

Functionalism, Mental Causation, and the Problem of Metaphysically Necessary Effects¹

Robert D. Rupert, Texas Tech University

The recent literature on mental causation has not been kind to nonreductive, materialist functionalism ('functionalism', hereafter, except where that term is otherwise qualified). The exclusion problem² has done much of the damage, but the epiphenomenalist threat has taken other forms. Functionalism also faces what I will call the 'problem of metaphysically necessary effects' (Block, 1990, pp. 157-60, Antony and Levine, 1997, pp. 91-92, Pereboom, 2002, p. 515, Millikan, 1999, p. 47, Jackson, 1998, pp. 660-61). Functionalist mental properties are individuated partly by their relation to the very effects those properties' instantiations are thought to cause. Consequently, functionalist causal generalizations would seem to have the following problematical structure: The state of being, among other things, a cause of \underline{e} (under such-and-such conditions) causes \underline{e} (under those conditions).³ The connection asserted lacks the contingency one would expect of a causal generalization. Mental states of the kind in question are, by metaphysical necessity, causes of \underline{e} ; any state that does not cause \underline{e} is thereby a different kind of state. Yet, a mental state's being the sort of state it is must play some causal role if functionalism is to account for mental causation.⁴

In what follows, I first articulate more fully the problem of metaphysically necessary effects. I then criticize three functionalist attempts to solve the problem directly. Given the failure of functionalist efforts to meet the problem head-on, I consider less direct strategies: these involve formulating functionalism or its causal claims in such a way that they appear not to generate the problem of metaphysically necessary effects. I argue against these indirect solutions, in each case concluding either that the problem still arises or that avoiding it requires the adoption of an unorthodox form of functionalism (itself a surprising result). In the final

section, I advocate a more straightforward solution to the problem: Because of the backward-looking manner in which causal-historical theories characterize mental properties, they are not subject to any version of the problem of metaphysically necessary effects, and this constitutes a reason to favor such theories over conceptual- or functional-role accounts of mental properties.

II. Functionalism and metaphysically necessary effects

Let us begin with an attempt simply to see how mental properties and their instantiations could fit into a material universe, absent a reduction of mental properties or states to physical ones, and how they might so fit in a way that allows for genuine mental causation. According to one prominent physicalist view, token physicalism, every event (or state—let this be understood hereafter) is a physical event, in the sense that each event in the universe instantiates at least one physical property (or has at least one true physical description). Token physicalism thus implies that every mental event (i.e., every event that instantiates at least one mental property) is a physical event.⁵

We now see how a mental event might possess causal power: on the token physicalist view it can possess causal power at least insofar as the physical event to which it is identical possesses causal power. This, however, leaves unanswered the question whether token physicalism allows mental properties to have distinctive causal efficacy or whether, vis-à-vis causal interactions, only events' physical properties matter. Many philosophers have been struck by the fact that, with respect to at least some causal transactions, not all of an object's properties are causally efficacious (Dretske, 1988, pp. 79-80, LePore and Loewer, 1987, p. 633): a large, smooth rock may crush a walnut when dropped upon it, but surely, in the typical case of this sort, the largeness of the rock, not its smoothness, accounts for the nut's collapse. Thus the contemporary epiphenomenalist malaise: Might it be that events' mental properties always play a

role analogous to the rock's smoothness, present but not causally relevant?⁶ Terence Horgan nicely expresses this as a concern about mental 'quausion', causation qua an event's mental properties (Horgan, 1989). It is one thing to say that a mental state causes behavior; it is another to say that the state causes the behavior in virtue of the state's mental properties.⁷

Functionalism, despite its status as a leading conception of mind, has particular difficulty accounting for mental quausion, partly because of the problem of metaphysically necessary effects. To see why, we must first get clear about the nature of functionally individuated mental states and the corresponding properties. To be in a functionally individuated mental state is to be in a state that figures into a certain pattern of causal relations involving inputs, outputs, and other mental states. Besides its being too rough, this formulation fails to capture clearly one of the central traits of functionally individuated states: that they are multiply realizable. Better this formulation, then: A system is in a given functional mental state if and only if it is in some realizer-state or other that plays the relevant causal role, i.e., that bears the right causal relations to the realizers of the appropriate inputs, outputs, and other mental states. Realizer states are usually thought of as physical states, but all that is necessary to a robust functionalism is that a realizer not itself be a functionally individuated mental state (or be characterized indirectly in terms of a functionally individuated mental state) (cf. Shoemaker, 1981, pp. 94-95). For a system to instantiate a given functionally individuated mental property is for the system to be in some realizer-state or other (or to instantiate some realizer property or other) that plays the mental property's individuating causal-functional role (the role that individuates the mental state corresponding to that mental property). If, for instance, the functional characterization of pain includes its causing aversive behavior, then attributing pain to a creature is to attribute to the creature some state or other that causes aversive behavior (under the right circumstances, where

such circumstances would include, for instance, the absence of a strong desire to persist in the task at hand). The nonmental (most likely neural) state that plays the causal role in question realizes pain. Instantiation of pain amounts to there being some realizer or other of the functional-role state of being in pain.⁸

Functional-role properties, so construed, appear to be causally irrelevant. Although explanation is not ultimately the issue, try explaining a subject's aversive behavior by appeal to her instantiation of the property of being in pain: the subject's being in some state or other that (among other things) causes aversive behavior caused her aversive behavior.⁹ The problem comes into focus more clearly when we consider the law of a supposedly autonomous empirical psychology meant to underwrite the single-case explanation: A subject that is in some state or other that causes, among other things, aversive behavior exhibits aversive behavior. Such a "law" appears vacuous, stating no more than that states that cause e cause e.¹⁰ Metaphysical necessity—not empirical law—grounds this claim. There being an instantiation of the property bachelorhood surely guarantees the instantiation of being unmarried, but this hardly suggests a causal relation. We should treat in the same fashion cases where properties are individuated by their realizers' causes and effects. Let lig be the property being in some state or other that causes bells to dax and let dax be the property being in some state or other caused to occur in bells by realizers of lig. It had better not be the case that either of these properties becomes causally efficacious simply because we have characterized it in terms of its realizers' causal relations.¹¹

Some philosophers have held that the laws of nature are metaphysically necessary (Swoyer, 1982, Fales, 1993, Shoemaker, 1998), and they, it would seem, will be bothered little by the problem of metaphysically necessary effects. Here is not the place for a full-scale critical discussion of the necessitarian view of natural law or of the causal theory of properties (CTP,

hereafter) that sometimes motivates it (and would seem to follow from it).¹² I hope, however, that the following remarks suffice to keep the present argument on track: First, we should want our theory of mental properties to remain, as much as possible, neutral with respect to such contentious metaphysical issues as the general nature of properties and laws. If nothing else, this essay should show that the functionalist's metaphysical options are surprisingly limited, perhaps to a necessitarian view, as opposed to other conceptions, of natural law.

Secondly, the view that causal necessity is a species of metaphysical necessity should not sit well with functionalists. The functionalist conception of mind gains much of its plausibility from a certain computationalist picture. Functionalist mental states are grounded in realizer states whose causal interactions mirror the relations between mental-cum-functional properties. Without the anchor provided by realizer-states, functionalist theorizing would seem to float freely (cf. Fodor, 1981, pp 12-14), but the necessitarian view of natural law threatens to pull the anchor. Here is how: Assume that causal laws are relations between properties. If causal laws hold necessarily, then properties necessarily enter into all of their causal relations. If we add the assumption that no two properties enter into the same range of causal relations, we have a quick route from the necessitarian view of natural law to the CTP. On the most straightforward interpretation of the CTP, though, no realizing structure adequately grounds—i.e., accounts for the causal powers of—functional-role mental properties: the CTP offers nothing more than further sets of ungrounded functional-role properties. Any given property is either nothing more than a set of relations to other properties that are nothing more than sets of relations or is individuated in terms of realizers that are nothing more than sets of relations to other properties that are nothing more than sets of relations, and so on—so that each property is either a set of ungrounded relations to other sets of ungrounded relations or is ultimately grounded in a set of

ungrounded relations to other sets of ungrounded relations. If the plausibility of philosophers' talk of functional-role mental properties presupposes the grounding of those properties in realizer-structures that possess something more than relational natures—if such grounding is part of what makes the functionalist picture of mind attractive—then the CTP is bound to disappoint.

Thirdly, and of most importance in the present context, the CTP fails to distinguish genuine properties from sham properties in a way that preserves the plausibility of functionalism. If only differences in causal relations distinguish one property from another, then on what basis can the CTP exclude from legitimacy mere-Cambridge properties, gruesome properties, and the whole lot of ligs and daxes? Sydney Shoemaker, the foremost proponent of the CTP, claims that sham properties are causally derivative (1998, p. 65): if their instantiations have any effects at all, these effects are parasitic upon the effects of the properties in relation to which terms for the sham properties are defined (grue's effects being parasitic upon green's and blue's, for example). This, however, is unsatisfactory in the present context, for functionalist properties seem to gain their effects in just such a parasitical manner: they are individuated by reference to the effects of something else, their realizers. If sham properties are those the “effects” of which derive by definition from the effects of other properties, functionalist mental properties are sham properties of the highest rank.

