Against A Posteriori Functionalism†

Marc A. Moffett


Abstract

As I will use the term, a posteriori functionalism is the doctrine that the characteristic pattern of the mental properties which is used as the basis for their functional definition will essentially involve a posteriori truths. I will argue that the a posteriori functionalist faces the following dilemma: either she must adopt the results of cognitive science wholesale (in which case her functional definitions will be unacceptably chauvinistic) or she must admit that the psychological principles relevant to giving functional definitions of the mental properties and relations can, at least in principle, be settled by a priori methods (in which case her position will be self-defeating).

There are two constraints on any functionalist solution to the Mind-Body Problem construed as an answer to the question, “What is the relationship between the mental properties and relations (hereafter, simply the mental properties) and physical properties and relations?” The first constraint is that it must actually address the Mind-Body Problem and not simply redefine the debate in terms of other, more tractable, properties (e.g., the species-specific property of having human-pain). Such moves can be seen to be spurious by the very multiple-realizability intuitions that motivate functionalism in the first place. For, according to those intuitions, it is possible for a being to experience pain, have beliefs, etcetera, and yet not only to be of a different species, but to have an entirely different material constitution from human beings. Such intuitions imply that our ordinary mental concepts are not species-restricted.¹

Second, in order to be properly a functional solution, it must in some way incorporate the idea that the mental properties display some characteristic pattern of relations which are both necessary and sufficient for their individuation.² This characterization of functional solutions to the Mind-Body Problem is sufficiently broad to capture most traditional theories, including machine state functionalism (Putnam

† A previous version of this paper was presented at the University of Wyoming. Thanks to George Bealer, John Bengson, Chad Charmichael, Franz-Peter Griesmaier, Dan Korman, Sydney Shoemaker, and two anonymous referees for helpful comments.

¹ See also the discussion in §4.4.

² The requirement that the pattern of relations be sufficient for the individuation of the mental properties is needed to distinguish functional solutions from hybrid solutions (e.g. Searle 1992).
1960), both American and Australian versions of Ramsified functionalism (Shoemaker 1981; Lewis 1972), language of thought functionalism (Fodor 1987), and non-reductive functionalism (Bealer 1997; Shoemaker 1999).

As I will use the term, a posteriori functionalism is the doctrine that the characteristic pattern of the mental properties which is used as the basis for their functional definition will essentially involve a posteriori truths. This is not to say that the a posteriori functionalist must eschew a priori truths or a priori methods altogether, but only that a posteriori investigation is necessary for establishing the truth of at least some of the principles which figure into the individuating pattern. In this paper, I develop and expand a familiar (though underappreciated) argument against a posteriori functionalism (Jackson and Pettit 1993; Jackson & Braddon-Mitchell 1996). (For ease of exposition and because of its familiarity, I will focus my discussion on Ramsified functionalism, but the argument can be easily generalized to cover all other genuine forms of functionalism. The reason for this generality is that the focus of the argument is the modal status of characteristic pattern and, as noted above, the delimitation of this pattern is essential any genuinely functionalist theory; the argument, thus, gets in “on the ground floor.”) The argument turns on the requisite modal status of the principles contained in the base psychological theory on which functionalists Ramsify. In order for the resulting functional definitions to be counterexample-free, these principles must be necessary in the sense that they must hold necessarily for every sentient creature at the requisite level of cognitive functioning. At the same time, we have ample reason to believe that the results of a posteriori scientific investigation will yield a significant number of non-necessary (contingent) principles. Consequently, we cannot simply carry over the results of scientific investigation unfiltered for the purposes of giving functional definitions of the mental; rather, we must have some way of sorting the core psychological principles (on which we may Ramsify) from the peripheral ones (on which we may not). Unless they are able to do this, a posteriori functionalists will not have an adequate account of the multiple-realizability of the mental.

Of course, it is widely accepted that Kripke has cleared the way for just such a theory by showing that the necessary and the a priori are conceptually independent. And, at least prima facie, the existence of necessary a posteriori truths provides the perfect framework for defending a posteriori functionalism. Surprisingly, however, there is no generally acceptable means of selecting necessary truths from the results.
of cognitive science which is consistent with their status as a posteriori truths, or so I shall argue. Consequently, the necessary a posteriori provides no comfort for the functionalist. Thus, the a posteriori functionalist faces the following dilemma: either she must adopt the results of cognitive science wholesale (in which case her Ramsified definitions will be mistaken) or she must admit that the psychological principles relevant to giving functional definitions of the mental properties can at least in principle be settled by a priori methods. On either horn of the dilemma, a posteriori functionalism is untenable. If this is correct, it follows that physicalists must either abandon functionalism altogether or accept the existence of highly non-trivial a priori truths.

I begin with a brief recapitulation of the functionalist strategy for defining mental predicates by way of Ramsification. As noted above, what I have to say extends equally to other functionalist approaches.

1. Functionalism: A Brief Recapitulation

The functionalist strategy for defining mental predicates begins with an articulation of a base psychological theory, \( \varphi(\{\text{believes, desires, hurts, } \ldots , \text{ sees}\}) \). Having arrived at some satisfactory base theory, the functionalist then removes the mental predicates from the theory and uniformly replaces them with predicate variables. For example, in the second step the functionalist uniformly replaces every occurrence of the predicate ‘believes’ with the predicate variable ‘\( R_1 \)’; every occurrence of the predicate ‘desires’ with the predicate variable ‘\( R_2 \)’; and so on. The result is a matrix theory, \( P \), in which every mental predicate is replaced with an appropriate predicate variable. Schematically: \( P[\{R_1, R_2, R_3, \ldots , R_n\}] \). Or, more simply, \( P[R] \), where \( R \) is the sequence of predicate variables \( \langle R_1, R_2, R_3, \ldots , R_n \rangle \). This gives us the characteristic pattern of psychological interaction.

