belief, Intention." *Proceedings of the Aristotelian Society, Supplementary Volume* 85(1): 201–221.

Byrne, Alex and James Pryor (2006). "Bad Intensions." In Manuel Garcia-Carpíntero and Josep Macià (eds.), *The Two-Dimensional Framework*. Oxford University Press.

Carey, Susan (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Carruthers, Peter (2009). "How We Know Our Own Minds: The Relationship between Mindreading and Metacognition." *Behavioral and Brain Sciences* 32(2): 121–138.

— (2010). "Introspection: Divided and Partly Eliminated." *Philosophy and Phenomenological Research* 80(1): 76–111.

— (2011). *The Opacity of Mind: An Intergrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

— (2013*a*). "Mindreading in Infancy." *Mind and Language* 28(2): 141–172.

— (2013*b*). "On Knowing Your Own Beliefs: A Representationalist Account." In Nikolaj Nottelmann (ed.), *New Essays on Belief: Structure, Constitution and Content*, chapter 7, 145–165. New York: Palgrave MacMillan.

Carruthers, Peter and Peter Smith (eds.) (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Chalmers, David (1998). "On the Search for the Neural Correlate of Consciousness." In S. Hameroff, A. Kaszniak and A. Scott (eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*, 219–229. MIT Press.

— (2000). "What Is a Neural Correlate of Consciousness?" In T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Issues*, 17–39.

MIT Press.
— (2010). "The Two-Dimensional Argument against Materialism." In *The Character of Conciousness*, chapter 6, 141–207. Oxford University Press.
Chalmers, David and Frank Jackson (2001). "Conceptual Analysis and Reductive Explanation." *The Philosophical Review* 110(3): 315–360.
Davidson, Donald (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
Davies, Martin and Tony Stone (eds.) (1995*a*). *Folk Psychology: The Theory of Mind Debate*. Oxford: Blackwell.
— (1995*b*). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
Dennett, Daniel (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.
Dretske, Fred (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
— (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.
— (1995). *Natualizing the Mind*. Cambridge, MA: MIT Press.
Evans, Gareth (1979). "Reference and Contingency." *The Monist* 62(2): 161–189.
— (1982). *The Varieties of Reference*. Ed. John McDowell. Oxford: Oxford University Press.
Fodor, Jerry (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
— (1990). *A Theory of Content*. Cambridge, MA: MIT Press.
Gallois, Andre (1996). *The World Without, the Mind Within: An Essay on First-Person Authority*. Cambridge: Cambridge University Press.
Goldman, Alvin (1989). "Interpretation Psychologized." *Mind and Language* 4(3): 161–185.
— (2006). *Simulating Minds: the Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
Gopnik, Alison (1993). "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality." *Behavioral and Brain Sciences* 16(1): 1–14.
Gordon, Robert (1986). "Folk Psychology as Simulation." *Mind and Language* 1(2): 158–171.
— (1995). "Simulation without Introspection or Inference from Me to You." In Martin Davies and Tony Stone (eds.), *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.
— (2007). "Ascent Routines for Propositional Attitudes." *Synthese* 159(2): 151–165.
— (2009). "Folk Psychology as Mental Simulation." In Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http: plato.stanford.edu/archives/fall2009/entries/folkpsych-simulation/.
Grandy, Richard (1973). "Reference, Meaning and Belief." *The Journal of Philoso-*

phy 70(14): 439–452.
Hawthorne, John (2002). "Deeply Contingent Apriori Knowledge." *Philosophy and Phenomenological Research* 65(2): 247–269.
Heal, Jane (1986). "Replication and Functionalism." In Jeremy Butterfield (ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press.
Hill, Christopher (1997). "Imaginability, Conceivability, Possibility, and the Mind-Body Problem." *Philosophical Studies* 87(1): 61–85.
Horgan, Terence (1993). "From Supervenience to Superdupervenience: Meeting the Demands of a Material World." *Mind* 102(408): 555–586.
— (1994). "Physicalism." In Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, 471–479. Cambridge University Press.
Horwich, Paul (1995). "Meaning, Use and Truth." *Mind* 104(414): 355–368.
— (1998). *Meaning*. Oxford: Oxford University Press.
— (2005). *Reflections on Meaning*. Oxford: Oxford University Press.
Hyslop, Alec and Frank Jackson (1972). "The Analogical Inference to Other Minds." *American Philosophical Quarterly* 9(2): 168–176.
Jackson, Frank (1996). "Mental Causation." *Mind* 105(419): 377–413.
Jackson, Frank and Philip Pettit (2004 [1993]). "Some Content is Narrow." In *Mind, Morality, and Explanation*, 69–93. Oxford University Press.
Kriegel, Uriah (2009). "Mysterianism." In Tim Bayne, Axel Cleeremans and Patrick Wilken (eds.), *The Oxford Companion to Consciosness*, 461–2. Oxford: Oxford University Press.
— (2011). *The Sources of Intentionality*. Oxford: Oxford University Press.
Kriegel, Uriah (ed.) (forthcoming). *Phenomenal Intentionality: New Essays*. Oxford: Oxford University Press.
Kripke, Saul (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
— (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
Lewis, David (1970). "How to Define Theoretical Terms." *The Journal of Philosophy* 67(13): 427–446.
— (1972). "Psychophysical and Theoretical Identifications." *Australation Journal of Philosophy* 50(3): 249–258.
— (1974). "Radical Interpretation." *Synthese* 27(3/4): 331–344.
Loar, Brian (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.
— (1990/1997). "Phenomenal States (Second Version)." In Ned Block, Owen Flanagan and Güven Güzeldere (eds.), *The Nature of Consciousness*. MIT Press.
Loewer, Barry (1997). "A Guide to Naturalizing Semantics." In Bob Hale and Crispin Wright (eds.), *A Companion to Philosophy of Language*, 108–126. Oxford: Blackwell.