How serious is the problem of metaphysically necessary effects? At the outset, I noted the extent to which recent discussions of functionalism and epiphenomenalism have focused on the exclusion problem. Why, then, worry much about the problem of metaphysically necessary effects with the seemingly insoluble exclusion problem looming? No doubt, the exclusion problem is quite serious; furthermore, in at least one respect, it is more general than the problem of metaphysically necessary effects: the exclusion problem arises with respect to all second-order

properties, not merely those that are functionally individuated (where a second-order property is the property of instantiating some property or other among a delineated set of properties—see Kim, 1998, p. 20, on second-order properties, and pp. 55, 77-87, on the scope of the problems under discussion).

Nevertheless, there is another respect in which the problem of metaphysically necessary effects is more serious than the exclusion problem: A number of proposals offered in response to exclusion-based worries fail to solve the problem of metaphysically necessary effects, at least when these proposals are set, as they sometimes are, within a functionalist framework. It is possible, for example, that an appeal to the part-whole relation solves the exclusion problem: perhaps a token mental state consists in the part of that state's realizer that is common to the corresponding mental state type's various realizers (Shoemaker [2001, pp. 80-81] and Stephen Yablo [1997, p. 257] have made suggestions along these lines). The relation between whole and proper part is not a relation of two independent candidates for causal responsibility and thus does not give rise to the exclusion problem; if an effect is due to the mental part of the realizer state, there is bona fide mental causation, even though there is still a clear sense in which the realizer state was causally sufficient for the effect in question. I omit detailed discussion of Shoemaker's and Yablo's rich proposals, hoping only to convince the reader that, although the part-whole strategy offers the functionalist some hope against the exclusion problem, it falls prey to the problem of metaphysically necessary effects: "The part of a brain state that is the part it is because it causes B, causes B" engenders the problem of metaphysically necessary effects to the same extent as versions of functionalism according to which realizer-states compete with functional states for causal efficacy.¹³ An advocate of the part-whole strategy might avoid this result by excluding effects from the individuation conditions of mental properties (thereby

excluding effects as a factor determining what counts as the common part of various realizer states), but that would be to reject functionalism.

David Robb's trope-based view fares no better when combined with the functional individuation of mental properties. According to Robb, although mental properties might be multiply realizable, each trope of a mental property is identical to a trope of a physical property (Robb, 1997, pp. 187-88); and it is the tropes that do the causing. Robb sees it as an advantage of his approach that it solves the exclusion problem within the functionalist framework (Robb, 1997, p. 190): identicals do not compete for causal efficacy. It follows from Robb's view, though, combined with the functionalist individuation of mental properties, that a physical trope is a trope of mental property \underline{F} only if it has certain effects. Thus, the trope-based view does not shake the problem of metaphysically necessary effects. A trope's being a trope of one mental property rather than another contributes nothing contingent to the causal interactions into which that trope enters. To say that a trope of type \underline{F} (which is also a physical trope) caused \underline{e} , and thus that the \underline{F} -ness is causally relevant to \underline{e} , is to say that a trope of the kind that causes \underline{e} caused \underline{e} . If the trope in question were not to have caused \underline{e} , it would not have been a trope of type \underline{F} .

Two points to keep in mind, then: (1) the respective scopes of the exclusion problem and the problem of metaphysically necessary effects are not such that one properly includes the other and (2) the most promising proposals for solving the exclusion problem in a way that is consistent with functionalism will not help the functionalist to get around the problem of metaphysically necessary effects. Thus, there seems no reason to privilege the exclusion problem and to view the problem of metaphysically necessary effects as any less pressing, at least if our concern is to see how functionalism might account for mental causation.

II. Causality and regularity

The functionalist might hope to account for quausality by appealing to an established theory of causality. Consider one venerable approach, the regularity theory. According to this view, an M-event causes a B-event if and only if events that instantiate property M are lawfully sufficient for events that instantiate B (or, if and only if that an M-event occurs lawfully entails that a B-event occurs). One might conceive in Humean terms of the covering laws involved, such that Ms' being lawfully sufficient for the occurrence of Bs consists in no more than that Ms are always followed by Bs. (I ignore many niceties because, for reasons that should presently become clear, the consideration of various refinements [e.g., as one sees discussed in Kim, 1973b] will not help the regularity theorist to establish the quausal power of functionalist mental properties.) Given, then, that instantiations of functionally individuated mental properties regularly precede instantiations of their metaphysically necessary effects, the former cause the latter, at least according to the simplest versions of a regularity-based view of causation. The requirements for quausality are then met regardless of what other properties are exemplified by events that instantiate M and B; Ms cause Bs and, therefore, are as causally relevant to Bs as one could wish that Ms be.¹⁴

Functionally defined properties appear to cause their metaphysically necessary effects, when causation is understood in keeping with the regularity theory. The state of being in some state or other that causes B regularly precedes B. Notice, however, the presence of the word 'causes' in the preceding sentence. Discharging it in accordance with a regularity-based analysis reveals an empty claim: to say that a functional property quauses its effects is to say something of the form "Instantiations of the property of being in some state or other that regularly precedes instantiations of B, regularly precede instantiations of B." One can qualify 'regularly precedes'

in whatever way one likes (requiring, for example, spatial contiguity of the causal relata); given, though, that the same qualifications must attach to both instances of 'regularly precedes', such refinements yield an empirically vacuous claim. If the regularity theorist insists that patterns thus described are causal patterns, so much the worse for a regularity-based view. The regularity view must in some way account for the noncausal nature of law-like patterns that hold in virtue of mathematical, logical, analytical, and supervenience-based truth; and although I am, admittedly, pressing an intuitive judgment about similarity, the form of the allegedly causal claims at issue suggests that any plausible version of the regularity theory should treat the patterns described by such claims in the way it treats these other, noncausal relations.

Consider a related objection that Block raises to the regularity-based defense of functionalist mental causation. Block recognizes that instantiations of functional properties and their alleged effects are covered by some sort of regularity, and he acknowledges that one might offer a causal-nomological account of these regularities (i.e., that one might count these as causal regularities). He notes, however, that there is an obvious alternative: that the regularities in question are grounded in logical relations (Block 1990, p. 158; cf. Kim 1973a, 1974). All widows have dead husbands. That is a regularity one could confirm by empirical means, were one so to choose, but doing so would not show that the regularity is part of the causal-nomological structure of the universe. It is obvious how to account for the regularity of all widows' having dead husbands: having a dead husband is what it is to be a widow. According to Block, the mere presence of a logical relation that holds between M and B does not absolutely preclude there being a causal relation between Ms and Bs, but the two kinds of relation co-occur only in very special cases (Block, 1990, pp. 157-58). In contrast, in typical cases the realizer of a functionally defined property causes the effects in terms of which that property is defined, and

the functionally defined property is related to those effects only by logical necessity. Insofar as there exists a regularity involving the functional property, it is fully grounded in logical relations; invoking causal-nomological relations is superfluous. If As are things such that in all possible worlds, if something does not cause a B, it is ipso facto not an A, then of course As regularly precede Bs in the actual world. It is gratuitous to invoke causal connections when logical connections obviously account for the patterns at issue. Thus, the appeal to a regularity-based theory of causality does not save functionalist mental quausion.

Focusing specifically on mental content properties, Block claims that one can escape the present problem either by accepting a counterfactual theory of causal relevance or by abandoning the functional-role approach to content. In due course, I entertain both suggestions, rejecting the former and accepting the latter. First, though, let us consider a different, and prima facie quite promising, defense of functionalist mental quausion: that instantiations of functional-role properties cause effects that are not individuating of those functional-role properties.

III. Realization-indifferent regularities

Louise Antony and Joseph Levine distinguish three sorts of regularity involving functionally defined states, the existence of only one of which would secure a quausal role for functionally individuated mental properties (Antony and Levine 1997, p. 92). At one extreme, Antony and Levine identify regularities that attach to specific realizations of a given functionally defined property. As pain is realized in humans, it may cause wincing; as realized in dolphins, a high-pitched screeching. Such effects do not underwrite any claim to the causal powers of pain, qua pain, because they vary from realizer to realizer.¹⁵ At the other end of the spectrum, Antony and Levine place regularities grounded in definition, which they call ‘analytic regularities’. They reject such regularities for reasons similar to those that motivate Block. We might,

Anthony and Levine imagine, give a functional specification of dormativity: it “is a property a substance possesses if and only if it possesses some first-order physical property that induces sleep” (Anthony and Levine, 1997, p. 92). We have not thereby located a causally efficacious property, however; we have merely stated an analytic truth (and note that, given the way the example is put, Anthony and Levine’s concern can just as well be understood as a worry about metaphysically necessary effects).

Anthony and Levine hope to show that mental properties possess distinctive causal powers by demonstrating the existence of regularities of a third kind: realization-indifferent regularities, which hold of various realizations of a given functionally defined property yet the holding of which is not part of the functional specification of that property. Dormativity, qua dormativity, does not cause sleep in any interesting sense. If, though, being asleep at the wheel causes car wrecks, then dormativity has a distinctive causal power: it causes car wrecks. Causing automobile accidents is not part of the analytical specification of dormativity (nor, to bring Anthony and Levine’s discussion into closer contact with the present one, is it metaphysically individuating of dormativity), yet events that instantiate dormativity seem to stand in a law-like relation to car accidents (Anthony and Levine, 1997, pp. 92-93).