Finally, the functionalist existentially quantifies over the (free) predicate variables in the matrix theory: \( (\exists R)(P[R]) \).\(^3\) Given this, one can provide functional definitions of the mental properties according to the following definitional schema: \( x \) has mental property \( m_i \) iff there exists a sequence of properties

\[^3\text{That is: } (\exists R_1)(\exists R_2)(\exists R_3) \ldots (\exists R_n)(P[R_1, R_2, R_3, \ldots , R_n]). \text{This method of defining the theoretical vocabulary was introduced by F. P. Ramsey (1931).} \]
that satisfy the matrix theory and \( x \) has the \( i \)th member of that sequence. In symbols: \( m_i(x) \iff (\exists R)((P[R]) \land R_i(x)) \).

2. The Problem of Psychological Chauvinism

Just as we have strong intuitions that mental properties may be instantiated in very different physical systems, so too we have strong intuitions that many psychological principles may vary considerably from world to world or species to species. That is, some of the law-like psychological principles that govern human behavior need not govern the behavior of an arbitrary sentient (or rational) being at the same general level of cognitive functioning. The domain specific modular processes posited in evolutionary psychology provide a detailed study of the sort of evolved variation that is possible (Barkow, Cosmides, & Tooby 1992; Pinker 1997).\(^5\) Such an evolutionary perspective effectively requires that different species, even different species at the same general level of cognitive sophistication, will vary considerably in their evolved psychological make-up as a result of differences in available genetic variation, selective pressures, genetic drift (and other nonselective forces in evolution), and developmental environment.

In fact, the very debate currently in play in cognitive science between domain-specific, modular theories of mind and general purpose systems is itself most naturally construed as a debate over which of two possible worlds we actually inhabit. (See Elman et. al (1996), Tomasello (1999, 2003), and Culicover & Jackendoff (1999) for a sense of the available alternative theories, especially with regards to the question of language.) And even if the debate can be relatively heated at times, it seems unlikely that either camp would be willing to accuse their opponents of defending a logically impossible theory of mind!\(^6\) But these competing theories posit distinct psychological laws. For example, modular theories of mind typically take the various mental modules to be informationally encapsulated to a greater or lesser extent (Fodor 1983). But then such theories will take the characteristic pattern of the interaction between various mental states

---

\(^4\) For example, \( x \) hurts \( \iff (\exists R)((P[R]) \land R_3(x)) \).

\(^5\) This is not intended as an endorsement of evolutionary psychology, see Buller (2005) for a recent critical discussion.

\(^6\) Here I follow tradition and use ‘logically impossible’ to mean ‘false at all possible worlds’ and not ‘false in virtue of logic alone’ (van Inwagen 1998). I am unwilling to follow the recent trend of characterizing modalities in terms of their epistemic ground.
(e.g., belief, learning, perception, etc.) to be different from the pattern derived from general processing theories.

To take a concrete case, consider the process of learning categories. One possible principle of categorization that has been posited by developmental psychologists is what I will dub the *basic category effect*:


[BCE] If \( x \) is presented with a group of objects \( \{o_1, \ldots, o_n\} \) from a given category \( C \) for perceptual inspection, then if \( C \) is a "basic" category \( x \) will learn \( C \) more rapidly without help than if \( C \) is a superordinate category (Markman 1989).

Principles such as [BCE] are routine in contemporary cognitive science. Nevertheless, it seems hardly worth mentioning that [BCE], if true, would merely be a contingent fact about human psychology. Surely it is possible for there to exist beings that satisfy the antecedent of [BCE] without satisfying the consequent (for instance, organisms who are better or worse than us at analogical comparisons or better or worse relative to certain domains but not relative to others).\(^7\)

Now consider what happens if we include [BCE] as an additional conjunct in the base psychological theory \( P \) from which we derive our functional definitions of the mental properties. In this case, we get the following base psychological theory (ignoring additional complexities): \( P[\{\text{perceptually inspects, learns}\}] \& \text{BCE[perceptually inspects, learns]} \). The corresponding functional definition of learning is as follows:

\[
\text{Learns}(x, C) \iff \text{def} \left( \exists \langle R_1, R_2 \rangle \right)
\]

(i) If \( x \) is presented with a group of objects \( \{o_1, \ldots, o_n\} \) from a given category \( C \) for \( R_2(x) \), then if \( C \) is a "basic" category \( R_1(x, C) \) more rapidly without help than if \( C \) is a superordinate category,

(ii) \( P[\langle R_1, R_2 \rangle] \) and

(iii) \( R_1(x, C) \)

---

\(^7\) It is possible that there is a reading of [BCE] according to which ‘basicness’ is relativized in such a way that it comes out analytic (or, at least, necessary) that those concepts which are learned first are basic. This is not the intended reading of [BCE]; the intended reading is one of ontological basicness. Nevertheless, the example here is purely illustrative; many other seemingly contingent principles would serve equally well.
This definition requires that there be properties which interact in the way specified by [BCE] but which otherwise satisfy the perceiving and learning roles. But since [BCE] is by hypothesis a contingent psychological principle, we know that a great many beings who should count as minded, will fail to satisfy the proposed definition in virtue of failing to satisfy condition (i). That is, for individuals who do not find basic categories easier to learn without coaching than superordinate categories, whatever realizer properties satisfy the matrix theory $P$ will fail to satisfy the pattern in [BCE], and conversely.

The point can be stated more generally. Let $P$ (the a posteriori base psychological theory) consist of two primary conjuncts, $U$ and $H$, where $U$ consists of all those psychological principles which we take to be universal principles of psychology and $H$ consists of all those principles that we take to be contingent principles of human psychology. Now, if we simply adopt $P$ as the base theory, the Ramsified definition for an arbitrary mental state $m_i$ will be as follows:

$$m_i(x) \iff_{\text{def}} (\exists R)((U \land H)[R]) \land R_i(x).$$

But this definition tells us that no state $r_i$ that does not satisfy the contingent principles of human psychology (that is, $H$) can be a realizer state of $m_i$. This result is, in its way, every bit as chauvinistic as the sort of matter chauvinism that rightly convinced many philosophers to eschew the identity theory in the first place.