Lycan, William (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Malle, Bertram and Sara Hodges (eds.) (2005). *Other Minds: How Humans Bridge the Divide between Self and Others*. New York: Guilford Press.

McDowell, John (1982). "Criteria, Defeasibility, and Knowledge." *Proceedings of the British Academy* 68: 455–479.

McGinn, Colin (1993). *Problems in Philosophy: The Limits of Inquiry*. Oxford: Blackwell.

— (1999). *The Mysterious Flame*. New York: Basic Books.

Millikan, Ruth Garrett (1984). *Language, Thought and Other Biological Catergories*. Cambridge, MA: MIT Press.

— (1989). "Biosemantics." *The Journal of Philosophy* 86(6): 281–297.

Moran, Richard (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.

Nichols, Shaun and Stephen Stich (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Nisbett, Richard and Timothy DeCamp Wilson (1977). "Telling more than we can know: Verbal Reports on Mental Processes." *Psychological Review* 84(3): 231–259.

Papineau, David (1984). "Representation and Explanation." *Philosophy of Science* 51(4): 550–572.

— (1987). *Reality and Representation*. Oxford: Basil Blackwell.

— (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Peacocke, Christopher (1998). "Conscious Attitudes, Attention and Self-Knowledge." In Crispin Wright, Barry Smith and Cynthia Macdonald (eds.), *Knowing Our Own Minds*, 63–99. Oxford: Oxford University Press.

Perner, Josef and Anton Küberger (2005). "Mental Simulation: Royal Road to Other Minds?" In Bertram Malle and Sara Hodges (eds.), *Other Minds: How Humans Bridge the Divide between Self and Others*, chapter 11. New York: Guilford Press.

Perry, John (2001). *Knowledge, Possibility and Consciousness*. Cambridge, MA: MIT Press.

Putnam, Hilary (1975). "The Meaning of 'Meaning'." In Keith Gunderson (ed.), *Language, Mind and Knowledge*, volume VII of *Minnesota Studies in the Philosophy of Science*, 131–193. The University of Minnesota Press.

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

Russell, Bertrand (1948). *Human Knowledge: Its Scope and Limits*. London: George Allen and Unwin.

Schiffer, Stephen (1993). "Yes, A Reply to Brian Loar's "Can We Confirm Superve-

nient Properties?"." *Philosophical Issues* 4: 93–100.
Schroeter, Laura (2006). "Against Apriori Reduction." *The Philosophical Quarterly* 56(225): 562–586.
Schwitzgebel, Eric (2010). "Introspection." In Edward Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2010/entries/introspection/.
— (2011). "Knowing Your Own Beliefs." *Canadian Journal of Philosophy* 35(Supplement, *Belief and Agency*, ed. D. Hunter): 41–62.
Setiya, Kirean (2011). "Knowledge of Intention." In Anton Ford, Jennifer Hornsby and Fred Stoutland (eds.), *Essays on Anscombe's* Intention. Cambridge, MA: Harvard University Press.
Shoemaker, Sydney (1993). "Special Access Lies Down with Theory-Theory." *Behavioral and Brain Sciences* 16(1): 78–79.
— (1996). *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
— (2009). "Self-Intimation and Second Order Belief." *Erkenntnis* 71(1): 35–51.
Silins, Nicholas (2012). "Judgment as a Guide to Belief." In Declan Smithies and Daniel Stoljar (eds.), *Introspection and Consciousness*. Oxford: Oxford University Press.
Soames, Scott (1998). "Skepticism about Meaning: Indeterminacy, Normativity, and the Rule-Following Paradox." *Canadian Journal of Philosophy* Supplementary Volume 23: 211–249.
Stalnaker, Robert (1984). *Inquiry*. Cambridge, MA: MIT Press.
Stampe, Dennis (1977). "Toward a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2(1): 42–63.
Stich, Stephen (1992). "What Is a Theory of Mental Representation?" *Mind* 101(402): 243–261.
Stich, Stephen and Shaun Nichols (1992). "Folk psychology: Simulation or Tacit Theory?" *Mind and Language* 7(1): 35–71.
Tye, Michael (1992). "Naturalism and the Mental." *Mind* 101(403): 421–441.
— (2000). *Conciousness, Color, and Content*. Cambridge, MA: MIT Press.
— (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts*. Cambridge, MA: MIT Press.
van Fraassen, Bas (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
Williamson, Timothy (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
Wittgenstein, Ludwig (1953). *Philosophical Investigations*. Oxford: Blackwell.