As promising as this tack may seem, it will not secure the quausal efficacy of functionally individuated properties. If the analytic (or metaphysically necessary) connection of dormativity to sleep precludes their causal connection, then so does it stand in the way of dormativity’s causing car accidents; for dormativity causes car wrecks only insofar as it causes drivers to fall asleep. Put in more general terms, one would expect the relations in question to follow this pattern: e₁ causes e₂ which causes e₃; e₁ exemplifies property A, instantiation of which causes the instantiation of property B in e₂; B’s instantiation in e₂ in turn causes the instantiation of C in e₃

(where a property-instantiation's causing the instantiation of another is equivalent to the former's quausing the latter). One might even think it a requirement of quausal relations that they respect in this way the transitivity of the causal relations between events that instantiate the relevant quausally related properties. (This requirement would only apply where transitivity in fact holds among the underlying causal relations, as in the present case; my point here does not depend on causal relations' being generally transitive.) If e_2 instantiates a property B that causes the instantiation of C (in e_3) and e_1 causes e_2 , then if e_1 instantiates no property causally relevant to B , then e_1 instantiates no property that quauses C absent some independent causal route to C 's instantiation in e_3 . Quausation is a kind of causal influence, after all. It is ad hoc to deny the need for quausal order where there is clearly causal order among the events that instantiate the alleged quauses.¹⁶

We should grant the possibility of causal intermediaries in the absence of quausal ones. Perhaps a certain property-instantiation remains constant throughout a causal interaction yet later has quausal impact: event e_1 causes e_2 , which causes e_3 ; e_1 and e_2 both instantiate A , which then causes the instantiation of B in e_3 . The causal order might, in such cases, be richer than the quausal one, but this is not the sort of case at hand. Antony and Levine posit a much more radical difference between the causal order and the quausal one. On their view, the cause-relation is a total order while the quause-relation is partial, at least with respect to the small sets under consideration: A quauses C ; B quauses C ; neither A quauses B , nor vice versa; yet there is only one totally-ordered underlying causal chain. This deep structural difference in the quausal and causal orders seems motivated solely by the hope of solving the problem of mental causation.

Let me put this concern less formally. Assume that A's instantiation causes C's instantiation and B's instantiation causes C's instantiation, yet there is no quausal connection between A and B. Perhaps this alone is no problem; there might be more than one cause of a given property-instantiation just as there is often more than one cause of a single event: a sedentary lifestyle and an unhealthy diet each can contribute causally to a single heart attack. Notice, however, that in such cases of multiple causal factors, there is, presumably, a distinct causal chain leading from each independent factor to the effect of interest. In contrast, in the kind of quausal case under consideration, there is an unmotivated leap-frogging of quausal influence from A to C. It is unmotivated in Antony and Levine's case precisely because we can see quite well how the events that instantiate the relevant properties are connected in the causal order. A good deal of the plausibility attached to Antony and Levine's example derives from the idea of a transitive chain: dormativity quauses wrecks because of its relation to drivers' being asleep at the wheel, which itself quauses wrecks. Sever the quausal connection between dormativity and sleep and we seem to have stripped the example of its force.

Moreover, Antony and Levine's case exemplifies a structure typical of psychological explanations. Consider a simple example. Joe comes to possess a bag of throat lozenges. Why? Because he wanted throat lozenges and this caused him to go to a store where they were available. Joe's wanting is parallel to the role of dormativity in Antony and Levine's example, and his trip to the store plays the role of a driver's falling asleep. Antony and Levine's position entails that Joe's desire for the medicine does not quause Joe's store-directed behavior, because of the way the two are analytically connected (at least on the plausible assumption that the functionalist specification of the property of wanting throat lozenges implies, under circumstances such as Joe's, at least an attempted trip to the store—other parts of the world must

cooperate if he is to arrive). Yet, Joe's trip to the store is undoubtedly causally relevant to Joe's later possession of throat lozenges (analogous to a car wreck in Antony and Levine's example). Is there any reason, other than a hope to save functionalist mental causation, to think that Joe's desire causes the procurement of lozenges but not via his trip to the store? I see none.

Antony and Levine propose what seems to be the most promising way a functionalist might confront squarely the problem of metaphysically necessary effects and still make way for a reasonably robust psychological theory. Strikingly, their approach does not pan out. On Antony and Levine's view, causal intermediaries facilitate the instantiation of realization-independent psychological regularities, while inexplicably, quausal intermediaries are unnecessary even in cases where there are obvious candidates; thus both the functionally defined state and what we would normally think of as the intermediary state independently quause (or contribute quausally to) the effect in question, but without the transitive structure that gives cases like Antony and Levine's what intuitive appeal they seem to carry and without any plausible independent route of quausal influence. If this be the price of functionalist mental causation, we ought to look elsewhere either for a theory of mind or a theory of causal relevance.¹⁷

IV. A counterfactualist defense

Perhaps a shift in focus, from regularities to claims about counterfactual relations, will help the functionalist to solve the problem of metaphysically necessary effects and locate properly the causal efficacy of functionally individuated properties. Indeed, as the primary positive message of "Can the Mind Change the World?", Block conditionally endorses a counterfactual-based account of causation: "Here, as earlier in the paper, the lesson is that if you want to avoid epiphenomenalism, go for a counterfactual theory of causal relevance, not a

nomological theory. This is, I suppose, the main positive point of the paper...” (Block, 1990, p. 159).

Block reaches this conclusion largely because he thinks an appeal to counterfactuals establishes the causal relevance of functional properties with respect to effects in terms of which such properties are individuated.¹⁸ If one integrates Block’s conclusion and Antony and Levine’s emphasis on realization-indifferent regularities, Block’s argument promises an important payoff beyond the vindication of analytically specified causal relations.¹⁹ The functionalist’s hope for a substantive and autonomous cognitive psychology rests partly on the conviction that there are realization-indifferent regularities of the sort described by Antony and Levine, but as argued above, the causal inefficacy of metaphysically necessary connections proscribes the existence of such realization-indifferent causal regularities. If, however, Block’s argument establishes the counterfactual relevance of functionally characterized properties with respect to the effects in terms of which those properties are individuated, then the path to realization-indifferent regularities opens. If dormativity quauses sleep after all, there is nothing to interrupt the quausal chain via which dormativity might quause car accidents, and the nonreductive functionalist will have located a rich quausal role for functionalist mental properties.²⁰

Consider the following counterfactual-based argument commonly used to motivate epiphobic hand wringing: the content of a mental state is not causally relevant to behavior because, were the realizing state, say, a neural state, to lack the content property in question, that realizer-state would nonetheless cause whatever behavior is in question.

Block responds by denying the counterfactual premise, at least for the case of metaphysically necessary effects. Functionalist content properties are individuated by the effects

of the states (or properties) that realize them. Thus, a neural state will not have the behavioral effects in question in the nearest possible worlds in which it lacks the relevant actual-world content property, at least if such effects are among those in terms of which the relevant functional-role content property is individuated. As Block puts it, “For if it had had a different meaning, or no meaning at all, its functional role would have been different, and since functional role is causal role, abstractly construed, a difference in functional role typically will include a difference in behavioral effects” (1990, p. 151).

Block’s argument faces at least two serious objections. Horgan’s appeal to pertinently similar worlds (1989, p. 58) as a way to ground claims of mental causation suggests the first. For the purpose of evaluating counterfactual claims about events \underline{c} and \underline{e} , Horgan considers worlds that contain events similar to \underline{c} or \underline{e} , without requiring that these events bear a relation of transworld identity to \underline{c} or \underline{e} . Pertinent similarity involves only a weak counterpart relation that tolerates divergence among essential properties. Allowing this divergence threatens Block’s argument, for it allows the consideration of a counterfactual event pertinently similar to the relevant real-world cause-event that (1) has the behavioral effects in question yet (2) lacks the overall conceptual role that determines the content of the real-world cause-event. Worse still for Block’s argument, the existence of such an event in a nearby possible world does not seem at all implausible. One might imagine a state that has the relevant behavioral effects of the actual-world realizer-state but is not caused by an input metaphysically necessary for the instantiation of the functional state in question (lacking its input relation, it has different content, but the resulting behavior is the same). Or one might imagine a state that has ninety-nine percent of the metaphysically necessary effects of the functionally defined content of the state in question, where the missing one percent seems entirely irrelevant to the token behavior to be explained

(the missing one percent might involve, for example, the effects a state with that content would have in circumstances quite different from those under consideration).

Block's counterfactual-based argument also must overcome an objection analogous to the one Block himself presses against the regularity-based defense of functionalist mental quausation. According to Block, a functionalist who embraces the regularity theory of causation can offer only an otiose account of the typical regularities involving metaphysically necessary effects: these regularities are unquestionably grounded in logical truth, and thus the assertion of causal status is gratuitous. A similar complaint holds against Block's counterfactualist tack. Let us assume a functionalist account of content and grant Block's counterfactual claim: if a content-bearing state were to lack its content properties, it would lack certain effects as well. This is so, however, only because, if a functionalist account of content is true, the content-bearing state has the effects in question by logical (or metaphysical) necessity: if the state were not to have those effects, it would not instantiate the relevant content-property. Hence, there exists a correct, noncausal explanation of the truth-values of the relevant counterfactuals. Surely patterns of counterfactual dependency sometimes constitute quausal connections, but in the cases of present concern, these patterns clearly manifest logical relations or relations of metaphysical necessity; thus, Block's causal gloss of such counterfactual dependencies is superfluous and should be rejected.²¹

One might worry that a certain pattern of truth-values of counterfactuals involving As and Bs is simply what it is for As and Bs to be causally (or quausally) related. Given, however, various problems faced by the counterfactual analysis of causality (see, for example, Kim, 1974, Hall, 2000), we should think that a pattern of counterfactual relations is no more constitutive of causality than is a law-like regularity. Certain patterns of counterfactual truth and falsehood are

grounded in, and we should normally think diagnostic of, causal relations, but some counterfactual dependencies result from or are indicative of noncausal facts; supervenience, for example, also involves a counterfactual-supporting form of dependency, but it is rarely taken to be a causal relation. In the cases at hand, then, it is overkill to invoke causal—or quausal—relations as the ground of the relevant counterfactual dependencies.²²

V. Assorted functionalist maneuvers

Regularity- and counterfactual-based attempts to solve the problem of metaphysically necessary effects appear not to pan out. In contrast to such direct responses, a functionalist might essay the strategy of dissolution. In this section, I consider three kinds of maneuver, each of which is intended to defuse the problem of metaphysically necessary effects by showing that functionalism or its causal claims, when properly formulated, emerge unscathed. In the end, the most any of these functionalist moves will show is that the problem of metaphysically necessary effects limits the acceptable forms of functionalism to a surprising heterodoxy.