Thus, it is extremely easy to build functional definitions that are inadequate because the base psychological theory on which one Ramsifies contains contingent, idiosyncratic clauses concerning how human psychology works (Jackson & Pettit 1993). The point here is a modal generalization of Shoemaker’s (1981) point that a functionalist theory built simply on commonsense psychological platitudes runs the risk of incorporating false clauses in the base theory. The generalization is simply that incorporating true, but possibly false, clauses will have essentially the same untoward consequences for functional definitions. The moral is that the base psychological theory $P$ from which our functional definitions derive must be restricted to psychological principles that hold of all conscious beings (that is, that hold necessarily of creatures with minds at the relevant level of cognitive functioning). Failure to satisfy this constraint will result in a theory that fails in an important respect to capture the multiple-realizability of the mental, specifically, by excluding individuals who differ from us in what are (by assumption) psychologically inessential ways.
3. The Necessary A Posteriori: Kripke’s Strategy

Given that the base theory from which functional definitions derive must be restricted to necessary psychological principles, is it nevertheless possible to maintain with the a posteriori functionalists that which principles these are can be known by wholly a posteriori means? At face value, recent work in the philosophy of language seems to clear the way for just such a position. Specifically, Kripke (1972), Putnam (1975) and others have argued for the existence of necessary a posteriori truths. And it would seem that an advocate of a posteriori functionalism could exploit this development in order to defend a view according to which the base theory is derived by a posteriori scientific means, but the resulting principles are nevertheless necessary psychological truths.

However, it is obvious that the mere existence of necessary a posteriori truths of one type (e.g., those expressed by nonredundant identity claims) does not justify the claim that truths of an entirely different logical category (e.g., nomological principles) may be necessary a posteriori as well. Thus, although a posteriori functionalists may find solace in the mere possibility of necessary a posteriori truths, it remains to be seen if they can muster the resources to take advantage of them without taking on board some hefty and rather implausible metaphysical commitments. In this section, I am going to argue that the Kripkean strategy for establishing necessary a posteriori truths is not adequate for the purposes of a posteriori functionalism. In §4 I will consider and reject some possible, nonKripkean attempts to establish necessary a posteriori truths. I will begin by laying out the Kripkean strategy.

The most convincing cases of the necessary a posteriori consist of identity statements involving rigid designators. Crucially, the necessity of such claims follows simply as part of the intuitive logic for identity; specifically, via the following logical theorem:

\[ \xi = \zeta \rightarrow \Box (\xi = \zeta) \]

where \( \xi \) and \( \zeta \) are arbitrary rigid designators. The case most relevant for our purposes concerns natural kind identities. We can think of such a posteriori necessities as arising from three factors, which together constitute the Kripkean Strategy:

---

8 General principles of necessitation such as [Id] may be established by less general logical intuitions. For instance, we have the intuition that if water is H2O, then necessarily it is. (And this, in turn, might be further supported by our concrete case intuitions about, for instance, Twin Earth.) These less general intuitions are adequate for generating
(i) an a priori understanding that the relevant kinds have (default) microstructural individuation conditions,

(ii) an a posteriori discovery of the microstructure, and

(iii) the logic of identity in an intensional setting; specifically, theorem [Id] above.

Taken together, the first two factors establish, by essentially a posteriori means, the truth of the bare identity that serves as the antecedent in an instance of [Id]. The third factor then provides for the blanket necessitation of the identity claim via logical inference. I will call this third factor the “formal component” of the Kripkean strategy because it allows us to derive the necessity claim without any independent direct assessment of the claim itself.\(^{10}\)

necessary a posteriori truths. The point I want to emphasize, however, is that these less general intuitions are systematic: if you have the water intuitions, you will also have the intuition that if gold is the element having atomic number 79, then necessarily it is and other similar ones. To my knowledge, there are no plausible examples where we have the intuition that if p, then necessarily p for some specific, isolated proposition p. I will return to this point in §4.2.

If, as Soames (2002, 241-253) argues, Kripke’s concept of rigid designation cannot be extended to natural kind terms, then the constraint on [Id] will need to be stated more carefully. However, I believe that Soames’ arguments can be adequately handled by way of a dual aspect syntax of the sort defined by Menzel (1993).

\(^{9}\) For the purposes of this paper I will assume that sentences such as ‘water is H\(_2\)O’ express identities. However, it is plausible that the relevant a posteriori necessity is actually a claim about composition, viz., that water is necessarily composed of H\(_2\)O (see Barnett 1998). In this case, we will need a corresponding general principle concerning the necessity of composition.

\(^{10}\) Essentially the same kind of formulaic derivation would work for (purported) a posteriori necessities involving composition (as in, e.g., Kripke’s table example). The crucial point in this case is that we have essentialist intuitions concerning origins: \((\forall X,Y)X\text{ is originally composed from }Y, \text{ then } ☐(X\text{ is originally composed from }Y)\). Of course, as an anonymous referee suggested, it may be that X need not necessarily be composed 100% from Y and that there may not be any exact percentage which is necessary. Even if this is correct, it does not undermine the point that our intuitions about the necessity of origins are general and systematic in the requisite way; it merely shows that there is a certain degree of vagueness in our concept of material composition.

Of course, it goes without saying that any proposition p that is entailed by an a posteriori necessity is itself an a posteriori necessity. For instance, it follows from ☐(water = H\(_2\)O) and ☐(H\(_2\)O contains hydrogen) that ☐(Water
The presence of a formal component in the Kripkean strategy is crucial for establishing the existence of necessary a posteriori truths. For, without such a component, the truth or falsity of the modal claim would require some independent direct assessment, for instance, by means of modal intuition. Call such methods for establishing necessary truths substantive a priori methods. Clearly, the exclusive use of substantive a priori methods in establishing a modal truth is inconsistent with the a posteriori standing of the embedded proposition. After all, if one can establish “It is necessary that p” a priori, then logic assures the truth of the embedded claim p. Thus, if condition (iii) were not merely formal, but relied on some substantive a priori method, then the mechanism for establishing that the identity is necessary would be inconsistent with its standing as a posteriori. In this event, while our means of discovering the proposition may have been a posteriori, the proposition itself would be a priori. The empirical and formal components of the Kripkean strategy work in concert to generate necessary a posteriori truths.