A. Causality and causal relevance:

I have argued that what are supposed to be functionalist causal generalizations are not causal generalizations at all. The form of such generalizations seems to make the problem especially clear: It is empirically empty to claim that the state of being in some state or other that causes e, causes e. In response, though, the functionalist might argue that I have gotten wrong the form of the relevant generalizations. She might insist instead on a gap between the causal characterization of a functional state and the functionalist's quausal claim. The functionalist might agree that mental properties are individuated causally, i.e., in terms of the causal roles of their realizers, but at the same time claim that the relevant connection between second-order functional properties and the relevant aspects of their “effects” is one of causal relevance, not

causality itself. The form of the functionalist's now-quausal claim becomes this: The state of being in some state or other that causes \underline{e} is causally relevant to \underline{e} 's instantiating \underline{G} . The idea would seem to be that it is one thing to cause \underline{e} , while it is another to be causally responsible for the existence of a \underline{G} -event. In order for a state to realize functional property \underline{F} , it must cause \underline{e} , but perhaps the fact that \underline{F} is instantiated specially accounts for the \underline{G} -aspect of \underline{e} (cf. LePore and Loewer 1987, 1989, McLaughlin, 1989, Horgan, 1989).

Such a move does not seem promising, however. Criterial to state \underline{c} 's being a realization of \underline{F} is that \underline{c} cause an event \underline{e} that instantiates \underline{G} . Event \underline{e} instantiates \underline{G} essentially, at least with respect to \underline{c} 's being a realization of \underline{F} . Thus, the functionalist proposal at hand would more honestly be put the following way: The state of being in some state or other that causes (among other things) a \underline{G} -event is causally relevant to that event's being \underline{G} . Despite the nominal difference between causality and causal relevance, there would appear to be nothing offered by the claim of causal relevance over and above what is metaphysically necessitated by \underline{F} 's instantiation. The instantiation of \underline{F} is causally relevant because it guarantees the occurrence of a realizer-event that causes a \underline{G} -event. It is one thing for the functionalist to avoid the appearance of redundancy in the form of her causal claims (or laws). It is another still for such claims to aver the metaphysically contingent.

B. Ceteris paribus clauses

Mental states can be instantiated without the occurrence of what I have been calling their 'metaphysically necessary effects'. One can be in pain and not exhibit aversive behavior (one might be incapacitated); one can want beer and have the belief that the refrigerator holds beer but, nevertheless, avoid the refrigerator (one might be on a campaign of abstention). Surely

functionalism can account for such cases; will not its doing so, however done, provide the functionalist with a solution to the problem of metaphysically necessary effects?

Three possibilities deserve consideration: First, there are cases of psychological interference, where the existence of one mental state is accompanied by the existence of another whose realizer prevents the typical effects of the first (cf. Horgan and Tienson's discussion of soft laws: 1996, chapter 7). Second, there are cases of physical interference, where a realizer-level mechanism malfunctions or the realizer state is somehow prevented by the instantiation of other first-order properties from having its standard effects (e.g., an area of the subject's brain suddenly hemorrhages). Third, there is a peculiar sort of case where a functional-role property is instantiated and its metaphysically necessary effect occurs, but that effect is caused by neither the realizer nor the instantiation of the functional-role property in question.

With respect to each kind of case, we should keep the following point in mind: The effects of the realizer of a given functionalist mental state are specified relative to the circumstances in which that mental state obtains. That certain effects occur in a particular milieu constitutes part of the individuation conditions of a functional-role mental property. Specification of the relevant milieu typically refers to very many other mental properties, a point from which follows a quick response to the first kind of case: Standard functionalist doctrine explicitly and purposively accounts for the fact of psychological "interference." Pain causes aversive behavior, so long as one does not have a stronger countervailing desire, e.g., a desire to persist in the behavior that is exposing one to the source of tissue damage (see note 3). Thus, cases of the first kind provide to functionalism no relief from the problem of metaphysically necessary effects. The functionalist's causal generalizations are of this sort: States that cause e (under certain conditions) cause e (under those conditions). The clause "under certain

conditions” builds into the very individuation of the mental property the conditions under which e occurs, including conditions pertaining to the subject’s overall psychological profile.

The challenge posed by the second case is also met in a fairly straightforward manner, at least on the reasonable assumption that all token causal interactions are covered by laws (see below for discussion of why the functionalist should find this assumption especially attractive); for then (1) the psychological laws involved will be ceteris paribus laws, and (2) the characterizations of causal roles in the individuation of functional properties will themselves contain ceteris paribus clauses (because the causal interactions referred to in the specification of causal roles must be covered by laws). The characterization of pain will include something to the effect that, ceteris paribus, it causes aversive behavior (where the ceteris paribus clause will require that, among other things, the realizing system remain intact in at least certain respects). Functionalist causal laws will thus take the following problematical form: “The property of being in some state or other that, among other things, causes e (ceteris paribus), causes e (ceteris paribus),” the two ceteris-paribus clauses covering the same range of other things.

In the third kind of case, a functionally characterized property is instantiated and one of its metaphysically necessary effects, e, occurs, yet the realizer-state in virtue of which the functionally individuated property is instantiated does not cause e. This sort of case is, like earlier ones, supposed to speak against the claim that what I have been calling the ‘metaphysically necessary’ effects of the instantiation of a functional-role property are in fact metaphysically necessary; after all, the present reasoning runs, the relevant effect can occur independently of the functional-role property in question, even in a case where that functional-role property is instantiated.

Notice, however, that the present case does not differ in essence from the earlier ones. If \underline{e} occurs and was not caused by the realizer of the relevant functional-role state \underline{F} , then \underline{e} was, presumably, caused by something else that preempted or in some other way interfered with \underline{F} 's realizer's causing of \underline{e} . How a given instance of this third kind of case is to be handled depends, then, on how the realization of \underline{F} was prevented from causing \underline{e} . If a subject experiences a stroke that completely destroys her \underline{F} -state's realizer, but by some neurophysiological fluke, \underline{e} occurs anyway, this is to be handled in two steps: first by the preceding treatment of cases of the second kind, and then, by noting that there is nothing in the author's position precluding the existence of more than one possible cause of \underline{e} . If, on the other hand, a different psychological state causes \underline{e} , say by preemption of \underline{F} 's realizer, then it is a matter to be captured by true psychological theory. The Ramsey-sentence expressing true psychological theory will imply that, when present with \underline{F} , the state in question causes \underline{e} in a way that is independent of, and preemptive of, the \underline{F} -state. It is not clear what sort of case remains where \underline{e} is to be accounted for in some way other than (1) as a side-effect of a breakdown or (2) as the effect of a psychological state that appears in the Ramsification of psychology and thus whose role as a cause of \underline{e} (under certain conditions) is an implicit part of the individuation conditions of property \underline{F} .

C. Other forms of functionalism

Antony and Levine (1997, p. 92) suggest that a functionalist might avoid the problem of metaphysically necessary effects by embracing empirical functionalism. Antony and Levine themselves reject this functionalist move but not because it fails to solve the problem of metaphysically necessary effects. Furthermore, given that many philosophers, including Block (1997, p. 120), endorse psychofunctionalism over analytical functionalism (for discussion of various forms of functionalism, see Block, 1978, 1980a, Shoemaker, 1981, Braddon-Mitchell

and Jackson, 1999), it would seem worthwhile, in the present context, to pursue Antony and Levine's suggestion.

The pursuit is short, however. According to psychofunctionalism (as Block conceives of it—1978, p. 272), the proper characterization of a mental property emerges from the toil of empirical research and related theorizing. Nevertheless, psychofunctional properties are second-order properties whose individuation conditions are captured by Ramsey sentences (or machine tables). Empirical work guides one's choice of functional-role properties to characterize, but the properties' profiles are functional-role profiles all the same. A given psychofunctional property is the property of being in some state or other that fits a certain causal profile. Whether that profile was generated by conceptual analysis is beside the point: the causal role of the mental property's realizers individuates that property. Thus, psychofunctional properties face the same epiphenomenalist threat faced by functionally individuated properties that, in keeping with analytical functionalism, issue from philosophical analysis of our mental concepts.