Now, to put it in a nutshell, the problem for a posteriori functionalists is that there is no plausible formal component for necessitating the sorts of law-like causal/nomological regularities offered up by cognitive science. To see why, consider the corresponding steps in the Kripkean strategy. Corresponding to the first factor stated above, functionalists maintain that we have an a priori understanding of the mental properties as having functional (vs. microstructural) individuation conditions. Corresponding to the second factor, a posteriori functionalists maintain that the functional pattern which forms the basis of the functional definitions of the mental properties is discovered by essentially a posteriori means (e.g., empirical investigation). Simplifying considerably, this pattern consists in large measure of causally or

contains hydrogen). But the proposition that water contains hydrogen is a posteriori. As far as I have been able to determine, the existence of this sort of derived necessary a posteriori truths does not affect the ensuing argument.

11 Of course, all necessary a posteriori truths make use of some substantive a priori methods in establishing the formal means by which the unconditional necessity is derived.

12 One should bear in mind here that, according to standard usage, a proposition p is a priori iff its truth can be established by wholly a priori methods.

13 Strictly speaking, I believe that the most we have settled to this point is that the mental properties are multiply realizable, a view which is consistent with the rejection of functionalism. The stronger claim that we can establish a priori that the mental properties have functional individuation conditions, is far more controversial. I will, however, set this issue aside here.
nomologically necessary material implications between events or properties: \( \bullet \varepsilon (\text{all } F\text{s are } G\text{s}) \).\(^{14}\) Thus, the first and second components of the Kripkean strategy exactly mirror the correlates in the case of identity claims.

However, few philosophers have been willing to swallow wholesale the logical principle which corresponds to the third factor in the Kripkean strategy. That principle would be the following: every causally or nomologically necessary correlation between events or properties is metaphysically necessary. In symbols:

\[
[\text{LN}] \quad \bullet \varepsilon (\text{all } F\text{s are } G\text{s}) \rightarrow \bullet (\text{all } F\text{s are } G\text{s}).
\]

Moreover, [LN] cannot be given the sort of intuitive corroboration that Kripke and Putnam provided for the corresponding logical principle for identity. Indeed, as discussed in §2, it would be highly counterintuitive to promote many of the law-like regularities discovered by cognitive science to the level of metaphysical necessities, that is, necessary claims about the behavior of sentient beings.

Thus, unlike the case of theoretical identities arising from a posteriori scientific investigation, the causal/nomological regularities discovered by cognitive science do not appear to admit of blanket necessitation. It seems that we must sort these regularities on the basis of their modal status. That is, it seems that we must be able to independently determine which regularities are necessary and which contingent. However, as noted above, if we can determine a priori which regularities hold necessarily, then a posteriori investigation is at least in principle unnecessary. Call this the Sorting Problem for a posteriori functionalism.

4. Responses

There are, as far as I have been able to determine, four possible ways for the a posteriori functionalist to respond to the Sorting Problem, all of which try to circumvent the need for sorting

---

\(^{14}\) Following Tooley (1977, 1987), I prefer a property-theoretic account of laws, namely, that laws are relations between natural properties or universals. Letting \( I \) be the relation of nomic implication, then we have \( \phi I \gamma \) for some properties \( \phi \) and \( \gamma \) determined by the respective kinds. It is this nomic relation between properties that determines the observed regularity. As far as I have been able to determine, nothing in the present discussion turns on these issues.
Against a posteriori functionalism

The first response is to argue that (intuitions to the contrary notwithstanding) law necessitarianism is true and, therefore, that [LN] or some suitably circumscribed variant of [LN] is true. The second response is to argue that some laws \( L \) intuitively necessitate, but to deny that the set of laws which necessitate in this way share in common feature that could be plugged into step (iii) in the Kripkean strategy. I will call this form of necessitarianism \textit{particularist necessitarianism}. The third response to the Sorting Problem is to argue that the sorts of principles ultimately adopted by a mature cognitive science will not (like [BCE]) be contingent. I will call this view \textit{de facto necessitarianism}. The fourth response is to try to make the psychological principles conditional on relevant background conditions. I will consider each of these responses in turn.

\subsection*{4.1. Law Necessitarianism}

According to law necessitarianism some or all of the laws of nature are logically necessary (see, e.g., Shoemaker 1980, 1998; Swoyer 1982; Fales 1993; Ellis & Lierse 1994; Ellis 1999, 2002). In its weak form (i.e., as a claim that some, but not all, of the laws of nature are logically necessary), law necessitarianism is not at issue. After all, a priori functionalists grant that some of the psychological laws are logically necessary; this much is entailed by functionalism itself. So the Sorting Problem is not solved by appeal to necessitarianism \textit{per se}, but only by those forms of necessitarianism that may be used to underwrite the formal component in the Kripkean strategy for establishing necessary, a posteriori truths.

There are two ways of doing this. One is to simply argue outright for a strong form of law necessitarianism according to which \textit{all} of the laws of nature are logically necessary; that is, to defend an unqualified version of [LN]. The second way is to argue that all the laws of nature possessing some feature \( \Phi \) are metaphysically necessary; that is, to defend a qualified version of [LN]. I will begin with the general (and more widely accepted) approach.

\footnote{There is one other, “mysterian” response (cf. McGinn 1991). On this view, mental properties would have functional definitions and these definitions would be given in terms of necessary psychological laws. However, the claim would be that we cannot know that these laws are necessary (either a priori or a posteriori). Consequently, we could never know that our functional definitions are correct. Discussion of this form of mysterian functionalism is beyond the scope of this paper, so I will simply leave it to the reader to decide for his or herself as to the plausibility of this view.}
Although a thorough, critical discussion of law necessitarianism is beyond the scope of this paper, I will briefly consider what I take to be the most important consideration in its favor, namely, the view that properties (generally) are individuated by their complete set of causal powers (Shoemaker 1998; cf. Ellis 2002). Call this view causal power essentialism or CPE. (Presumably, Ramsifying functionalists accept that mental properties are individuated in terms of some of their causal powers.16) Let us grant that if CPE is true then so is [LN]; thus, if CPE is true, then a posteriori functionalism is vindicated. Conversely, if properties (in particular, mental properties) are not individuated by their causal powers or if they are individuated only by some subset of their causal powers, then the Sorting Problem will remain in tact.