There is a distinct view sometimes called 'empirical functionalism' according to which functional-role specifications are used to fix rigidly the reference of mental state terms to nonmental states that play, in the actual world, the causal roles in question (Braddon-Mitchell and Jackson, 1999, pp. 81-85; Braddon-Mitchell, 2003, p. 121). Those nonmental states can, however, play different causal roles in other possible worlds, yet still be designated by our mental state terms (referred to counterfactually, as it is sometimes said). Have we found a version of functionalism not subject to the problem of metaphysically necessary effects? It would appear not, for the view in question abandons the basic functionalist vision. According to what was supposed to be the essential functionalist insight, mental properties are, in some deep sense, functional-role properties. If functional-role descriptions are used merely as actual-world

handles to achieve reference to states that need not fit those descriptions, then in what sense are mental properties “really” functional-role properties? The view on offer departs so radically from the core functionalist claim that it hardly seems to offer the functionalist a way out of the problem of metaphysically necessary effects—even though it might provide a satisfying causal-descriptivist account of how the reference of mental states terms is fixed (but see Braddon-Mitchell and Jackson’s worries about chauvinistic consequences—1999, pp. 81-82).²³

A further functionalist variant worth considering offers probability-based characterizations of mental states’ causal-functional roles. It was common in the early days of functionalist theorizing to think of mental states as machine-table states. Machine-state transitions are sometimes characterized probabilistically, and thus we might also think of functionalist mental states as related probabilistically, as states of probabilistic automata (Putnam, 1967). (The case where one and zero are the only transition probabilities included on the table is the limiting case; thus, talk of probabilistic automata is the most general way to state machine-table functionalism.) On such a view, being in a particular functional state is a matter of being in some state or other that plays a particular causal role but a causal role characterized in probabilistic terms: given the context, the state is likely, to a certain degree of probability, to have certain effects. This seems to introduce a gap between probabilistic causal relations at the type-level and causal interactions on particular occasions.²⁴ At least in straightforward cases, whether a particular realizer-state of F contributed causally to the occurrence of e is a yes-or-no proposition, in contrast to the case of the probabilistic causal relations used to characterize kinds of functional states. “The state of being in some state or other that is likely to cause e caused e,” a particular functionalist causal claim might run, and this claim seems to be contingent in a way that previously considered functionalist causal claims have proven not to be.

My surrejoinder begins with a pair of preliminary responses: First, if a functionalist theory couched in probabilistic terms is to defuse the problem of metaphysically necessary effects, the probabilistic characterization of functional-role properties must not be a matter of epistemological limitation. If functional-role properties are, in reality, individuated by their deterministic causes and effects, it is of no present import that, because of our epistemic limitations, we must characterize them in terms of probable causes and effects in order, say, to make useful predictions. Probabilistic causation or state-transitions create a gap between probabilistic laws and token causal relations only if the probability involved is a metaphysical fact about the individuation of the functional-role properties involved.

Second, if the functionalist talk of probabilities is meant only to accommodate probabilistic relations at the lower level, it is not clear that probabilistic causal relations are genuinely constitutive aspects of functional-role properties themselves. If, for example, functionalists include probability specifications as a way to correct for the disruptive effects of unlikely physical interactions among realizers (and interactions between parts thereof), the probabilities would not seem to play a role in individuating functional-role properties; rather, the inclusion of nonzero, nonunit transition-probabilities would constitute a particular way of formulating the ceteris paribus clauses discussed above. The functionalist would be saying, in essence, that things that cause e when no unlikely lower-level interactions occur cause e when no unlikely lower-level interactions occur. In the token case, this amounts to saying, “A state that causes e when no unlikely lower-level interactions occur caused e in a case where no unlikely lower-level interactions occurred.”

Furthermore, the problem of metaphysically necessary effects continues to cause trouble, even for a functionalist theory cast in terms of irreducibly probabilistic causal relations. Notice

first that at the level of psychological law, the problem of metaphysically necessary effects undoubtedly persists; the functionalist's allegedly causal generalizations will be of the form "The state of being in some state or other that is likely (to a definite degree) to cause e is likely (to the same definite degree) to cause e ." Were a state not likely to cause e with the appropriate degree of probability, then it would ipso facto not instantiate the functional-role property in question. Functionalism now faces a dilemma: Either a token causal interaction must be covered by a causal law, or it need not be. Given that the functionalist generalizations in question do not state causal laws (or causal generalizations), they cannot cover token causal interactions; therefore, they cannot cover token causal interactions of the sort at issue, and thus, those token interactions are not causal. Now consider the dilemma's second horn: by hypothesis, causal interactions are fundamental in the sense that their occurrence does not require the existence of general laws (cf., Anscombe, 1971). This picture, however, jeopardizes the functionalist's autonomous psychological theory. What will be the kinds of said psychology? What will be the relation of psychology's causal laws? Some philosophers of science may wish to accord fundamental status to token causal interactions and to deflate the role of causal laws or generalizations. Functionalism, though, should have no truck with such a program, for it is part of the functionalist's very conception of mental properties that they are causal types—and causal types, i.e., types individuated in terms of causes and effects, have no integrity absent causal laws or generalizations.

In response, the functionalist might suggest that the individuation of mental kinds proceed by the grouping together of objects or events with the same irreducible causal propensities; fine, but (1) this is quite a metaphysical cost for the preservation of a functionalist theory of mind and (2) it remains unclear how helpful it is to construct a psychological theory

with laws of the form “Things of the group whose members all have a significant propensity toward the causing of e (under such-and-such circumstances) sometimes cause e (under those circumstances).” No single effect is metaphysically necessary, but as a psychological theory, functionalism seems decidedly unpromising.

VI. The causal efficacy of mental content and state-types

A distinction is commonly made between state-types, for instance, belief (the sort of state that combines with desire to cause action) and content properties, e.g., the property of having the content that Bush is president, which can be shared by various mental state types (Field, 1978, Fodor, 1986, p. 14, 1987, p. 17) and perhaps linguistic items. A functionalist theory of either aspect of mental states—the attitudinal element or the representational one—faces the problem of metaphysically necessary effects. Thus, we should hope to find an alternative account of each. In the remainder, I briefly remark on what I take to be the most promising nonfunctionalist theories of content properties and state-types, theories not subject to the problem of metaphysically necessary effects.

Although I have not proven it impossible for the functionalist to solve the problem of metaphysically necessary effects, the present difficulty should be considered in conjunction with other longstanding problems for functional-role theories of mental content, including their holist implications and the difficulty they have in assigning determinate and correct truth-conditions to mental states.²⁵ This constellation of problems provides good reason to seek a less problematical approach, and a causal theory provides an obvious alternative (Fodor, 1987, 1990, Dretske, 1988, Millikan, 1984). The problem of metaphysically necessary effects gains no purchase against the causal approach.²⁶ Some causal theories assign a content-determining role to effects, by, say, appealing to the success or adaptive value of the behavior caused—this is especially true of

teleologically oriented causal-historical theories (Millikan 1984, Dretske 1988; cf. note 1 above). These appeals do not, however, imply that any particular effected behavior issues necessarily from a given contentful state.²⁷

Ideally, a causal account will satisfy other desiderata emerging from the present discussion of functionalism and mental causation. Most importantly, we should want a causal theory to (help to) explain how it is that mental states, qua their content, enter into counterfactual-supporting realization-indifferent regularities. This would seem to present a stiff challenge to causal theories; although causal theories of content do well in accounting for the truth- or satisfaction-conditions of mental states, they seem to lack the resources to explain mental processing (cf. Cummins, 1996) and thus the resources to account for causal efficacy of content. I suspect that success on this score requires a causal theory of a specific kind, a causal-historical theory that pays special attention to the development of the underlying neuro-computational substrate. Of particular strength would seem to be a theory that binds content fixation to the very causal interactions with the environment that shape the underlying neural resources (Rupert, 1998/99; 2001). Such a theory might, at least at the level of concepts acquired early in development, effect a tight connection between the content fixed by causal interaction and the patterns of neurally controlled effects that are reinforced by those same causal interactions—a connection tight enough to support regularities adverted to in an autonomous psychology, without being so tight as to individuate the very kinds at issue (cf. Fodor and LePore, 1992, pp. 173-74). There are, of course, complications. One would, for example, like to know whether it is nomologically possible for a representation-using system to emerge absent the single developmental process that is supposed both to fix content and at the same time wed contentful states to apposite behavior. Spelling out and supporting such a causal-developmental

approach is, however, a substantial project that must be undertaken elsewhere (Rupert, 1999; op. cit.).

We should also like to account for the causal power of mental state-types qua state-types. Kinds of mental states are widely thought to be functional-role kinds: being a belief just is being a state that combines in a certain way with motivational states to produce action (see, e.g., Fodor, 1986, p. 14). Something seems right about this view: it surely makes a causal difference whether I believe that P or instead desire that P (whether, for example, I believe that Nader is president or merely desire that this be so); yet, if this essay's main line of argument holds up, functional-role individuation of state-types does not allow for such causal efficacy.

Is there a plausible nonfunctionalist account of state-types? One might hope for a theory cut from the same cloth as the causal account of mental content. On such a view of state-types, that a content-bearing state is of a particular kind might be a function of its causal genesis. If, for example, the state in question acquired a role in the cognitive system because it indicated something about the world, it is a belief (Dretske, 1988). If, instead, the state's content is systematically tied to the need that gave rise to that state (Stampe, 1986) or the state has altered the system's behavior as the result of being reinforced (Dretske, 1988), it is a desire. (Better still, one might advert to the history of the internal mechanism that produced a given state, thus substituting a story about the history of the mechanism for commitment to a very specific story about the causal genesis of each belief and desire.) So long as there is a genuine difference between the two kinds of causal history, a current state can be of a determinate mental type regardless of whether it has effects typical of that type—and thus such a view avoids the problem of metaphysically necessary effects. Although Stampe and Dretske have done significant work in this regard, causal theories of state-types have not had the rigorous development and

examination that causal theories of content have been given. It remains to be seen, then, whether a satisfactory causal account of mental state-types can be fully worked out.