Prima facie the latter conclusion seems most warranted. After all, as noted in §2, situations in which various laws of psychology are different from the way they actually are seem intuitively possible. Moreover, because the Kripke-Putnam arguments for the necessary a posteriori rely essentially on the reliability of modal intuition, a posteriori functionalists cannot simply ignore these counterintuitive consequences.17

Shoemaker attempts to avoid this problem by invoking a rephrasal strategy similar to that used by Kripke to blunt the descriptivist intuitions that certain identity claims are contingent. According to Kripke, the alleged intuition that, say, Hesperus might not have been Phosphorus is actually being misreported. Our actual intuition is that we might have been in an epistemically identical situation where the thing we named “Hesperus” was not the thing we named “Phosphorus”. A similar response, Shoemaker suggests, can be given for the apparent contingency of laws.

Let the law be that strychnine in a certain dosage is fatal to human beings. We can grant that it is imaginable that ingesting vast amounts of what passes certain tests for being strychnine should fail to be fatal to what passes certain tests for what passes for being a human being, but deny that this amounts to imagining a human being surviving the ingestion of that much strychnine (1998, 62).

16 Though, as Yablo (1992) shows, care is required if this claim is to be defensible.

17 For a general discussion of law necessitarianism and the problem of “cherry picking” intuitions, see Korman (2005).
Thus, according to Shoemaker, our intuition that the law in question is contingent can be rephrased in terms of a corresponding epistemic possibility.¹⁸

There are two serious problems with this type of rephrasal strategy in the present context. First, Kripke and Putnam were able to elicit intuitions to the effect that identity claims are necessary; for instance, the well-known Twin Earth intuitions. These intuitions, however, are in direct conflict with the contingent identity intuitions. Kripke’s rephrasal strategy, therefore, was not merely an ad hoc response to a group of problematic intuitions, but a necessary means for resolving this stalemate. By contrast, neither Shoemaker nor (to my knowledge) anyone else, has made it seem intuitively plausible that all or even most of the laws of nature are necessary. Consequently, invoking the rephrasal strategy in the case of laws appears to be unacceptably ad hoc.

The second problem with Shoemaker’s rephrasal strategy is that, like Kripke’s original one, it doesn’t work in the psychological context. Consider Shoemaker’s strychnine example. This example works by exploiting the epistemic gap between the macro properties of strychnine (i.e., “what passes certain tests for being strychnine”) and its microstructure. Since the former are not sufficient for determining the latter and since strychnine is individuated in terms of its microstructure, we can readily rephrase the intuition in terms of strychnine-like stuff.¹⁹ Mental properties, however, are not “twin-earthable” in the way that microstructural properties are (Bealer 1987; Chalmers 2004). Take, for instance, a situation in which you are in pain. We can’t imagine being in exactly this epistemic situation but, nevertheless, not being in pain! In this case, there is no gap to exploit between the phenomenal properties of pain and pain itself. As a result, our imagined counterexamples can’t be cases where something that phenomenologically “passes for pain (but isn’t)” is implicated in different laws than the ones pain is implicated in.²⁰

¹⁸ I assume here that Shoemaker’s response may with equal felicity be stated in terms of modal intuition rather than “imagination.” If not, then the response fails in any event.

¹⁹ If by “human beings” in the example Shoemaker intends the biological kind Homo sapiens and if H. sapiens is individuated in causal-historical terms (Ghiselin 1974), then the same point holds mutatis mutandis here.

²⁰ Of course, mental properties that are both externalist and prime (Williamson 2000) might be twin-earthable. Even if this is correct, however, many mental properties on which we will want to base our Ramsified definitions will not have these properties. I owe this point to John Bengson.
It appears, therefore, that Shoemaker is wrong in his assessment that Kripke-style rephrasal strategies are sufficient for warding off apparent counterexamples to strong law necessitarianism. My conclusion, therefore, is that a commitment to an unqualified version of [LN] is not warranted.\(^\text{21}\) Given this, the advocate of a posteriori functionalism might prefer a more modest approach. Rather than defend [LN] simpliciter, she might instead argue that we are justified in necessitating all and only those laws which possess some further characteristic \(\Phi\). Specifically, she might defend the following qualified version of [LN]:

\[
[\text{LN}^*] \quad (\exists \Phi)(\forall p)(p = \Box, (\text{all Fs are Gs}) \land \Phi(p)) \rightarrow \Box p.
\]

Given [LN\(^*\)], the a posteriori functionalist can then restrict the base theory of her Ramsified definitions to those laws discovered by cognitive science that possess the feature \(\Phi\).

Obviously, given the existential quantifier in [LN\(^*\)], any discussion of this strategy will be necessarily incomplete. One obvious problem with [LN\(^*\)] is that it is difficult to provide any clear candidate for \(\Phi\). In part, this is because the demands on an adequate feature \(\Phi\) are pretty strong. It must be the case that every law having \(\Phi\) must be a plausible candidate for necessitation. What is more, evidently many of the more obvious candidates for \(\Phi\) don’t work, or at least we have no reason for thinking that they will. Consider, for instance, the possibility of letting \(\Phi = \text{the property of being a basic law}\). We have already noted that it would be highly counterintuitive to take many higher-level or derived laws to be anything but contingent. As Swoyer (1982, 211–216) notes, however, many derived laws follow necessarily from laws which are the most plausible candidates for being basic laws. But clearly, if the basic laws were necessary then any laws derived from them would also have to be necessary. Now, for virtually all of these derived laws, we have fairly strong contingency intuitions.\(^\text{22}\) Moreover, we lack strong intuitions about the necessity of the basic laws from which they derived. Consequently, the intuitive contingency of the derived laws provides a reason for thinking that the basic laws are themselves contingent.

\(^{21}\) If, despite these problems, you are convinced of the virtues of necessitarianism, then you can take the moral of this paper to be the claim that a posteriori functionalism entails strong necessitarianism.