Works Cited

- Anscombe, G. E. M. (1971) Causality and Determination, (Cambridge: Cambridge University Press)
- Antony, Louise, and Joseph Levine. (1997) "Reduction with Autonomy," in Tomberlin, 1997, pp. 83-105
- Block, Ned. (1978) "Troubles with Functionalism," in C. W. Savage (ed.), Perception and Cognition. Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science, vol. 9, (Minneapolis: University of Minnesota Press), pp. 261-325. Reprinted in Block, 1980b, pp. 268-305; page references are to Block, 1980b
- Block, Ned. (1980a) "Introduction: What is Functionalism," in Block, 1980b, pp. 171-84
- Block, Ned (ed.). (1980b) Readings in the Philosophy of Psychology, Volume one, (Cambridge: Harvard University Press)
- Block, Ned (ed.). (1980c) Readings in the Philosophy of Psychology, Volume Two, (Cambridge: Harvard University Press)
- Block, Ned. (1986) "Advertisement for a Semantics for Psychology," in P. French, T. Uehling, and H. Wettstein (eds.), Midwest Studies in Philosophy, Vol. 10: Studies in the Philosophy of Mind, (Minneapolis: University of Minnesota Press), pp. 615-78
- Block, Ned. (1990) "Can the Mind Change the World?" in G. Boolos (ed), Meaning and Method: Essays in Honor of Hilary Putnam, (Cambridge: Cambridge University Press), pp. 137-

- Block, Ned. (1997) "Anti-Reductionism Slaps Back," in Tomberlin, 1997, pp. 107-32
- Braddon-Mitchell, David. (2003) "Qualia and Analytical Conditionals," Journal of Philosophy 100: 111-35
- Braddon-Mitchell, David, and Frank Jackson. (1999) "The Divide and Conquer Path to Analytical Functionalism," Philosophical Topics 26: 71-88
- Cummins, Robert. (1996) Representations, Targets, and Attitudes, (Cambridge: MIT Press)
- Davidson, Donald. (1980) Essays on Actions and Events, (Oxford: Oxford University Press)
- Devitt, Michael. (1990) "A Narrow Representational Theory of Mind," in W. G. Lycan (ed.) Mind and Cognition: A Reader, (Cambridge: Blackwell), pp. 371-98
- Devitt, Michael, and Kim Sterelny (1987) Language and Reality: An Introduction to the Philosophy of Language, (Cambridge: MIT Press)
- Dretske, Fred. (1988) Explaining Behavior: Reasons in a World of Causes, (Cambridge: MIT Press)
- Fales, Evan. (1993) "Are Causal Laws Contingent?" in J. Bacon, K. Campbell, and L. Reinhardt (eds.), Ontology, Causality and Mind: Essays in Honour of D. M. Armstrong, (Cambridge: Cambridge University Press), pp. 121-44.
- Field, Hartry. (1978) "Mental Representation," Erkenntnis 13: 9-61. Reprinted with a postscript in Block, 1980c, pp. 78-114; page references are to Block, 1980c
- Fodor, Jerry. (1981) Representations, (Cambridge: MIT Press)
- Fodor, Jerry. (1986) "Banish DisContent," in J. Butterfield (ed.), Language, Mind, and Logic, (Cambridge: Cambridge University Press), pp. 1-23
- Fodor, Jerry. (1987) Psychosemantics: The Problem of Meaning in the Philosophy of Mind, (Cambridge: MIT Press)

- Fodor, Jerry. (1990) A Theory of Content and Other Essays, (Cambridge: MIT Press)
- Fodor, Jerry, and Ernest LePore. (1992) Holism: A Shopper's Guide, (Cambridge: Blackwell)
- Gillett, Carl, and Barry Loewer (eds.). (2001) Physicalism and Its Discontents, (Cambridge: Cambridge University Press)
- Hall, Ned. (2000) "Causation and the Price of Transitivity," Journal of Philosophy 97: 198-222
- Harman, Gilbert. (1982) "Conceptual Role Semantics," Notre Dame Journal of Formal Logic 23: 242-56
- Horgan, Terence. (1989) "Mental Quausation," in Tomberlin, 1989, pp. 47-76
- Horgan, Terence. (1997) "Kim on Mental Causation and Causal Exclusion," in Tomberlin, 1997, pp. 165-184
- Horgan, Terence, and John Tienson. (1996) Connectionism and the Philosophy of Psychology, (Cambridge: MIT Press)
- Jackson, Frank. (1998) "Causal Roles and Higher-Order Properties," Philosophy and Phenomenological Research 58: 657-61
- Jackson, Frank, and Philip Pettit. (1988) "Functionalism and Broad Content," Mind 97: 381-400
- Jacob, Pierre. (2002) "Some Problems for Reductive Physicalism," Philosophy and Phenomenological Research 65: 649-54
- Kim, Jaegwon. (1973a). "Causes and Counterfactuals" Journal of Philosophy 70: 570-72; reprinted in E. Sosa and M. Tooley, Causation (Oxford: Oxford University Press, 1993), 205-7
- Kim, Jaegwon. (1973b) "Causation, Nomic Subsumption, and the Concept of Event," Journal of Philosophy 70: 217-36; reprinted in Kim, 1993
- Kim, Jaegwon. (1974) "Noncausal Connections," Noûs 8: 41-52; reprinted in Kim, 1993

- Kim, Jaegwon. (1976) "Events as Property Exemplifications," in M. Brand and D. Walton (eds.), Action Theory, (Dordrecht: Reidel), pp. 159-77; reprinted in Kim, 1993
- Kim, Jaegwon. (1979) "Causality, Identity, and Supervenience in the Mind-Body Problem" in French, Uehling, and Wettstein (eds.), Midwest Studies in Philosophy vol. 4, (Minneapolis: University of Minnesota Press), pp. 31-49
- Kim, Jaegwon. (1993) Supervenience and Mind: Selected Philosophical Essays, (Cambridge: Cambridge University Press)
- Kim, Jaegwon. (1998) Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation, (Cambridge: MIT Press)
- Kim, Jaegwon. (1999) "Making Sense of Emergence," Philosophical Studies 95: 3-36
- Kim, Jaegwon. (2002) "Responses," Philosophy and Phenomenological Research 65: 671-80
- LePore, Ernest, and Barry Loewer. (1987) "Mind Matters," Journal of Philosophy 84: 630-42
- LePore, Ernest, and Barry Loewer. (1989) "More on Making Mind Matter," Philosophical Topics 17: 175-91
- Lewis, David. (1966). "An Argument for the Identity Theory," Journal of Philosophy 63: 17-25.
Reprinted in Lewis, 1983, pp. 99-107; page references in the text are to Lewis, 1983
- Lewis, David. (1980) "Mad Pain and Martian Pain," in Block 1980b, pp. 216-22
- Lewis, David. (1983) Philosophical Papers, vol. I, (Oxford: Oxford University Press)
- Lewis, David. (1986) Philosophical Papers, vol. II, (Oxford: Oxford University Press)
- Lewis, David. (1994) "Reduction of Mind," in S. Guttenplan (ed.), A Companion to the Philosophy of Mind (Cambridge: Blackwell), pp. 412-31
- Loewer, Barry. (2001a) "From Physics to Physicalism," in Gillett and Loewer, 2001
- Loewer, Barry. (2001b) Review of Kim, 1998, Journal of Philosophy 98: 315-24

- Lycan, William. (1981a) "Form, Function, and Feel," Journal of Philosophy 78: 24-50
- Lycan, William. (1981b) "Toward a Homuncular Theory of Believing," Cognition and Brain Theory 4: 139-59
- Lycan, William. (1981c) "Psychological Laws," Philosophical Topics 12: 9-38
- McLaughlin, Brian. (1989) "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical," in Tomberlin, 1989, pp. 109-135
- Millikan, Ruth. (1984) Language, Thought, and Other Biological Categories: New Foundations for Realism, (Cambridge: MIT Press)
- Millikan, Ruth. (1999) "Historical Kinds and the Special Sciences," Philosophical Studies 95: 45-65
- Paul, L. A. (2000) "Aspect Causation," Journal of Philosophy 97: 235-56
- Papineau, David. (2001) "The Rise of Physicalism," in Gillett and Loewer, 2001, pp. 3-36
- Pereboom, Derk. (2002) "Robust Nonreductive Materialism," Journal of Philosophy 99: 499-531
- Putnam, Hilary. (1960) "Minds and Machines," S. Hook (ed.), Dimensions of Mind, (New York: New York University Press); reprinted as chapter 18 of H. Putnam, Mind, Language, and Reality: Philosophical Papers, vol. 2, (Cambridge: Cambridge University Press, 1975)
- Putnam, Hilary. (1967) "Psychological Predicates," in W. H. Capitan and D. D. Merrill (eds.), Art, Mind, and Religion, (Pittsburgh: University of Pittsburgh Press), pp. 37-48.
Reprinted as "The Nature of Mental States" in Block, 1980b, pp. 223-31
- Quine, W. V. (1969) Ontological Relativity and Other Essays, (New York: Columbia University Press)
- Robb, David. (1997) "The Properties of Mental Causation," Philosophical Quarterly 47: 178-94