\(^{22}\) Again, recall that the a posteriori functionalist is in no position to ignore or downplay the epistemic significance of these intuitions.
There is, however, a more general reason for thinking that an appeal to [LN*] will not help the a posteriori functionalist. One of the main motivations for adopting a necessitarian theory of laws and CPE (indeed, the main reason) is to give a principled, uniform account of their metaphysical basis. Since the advocate of [LN*] is committed to the existence of contingent laws (viz., those laws not possessing Φ), she must give a separate account of the metaphysical basis of these types of laws. But if she is able to do this, then it is unclear why the account wouldn’t be adequate for those laws which are purported to be necessary. But if this is so, then the move to restricted necessitarianism is unmotivated. The problem, in effect, is that lacking intuitive support, the plausibility of necessitarianism rests almost entirely on the theoretical work it does. Unless one adopts a strong necessitarian thesis, however, one significantly undermines the main theoretical motivation for adopting a necessitarian stance to begin with. For one is still left with the theoretical problem of accounting for the metaphysical ground of non-necessary laws.

Given this problem of providing a cogent (intuitive or theoretical) motivation for the view and given that there are no clear candidates for Φ, I take it that restricted necessitarianism cannot underwrite a satisfactory formal mechanism for the a posteriori functionalist to exploit in the third step of the Kripkean strategy outlined in §3.

4.2 Particularist Necessitarianism

At this point the a posteriori functionalist might opt for a particularly radical form of necessitarianism, particularist necessitarianism. Recall that in laying out the Kripkean strategy I took the formal component to invoke a general principle of necessitation. In the traditional a posteriori identities, this claim was that if \( x = y \), then \( \Box(x = y) \). At this point, however, it might be suggested that such general principles are not needed; all that is required, it will be said, is that we have specific intuitions of the form \( \text{if } A = B, \text{ then } \Box(A = B) \) for particular As and Bs. Strictly speaking, this is correct and provides an adequate formal component for the derivation of necessary a posteriori laws. In uncontroversial cases of the necessary a posteriori, however, there exists some more general principle of necessitation (such as [Id]); principles which arguably provide the explanatory basis for our more specific intuitions. The particularist necessitarian, however, abandons any claim to these more general necessitation principles. Instead, she maintains that it is a brute fact that for at least some laws L, it will seem intuitively true to us that if L, then \( \Box L \). More specifically, the particularist necessitarian maintains the following pair of theses:
[PN1] There are certain, specific sets of alternative competing laws that, if true, intuitively necessitate.23

[PN2] There is no interesting general relation amongst these sets of laws that can be captured by a general necessitation principle such as [LN].

The first significant problem for the particularist necessitarian is that the intuitions required by [PN1] are scarce to nonexistent. Consider, for instance, Coulomb’s law which governs the electric force acting on a point charge $p$ as a result of the presence of a second point charge $q$:

$$ F = \frac{pq}{4\pi \varepsilon_0 r^2} $$

(where $\varepsilon_0$ is the electric permittivity of space having the value $8.854187817 \times 10^{-12}$ F/m). Surely if we are going to have any necessitarian intuitions about particular laws, this is such a case. So is it true that if in fact $F = \frac{pq}{4\pi \varepsilon_0 r^2}$, then necessarily $F = \frac{pq}{4\pi \varepsilon_0 r^2}$? Intuitively not. Intuitively, the electric permittivity of space could have been at least slightly greater or slightly smaller than $\varepsilon_0$, say, $8.85418781659 \times 10^{-12}$ F/m; or the force exerted on interacting particles could have been a function not of the distance between the particles squared, but raised to the $1.9999999999999$th power. Thus, even in the “best case scenarios,” the relevant necessitarian intuitions do not appear to be forthcoming. As a result, the particularist necessitarian faces the same basic set of problems that the strong necessitarian faces: no clear rationale for rephrasing the contingency intuitions and no clear way of doing so.24

---

23 The idea is that we have various alternative possible laws $L_1, \ldots, L_n$ such that, at most, one of the $L_i$s is true and for each $L_i$ we have the intuition that if $L_i$, then $\Box L_i$.  

24 Note that these alternative scenarios were certainly both epistemic possibilities in the not-too-distant past. But given the traditional assumption that the epistemically possible worlds at any given time are a subset of the logically possible worlds, the particularist necessitarian is forced to rephrase these epistemic possibilities along the lines suggested by Shoemaker (see §4.1 above). But notice how implausible such rephrasals are in this case. For instance, consider the apparent epistemic possibility of discovering that $F = \frac{pq}{4\pi \varepsilon_0 r^{1.9999999999}}$. According to the particularist necessitarian, it is not a logical possibility that $F = \frac{pq}{4\pi \varepsilon_0 r^{1.9999999999}}$, and so not an epistemic possibility that $F = \frac{pq}{4\pi \varepsilon_0 r^{1.9999999999}}$. The actual epistemic possibility, which we have misreported, is really .... What? That something which passes for force (but isn’t) is a function of $r^{1.9999999999}$? Or that we might express the law that $F = \frac{pq}{4\pi \varepsilon_0 r^2}$ by uttering the sentence ‘$F = \frac{pq}{4\pi \varepsilon_0 r^{1.9999999999}}$’? These rephrasals stretch our credulity beyond its limits. Perhaps the most credible rephrasal is that we can imagine the evidence being misleading in such a way that we come to believe
In addition, the a posteriori functionalist who adopts this approach apparently cannot settle on one or a few general psychological principles. For the base theory on which the Ramsified definitions are built must be sufficiently robust to guarantee uniqueness; the pattern has to be the pattern characteristic of the mental properties and nothing else. But it is unclear why we should think that, even if we do have intuitions that some specific psychological laws are necessary if true, we have sufficiently many of them.

Moreover, [PN2] places the particularist necessitarian on uncertain epistemic ground. In the case of traditional a posteriori necessities, we believe that there are general facts about our referential intentions, natural kind concepts, and the concept of identity (or composition) which provide an explanatory basis for whatever specific intuitions we have. Our more specific intuitions, thus, give us reason to posit the general principles (perhaps as some sort of inference to the best explanation); but as an explanatory matter, it is the general principles which account for our specific intuitions. The particularist necessitarian, however, abandons this satisfying theoretical framework and leaves her specific intuitions as stand alone facts that are not explained by more general logical or metaphysical principles. As a result, she is apparently forced to claim that (some of) our concepts have highly idiosyncratic essentialist content. But it is unclear what plausible sort of account can be given of this kind of content. It is certainly implausible that it is determined by our referential intentions. And even if one adopts something like a Lewisian (1983) theory of properties according to which we intend to refer to the most natural corresponding property, this view entails a disturbingly disuniform underlying metaphysics whose most natural properties at a given ontological level don’t pattern in systematic ways.

4.3. De Facto Necessitarianism

There is one final form of necessitarianism which the a posteriori functionalist might attempt to exploit, de facto necessitarianism. According to de facto necessitarianism, the laws posited by a completed or mature cognitive science will all be, as a matter of fact, metaphysically necessary. In effect, de facto necessitarianism is a special case of restricted necessitarianism with \( \Phi \) being the property of being a law of

\[
F = \frac{pq}{4\pi e^1.99999999999},
\]

that in fact it doesn’t. But even this rephrasing rings a bit hollow by comparison, in part no doubt, because we have no countervailing pro-necessitarian intuitions that must be accommodated. For further discussion of these sorts of examples see (Bird 2001, 2002; Beebee 2002).
a completed cognitive psychology. I take it that Georges Rey gives voice to the _de facto_ necessitarian view in the following passage:

Vesuvians might differ from humans at _many_ different levels of description: substance, physical arrangement of their “nervous system,” input system, monocular vs. binocular vision, even at many levels of “information processing.” The question is whether they differ at the level at which a mature psychology will define psychological phenomena [1997: 190; emphasis in the original].

If _de facto_ necessitarianism is correct, then once we have a mature cognitive science in our hands, we can simply carry over all of the laws for the purposes of Ramsification, since they will be pitched at a level which (allegedly) makes it plausible to hold them to be necessary, a posteriori truths.

I confess that _de facto_ necessitarianism sounds a bit far-fetched to me on its face. If anything, current work in cognitive psychology (which is now heavily informed by evolutionary biology) seems to be heading in exactly the opposite direction—toward largely species-specific psychological principles. Moreover, human psychologists are likely to remain interested not just in those elements of human cognition that are universal, but also those that are specific to human beings. From the point of view of psychology (and, in particular, applied psychology), there is little real interest in the distinction between causally necessary but contingent principles and metaphysically necessary ones. So, there is little reason to think that a mature/completed cognitive science will pitch the psychological laws at the level of generality necessary to make it plausible to hold them to be metaphysically necessary.  

But even if this is not so, _de facto_ necessitarianism does not provide a solution to the Sorting Problem. As the preceding discussion suggests, there does not seem to be any epistemic guarantee that a mature psychology will be formulated entirely in terms of logically necessary truths. Consequently, even if things turn out that way—even if cognitive science actually will hit upon all and only (or at least only) the logically necessary truths of psychology—the a posteriori functionalist can’t simply assume this. It is at the very least an epistemic possibility that a mature cognitive science will continue to be a mixed bag of

---

25 Granted, some of the contingent principles might be weeded out by researchers in AI, specifically, those that are in some way grounded in implementation. But AI researchers are as likely as not to simply incorporate many contingent components of human cognition.
necessary and contingent principles; in fact, it is arguably a good bet that it will be. But if there is no such guarantee, then the Sorting Problem is still in play.

Let me spell this worry out in more detail. Suppose that we let \( \mathcal{P} \) (the a posteriori psychological theory we actually arrive at in a mature cognitive science) consist of two primary conjuncts, \( \mathcal{U} \) and \( \mathcal{H} \), where \( \mathcal{U} \) consists of all those psychological principles which are universal principles of psychology and \( \mathcal{H} \) consists of all those principles which are contingent principles of human psychology. The \textit{de facto} necessitarian maintains that, as a matter of fact, \( \mathcal{H} \) might (will?) turn out empty in a mature cognitive science. But given that it is an epistemic possibility that \( \mathcal{H} \) will be not be empty, we need some reason for thinking that all of the principles of \( \mathcal{P} \) should be included in \( \mathcal{U} \). This is clearly just the limit case of the Sorting Problem where, as a matter of fact, the entirety of \( \mathcal{P} \) is selected for inclusion in the base theory of the Ramsified definitions. But the mere fact that the entirety of \( \mathcal{P} \) is selected does not explain the epistemic basis for this selection. And given that it is epistemically possible for a mature/completed cognitive science to yield something other than universal, necessary truths, the mere fact that \( \mathcal{P} \) is the result of a mature/completed cognitive science does not justify us in Ramsifying over \( \mathcal{P} \).

The upshot is that the \textit{de facto} necessitarian actually needs a much stronger, and even more implausible, thesis in order to avoid the Sorting Problem, namely, that the psychological theory we actually arrive at in a mature cognitive science \textit{must} consist entirely of universal principles of psychology. Pending some compelling reason for making this assumption, I conclude that \textit{de facto} necessitarianism does nothing to solve the Sorting Problem.

4.4. Conditionalization

It might be thought that the Sorting Problem can be handled without appeal to any sort of substantive \textit{a priori} methods by means of conditionalization. The idea would be that, rather than Ramsifying the raw conclusions of a posteriori psychological investigation, we would first make some (or, possibly, all) of the regularities conditional on background neural or historical states of the organism and background physical laws. For example, let \( \text{Cond} \) be the set of background conditions relevant to the (above posited) law of categorization, BCE. Conditionalizing on \( \text{Cond} \) will yield the following principle \([\text{BCE}^*]\):
[BCE*]  Given Cond, if \(x\) is presented with a group of objects \(\{o_1, \ldots, o_n\}\) from a given category \(C\) for perceptual inspection, then if \(C\) is a “basic” category \(x\) will learn \(C\) more rapidly without help than if \(C\) is a superordinate category.