- Rupert, Robert D. (1998/99) "On the Relationship between Naturalistic Semantics and Individuation Criteria for Terms in a Language of Thought," Synthese 117: 95-131
- Rupert, Robert D. (1999) "The Best Test Theory of Extension: First Principle(s)," Mind & Language 14: 321-55
- Rupert, Robert D. (2001) "Coining Terms in the Language of Thought: Innateness, Emergence, and the Lot of Cummins's Argument against the Causal Theory of Mental Content," Journal of Philosophy 98: 499-530
- Shoemaker, Sydney. (1980) "Causality and Properties," in Peter van Inwagen (ed.), Time and Cause, (Dordrecht: Reidel), pp. 109-35
- Shoemaker, Sydney. (1981) "Some Varieties of Functionalism," Philosophical Topics 12: 93-118
- Shoemaker, Sydney. (1998) "Causal and Metaphysical Necessity," Pacific Philosophical Quarterly 79: 59-77
- Shoemaker, Sydney. (2001) "Realization and Mental Causation," in Gillett and Loewer, 2001, pp. 74-98
- Sober, Elliott. (1985) "Panglossian Functionalism and the Philosophy of Mind," Synthese 64: 165-93
- Stampe, Dennis. (1986) "Defining Desire," in Joel Marks (ed.), The Ways of Desire: New Essays in Philosophical Psychology on the Concept of Wanting (Chicago: Precedent), pp. 149-73
- Swoyer, Chris. (1982) "The Nature of Natural Laws," Australasian Journal of Philosophy 60: 203-23

Tomberlin, James (ed.). (1989) Philosophical Perspectives, 3: Philosophy of Mind and Action Theory, (Boston: Blackwell)

Tomberlin, James (ed.). (1997) Philosophical Perspectives, 11: Mind, Causation, and World, (Boston: Blackwell)

Yablo, Stephen. (1992) "Mental Causation," Philosophical Review 101: 245-80

Yablo, Stephen. (1997) "Wide Causation," in Tomberlin 1997, pp. 251-81

¹ I would like to thank the National Endowment for the Humanities for a Summer Stipend (2002) that supported work on this essay. Thanks also to Edward Averill, Carl Gillett, Sungsu Kim, Aaron Meskin, Doug Kutach, Mariam Thalos, Eric Hiddleston, Rob Lovering, and two anonymous referees for *Noûs* for helpful comments on earlier drafts of this paper. Shorter versions of the paper were presented at the 2003 meeting of the Eastern Division of the APA, at North Carolina State University, and, in poster form, at the 2003 meeting of the Society for Philosophy and Psychology. Members of all three audiences gave helpful feedback, for which I am grateful; special thanks to John Carroll, Jonathan Cohen, Phil Robbins, Ron Endicott, Anthony Dardis, Jay Newhard, and Colin Allen for stimulating questions and discussion.

² See Kim, 1993, chapter 13, 1998, chapter 2. The exclusion problem arises from an apparent competition: In any particular case, the realizer of a functionally individuated mental state provides a sufficient cause of what are thought to be the effects of that functional state. The physical nature of the realizer-state together with a plausible principle of physical priority appear to exclude the functionalist mental state from the causal story; there is no causal work left to be done by any distinctively mental property or aspect of that state.

³ I often omit reference to these conditions where mentioning them would be tangential to the point at issue. The realizer of a functional-role state should not, however, be understood as a state that, ipso facto, has certain effects; rather, the realizer has certain effects under certain conditions, the details of which are part of the individuation conditions of the functional-role property realized. This, the common understanding of functional-role properties, leaves room for their characterization as essentially disposition-conferring. That is, one can think of a realizer-state as possessing, qua realizer, a disposition to have certain effects under certain conditions. As should become clear in what follows, commitment specifically to the disposition-conferring view of functional properties does not provide the functionalist with a solution to the problem of metaphysically necessary effects.

⁴ Note that the sort of functionalism I have in mind, and the sort most obviously subject to the exclusion problem, is the sort inspired by computational theory and, more generally, the logico-mathematical characterization of functions. This includes what is often called ‘machine-table’ functionalism, inspired largely by the writings of Hilary Putnam (1960, 1967); it also includes versions of functionalism formulated using Ramsey sentences (Lewis, 1966—for a survey of forms of functionalism, see Block, 1980a), so long as the relevant Ramsey sentence is taken to characterize the various mental properties referred to by our psychological theory (folk or otherwise) (for more on this restriction, see note 8 below). The sort of functionalism at issue does not subsume teleological functionalism (Sober, 1985)—‘teleofunctionalism’, as it is often called—and included here is its homuncular incarnation (Lycan, 1981a, 1981b). Whether the present problem also affects teleofunctionalism depends on how a teleofunctionalist theory characterizes mental functions. If the function of, for example, the speech-parser is characterized

such that it necessarily does what speech-parsers do, a version of the present problem threatens (cf. Lycan 1981c, p. 30). If, instead, the teleological approach assigns functions on the basis of evolutionary history (Sober, 1985), the problem of metaphysically necessary effects might be avoided—but in that case, teleofunctionalism is a version of the causal-historical account of mental properties that I advocate at the end of the present essay (and which I take to be the primary competitor to the kind of functionalist view criticized herein).

⁵ Donald Davidson is primarily responsible for developing the token physicalist view (see various essays in Davidson, 1980, especially “Mental Events”). As a still materialist alternative to the token-identity thesis, we should consider Jaegwon Kim’s view of events as property exemplifications (Kim, 1976) together with a claim to supervenience. This view allows for what might be considered uniquely mental events but is nevertheless physicalist in that each event is the instantiation of a property in a physical object and mental events supervene on physical events (Kim, 1979) (for more on supervenience, see Kim, 1993; for a collection of essays representing the current state of debates surrounding physicalism, see Gillett and Loewer, 2001). The decision to work within the token-physicalist framework makes little difference to the arguments that follow, although at certain places where it seems important to do so, I reformulate essential points in terms of supervening sets of events.

⁶ As has been widely noted (see, e.g., Block, 1990, p. 140), the form of epiphenomenalism of present concern is distinct from traditional epiphenomenalism; the latter view was formulated in terms of causal interactions, or lack thereof, between physical substance and nonphysical, mental substance (and thus was not a problem that arose within the framework set by materialist ontology). All the same, present in traditional and contemporary forms of epiphenomenalism is

a common worry: that distinctively mental items—properties, substances, or whatnot—have no effects of their own on the physical world.

⁷ I will move freely between talk of mental quausation, talk of the causal relevance of mental properties, and talk of their causal efficacy. Ned Block—whose discussion of these issues provides the point of departure for the present study—does not use the term ‘quausality’ or its cognates, but it is clear that he has in mind the problem of mental quausation. For Block, saying that “property P of event c is causally relevant to effect e” is equivalent to saying “that c causes e in virtue of P” (Block, 1990, p. 140). Furthermore, note the way in which Block introduces the epiphenomenalist concern: he raises the worry that computational processes might be sensitive only to syntax, not semantics, even in the case of computational states that possess semantic properties (pp. 138-39). The question whether computational processing is sensitive to semantic properties is essentially the question whether computational states have their effects qua their semantic, or only their syntactic, properties.

Note also that, although causal relevance is sometimes taken to be an epistemic affair (Jacob, 2002, p. 651), the sort of causal relevance Block has in mind is a metaphysical matter. Similarly for quausality: Horgan talks of “explanatory relevance” (1989, p. 50) and accepts that explanation is context relative in certain respects (1997, p. 179), but he denies explanatory irrealism (1997, n15) and states explicitly that quausality is at root a metaphysical, not an epistemological, matter (1989, pp. 53-54).

⁸ The exposition given in the text reflects a widely held understanding of functionalism, the role interpretation (Jackson and Pettit, 1988, Block, 1990, Kim, 1998, 1999, Papineau, 2001). The “realizer” interpretation, according to which mental states just are the realizer states (Lewis

1966, p. 101, 1983, *xi*, 1994, pp. 412-21), provides an alternative but can for present purposes be set aside. It seems immaterial whether the problem of metaphysically necessary effects afflicts functionalism given the realizer interpretation, for the realizer version of functionalism does not offer what nonreductivist functionalists are after: a theory that allows causal autonomy and ontological independence to mental properties and the laws that hold between them. General descriptions of psychological roles do not, on the realizer account, designate properties; they provide mere concepts that pick out local realizers (*ibid.*, p. 420), and there is reason to doubt that autonomous psychological laws can hold between concepts. Admittedly, it might be that states to which a given mental concept rightly applies regularly precede occurrences of events of some type \underline{e} and thus, assuming a regularity-based view of natural laws, are related to \underline{e} -events by psychological law. This tack would be taken for naught, though, if it were meant as a functionalist way around the problem of metaphysically necessary effects. A given mental concept \underline{P} , as a concept of a functionally individuated kind, would not rightly apply to a given state if it were not the case that \underline{e} follows that state (in the circumstances in question).

⁹ Explanations that appeal to functionally individuated properties can seem to provide at least contrastive or negative information about events' causal histories (*cf.* Lewis, 1986, pp. 220-21, 223-24). A desire for a sandwich might have led to Fred's opening of his refrigerator door, but so might have a desire to clean out the fridge. The two states might eventuate in identical metaphysically necessary effects in respect of Fred's kitchen behavior (at a particular time); but given that these states have differing metaphysically necessary effects under other circumstances, it may be useful for someone to explain Fred's trip to the kitchen by saying, in effect, that Fred is in a state that partly realizes one pattern of functional relations rather than another; knowing this

might help the hearer to predict, for example, what Fred will do after he opens the refrigerator door. As has been widely remarked, explanation, like a significant range of communicative activity (see Lewis, 1986, p. 227), is context-sensitive. I do not intend to take any particular stand regarding the nature of explanation, only to point out that, given variations in speakers' and thinkers' purposes and audiences' interests, it should be no surprise if it turns out that one can convey useful information by giving an explanation that does not cite a genuine cause of the explanandum.

¹⁰ Concern about such empty claims of causal relation may well motivate the common practice of requiring, as part of an analysis of causation, the metaphysical or logical independence of causal relata; see Paul, 2000, pp. 245-46, Hall, 2000, p. 198, Loewer, 2001a, p. 46, and, in the context of discussions of mental causation and causal relevance, Horgan, 1989, pp. 56, 58, 61, and LePore and Loewer, 1987, p. 635.

¹¹ cf. Antony and Levine's formula for constructing pairs of sham causal-role properties—1997, pp. 91-92 (also Quine, 1969, pp. 132-33). More than one reader has thought that there must be something wrong with the argument in the text, for functional-role properties just are causal-role properties. "How," it is asked, "could a causal-role property not play a causal role?" According to role-functionalism, functional properties are individuated by the causal roles of their realizers, which leaves open the question whether the functional properties themselves (or their instantiations) have causal efficacy. One might, instead, choose to characterize functional-role properties in such a way that their instantiations must have effects if they are instantiated at all, perhaps by saying, "Property F is simply whatever property is such that its instantiations cause the instantiation of property G." Then the operative questions become "Are there any such

properties?” and “How would we be justified in claiming that such properties are instantiated?” (For further discussion, see the concern expressed below about the anchoring role of realizers.)

¹² The CTP holds that a property is individuated by the causal relations into which it (or its instantiations) enters; see Shoemaker 1980, 1998. On this view, the causal relations into which a property enters are essential to it.

¹³ Yablo’s earlier presentation suggests only an analogy between the part-whole relation and the relation in which he is primarily interested: the relation holding between determinables and determinates (1992, p. 259). Even so, a version of the problem raised in the text would seem to hold against Yablo’s earlier view when that view is wed to functionalism. According to Yablo, a mental property causes e if and only if the mental aspect of the state—the mental determinable—causes e. Combined with the functionalist individuation of properties we get the following: “The aspect of a determinate state—its being a determinate of a particular mental determinable—which is the aspect it is (i.e., the determinable it is) partly because it causes e, causes e.”

¹⁴ Modulo concerns about causal overdetermination; these, however, pertain more directly to the exclusion problem and will be set aside.

¹⁵ One might embrace a local reduction of mental states, perhaps a species-specific type-identity theory of the sort proposed by Lewis (1980) and Kim (1998, although note Kim’s discomfort with this label—2002, response to Jackson), and in this way, eliminate the need to find a causally relevant property pain simpliciter. This may ultimately be the correct position, but we are now considering the program of those philosophers attempting to locate a distinctive causal role for mental properties, one that confers on such properties a form of metaphysical autonomy that local reduction does not offer (see note 8, above).

¹⁶ Compare this to Horgan's analysis of quausion, according to which quausal relations always involve causal relations (1989, pp. 50, 58). On Horgan's view, when we ask about quausion, we ask after the relevance of a certain property to a causal transaction (Horgan 1989, p. 69, n21); we want to know whether a certain property of the cause-event makes any causal contribution to the bringing about of the effect-event. Thus, if A quauses C, the event that instantiates A must *cause* the event that instantiates C. This makes it all the more mysterious how one's quausal gloss could plausibly ignore the structure of the causal connections running from events that instantiate dormativity to those that instantiate being a car wreck.

One might go in for a weaker version of causal relevance, one that eschews any commitment to causal relations among the events that realize functional states or that realize or instantiate inputs and outputs; one might instead ground causal relevance only in nomic relations or relations of counterfactual dependency among property-instantiations. We should, however, bear in mind the point of introducing quausal talk: to understand mental causation, i.e., to understand how mental aspects of human mental states can be responsible for behavior in the way we normally take causes to be. The weaker one's requirements for quausality, the less likely quausion is to yield a robust account of mental causation. Without a demand for causal relations among the realizing (or subvenient) events, it is not clear how one will preserve the materialist framework for causation; there is a mysterious 'direct causation' among mental aspects of events, one that would seem to yield more autonomy than even a nonreductive materialist would want to grant. Thus, it is incumbent on the nonreductive materialist to provide the sort of story McLaughlin demands of type-dualism (1989, pp. 130-31), and it is anyone's guess how one might do so without requiring causal relations among the events, states, or objects

that instantiate quausally related mental properties (or among the events on which events that instantiate mental properties supervene).

¹⁷ One might think that in some cases functionally characterized mental states directly cause certain forms of behavior not implicated by the individuation conditions for those mental states, and that one could formulate a worthwhile psychological theory covering these cases. This would, however, require (1) that the mental states in question have the relevant effects absent any “causal” intermediaries—say, other mental states—to which the alleged causes are necessarily connected and (2) that the forms of behavior in question be unimportant enough to our psychological theorizing about (or to our conception of) the mental states involved that our functionalist theory individuating the mental states in question makes no mention of these forms of behavior. A functionalist psychological theory meeting these constraints would be severely impoverished.

¹⁸ My presentation of Block’s line of reasoning, while fair enough, subverts its order.

¹⁹ Although in “Can the Mind Change the World?” Block does not advocate the strategy of looking for realization-indifferent empirical regularities, in “Anti-Reductionism Slaps Back” (Block, 1997), his discussion of the “Disney Principle” (p. 120) suggests a view similar to Antony and Levine’s. In that essay, however, Block concerns himself primarily with questions about projectibility and natural kinds, not causation: in his closing paragraph, he explicitly distances his position on the former issues from any claim about causal efficacy.

²⁰ Granted, Antony and Levine argue against quausal relations between analytically connected properties. Nevertheless, a counterfactual-based approach to metaphysically necessary “causal” connections might give them reason to change their minds, should such an approach succeed.

Antony and Levine are, after all, generally sympathetic to a counterfactual theory of causation (1997, p. 96).

²¹ This objection to Block's counterfactual account would seem to extend to other counterfactual accounts of mental causation (LePore and Loewer, 1987, 1989, Horgan, 1989, Loewer, 2001b), when they are wed to functionalism.

²² Various problems with extant reductivist views of causality suggest that causality involves an irreducible relation between properties, universal, or kinds. On such a view, although leading reductivist theories of causality likely capture some of what is central to causal relations, attempted analyses of causation are bound to fail.

Even if one stands by a counterfactual analysis of causation, remarks similar to those made above about regularity theories apply here: The counterfactual analysis must in some way account for the noncausal nature of counterfactual dependencies that hold in virtue of logical, mathematical, analytical, and supervenience-based truth, and one should expect any plausible version of the counterfactual analysis to treat Block's patterns as it treats these other, obviously noncausal relations.

²³ The realizer-interpretation of functionalism is sometimes understood as a theory of reference-fixing for mental terms. According to realizer-functionalism, a Ramsey sentence serves to fix nonrigidly the reference of a mental state term to whatever kind of state locally plays the functional-role in question, allowing differences of reference in different contexts. Here our topic is something closer to what functionalists should think of as functional properties—where we focus on the “diagonalized” senses of mental state terms (Lewis, 1983, xi). There are, however, other reasons to set aside realizer-functionalism in the present context (see note 8).

²⁴ This concern was raised by Sungsu Kim

²⁵ Versions of functionalism that make room for so-called long-arm functional roles account more effectively than other forms of functionalism for externalist grounding and truth-conditions (Harman, 1982, Block, 1986, Devitt, 1990). Such theories, however, face their own problems and must overcome residual difficulties faced by functionalist semantics in general (Fodor, 1990, pp. 25-29, Fodor and LePore, 1992, chapter 6, and Cummins, 1996, chapter 4). Given our effort to consider widely the various nonteleological forms of functionalism, this might seem to be a good time to revisit the realizer-version of functionalism (see note 8). Our present focus on content, though, leads in the other direction: problems related to holism and compositionality speak as strongly against realizer-functionalism as they do against role versions.

²⁶ Causal-historical theories of content are not without their own difficulties, for example, the qua problem (Devitt and Sterelny, 1987, pp. 63-65, 72-75) and the disjunction problem (Fodor, 1987, pp. 101-2). These cannot be addressed here (for one proposed solution, see Rupert, 1999).

²⁷ Millikan, for example, emphatically distinguishes the content-determining (or more broadly, function-determining) effects of a state from its statistically normal or usual effects, arguing that the former can be relatively rare effects of the behavior in question.

Note also that causal theories avoid the problem of metaphysically necessary effects regardless of the extent to which Robert Cummins is right in his characterization of such theories as functional-role theories of a sort: ‘use’ theories that premise content upon the role of mental representations in detection (Cummins, 1996, chapter 5). Even if the use of a mental representation in detection determines that representation’s content, it is not the case, on a causal approach, that the effects of such uses constitutively determine whether a given use was correct

or determine the content of that mental representation, as would have to be the case in order to generate the problem of metaphysically necessary effects.