The claim is that principles such as [BCE*] can be included in the Ramsified definitions without loss since organisms differing from us with respect to the background conditions will satisfy [BCE*] vacuously, in virtue of making the antecedent of the conditional false and [BCE*] will be necessary for all organisms that do satisfy the background conditions. As a consequence, organisms that differ from us with respect to contingent background conditions are not arbitrarily excluded from satisfying the base psychological theory \(P\).

Unfortunately, conditionalizing presents the Ramsifying functionalist with a dilemma: either she must conditionalize all the principles provided by empirical psychology (in which case her definitions will be open to counterexamples) or she must conditionalize only some of them (in which case she must have some substantive a priori means for sorting the principles).

Once again, let us take \(P\) to be the psychological theory arrived at by a posteriori psychological investigation, \(U\) to be the universal component of \(P\) and \(H\) to be the contingent component of \(P\). On the first horn of the dilemma, the Ramsifying functionalist conditionalizes every principle in \(P\) on the relevant physical, historical, and nomological background conditions. The resulting Ramsified theory can be written schematically as follows: \((\exists R)(\text{Cond} \rightarrow P[R]).\)

But this way of incorporating background conditions into the Ramsified definitions faces a further fatal dilemma. On the one hand, we could allow for vacuous satisfaction of the background conditionals. This allows nonhuman Martians to vacuously satisfy the principles of human psychology. However, since we have conditionalized every principle in \(P\) on background conditions, it allows all (or, at least, many)

---

26 Of course, one does not conditionalize the entirety of \(P\) on a uniform set of background conditions. Rather, the schema should be understood as the conditionalization of a representative principle from \(P\) on background conditions.
nonhuman, *nonsentient* objects to satisfy the Ramsified definitions!\(^{27}\) Consequently, allowing for vacuous satisfaction of the background conditionals yields Ramsified definitions that are not sufficient for instantiating the associated mental properties. This suggests (and this is the other horn of the second dilemma) that we should disallow vacuous satisfaction of the background conditionals. But if we do this, then sentient, nonhuman Martians will *fail* to satisfy the Ramsified definitions and we will have a counterexample to the necessity claim.

It looks, therefore, as if universally conditionalizing \(\mathcal{P}\) on background conditions is unworkable. What is needed is a way of conditionalizing some (i.e., the ones governing human psychology), but not all, of the principles in \(\mathcal{P}\).

This brings us to the second horn of the first dilemma. What we want are Ramsified definitions having the following schematic form: \((\exists R)((U \& \text{Cond} \to H)[R])\). Unfortunately, we must now decide which principles to conditionalize and which to leave alone. Specifically, if we conditionalize too many of the principles of universal psychology, we run the risk of generating Ramsified definitions that are too permissive and, hence, generate counterexamples to the sufficiency claim; similarly, if we fail to conditionalize enough of the principles that we intuitively take to be principles of human psychology, we run the risk of generating Ramsified definitions that are insufficiently general and, hence, generate counterexamples to the necessity claim. Thus, in order for the conditionalization strategy to work, the Ramsifying functionalist requires a *reliable* way of separating the universal principles of psychology from the contingent ones.\(^{28}\) Thus, conditionalization does not provide a satisfactory formal solution to the problem of psychological variability.

\(^{27}\) Let \(r\) be an n-tuple of arbitrary first-order physical geological properties (e.g., being composed of 67.5% quartz). Then, since Cond and \(r\) never coinstantiate, \((\text{Cond} \to \mathcal{P}[r])\) is vacuously true. But then, given the associated Ramsified definitions, any rock instantiating one of these states, \(r_i\) will have the associated mental state \(m_i\).

\(^{28}\) Of course, there will always be formal mechanisms that might *by pure chance* yield the correct separation of principles. For instance, we could randomly assign psychological principles to either \(U\) or \(H\) and the result *might* yield correct Ramsified definitions. It is clear, however, that any such process could never yield definitions that we were justified in accepting. In order for us to be justified in accepting the definitions, the sorting process must (at a minimum) be a reliable one.
5. Conclusion

According to functionalism, mental properties are individuated by way of their causal role. But this way of putting things is misleading, for it suggests that the causal relations relevant to individuating mental properties can be identified with the set of causal relations those properties in fact exhibit. Put somewhat differently, the claim that mental properties are individuated by their causal role suggests that the causal role of the mental properties “falls out” of empirical science. This way of thinking about things, however, is wrong and leads to a form of psychological chauvinism that is every bit as objectionable as the matter chauvinism which motivated philosophers to adopt functionalism in the first place.

In order to avoid this problem of psychological chauvinism, functionalists are forced to Ramsify only over universal, necessary truths of psychology. In this paper, I have argued that this constraint causes a serious problem for a posteriori functionalists: either they must find some formal mechanism for necessitating the appropriate laws of cognitive psychology or they must rely on substantive a priori methods (e.g., rational intuition) for this purpose. I have argued, however, that there is no plausible formal mechanism to which the a posteriori functionalist can appeal. Moreover, use of substantive a priori methods in deciding which laws to use in the base theory of our Ramsified definitions is incompatible with the a posteriori standing of those laws. Thus, if functionalism is defensible at all, it must be some form of a priori functionalism.  

References


29 It might be thought that various thought experiments show that there are no necessary psychological principles whatever; for instance, one might try to press Lewis’s examples of mad pain into service of this claim (1980). If such a case can be made, it would show that functionalism (as a general thesis) fails tout court. Nevertheless, I am doubtful that such an argument can be made rigorous and would, in any event, be far beyond the scope of this paper.
Against a posteriori functionalism


Against a posteriori functionalism


Biographical Notes

Marc Moffett is Assistant Professor of Philosophy at the University of Wyoming. His primary interests are in the philosophy of language and epistemology. He is co-editor of Know-how: Essays on Knowledge, Mind and Action (Oxford, forthcoming). Recent publications include, “Radical Intellectualism” (Know-how, forthcoming; with John Bengson), “The Folk on Knowing How” (Philosophical Studies, forthcoming; with John Bengson and Jen Wright), and “Reasonable Disagreement and Rational Group Inquiry” (2007, Episteme).