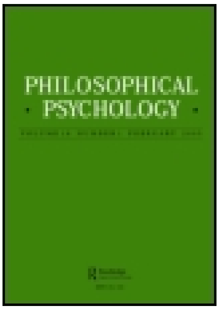

Philosophical Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cphp20

Two tales of functional explanation

Martin Roth & Robert Cummins

Published online: 17 May 2013.

To cite this article: Martin Roth & Robert Cummins (2013): Two tales of functional explanation, Philosophical Psychology, DOI:10.1080/09515089.2013.800454

To link to this article: http://dx.doi.org/10.1080/09515089.2013.800454

Philosophical Psychology, 2013
http://dx.doi.org/10.1080/09515089.2013.800454

Two tales of functional explanation

Martin Roth and Robert Cummins

This paper considers two ways functions figure into scientific explanations: (i) via laws—events are causally explained by subsuming those events under functional laws; and (ii) via designs—capacities are explained by specifying the functional design of a system. We argue that a proper understanding of how functions figure into design explanations of capacities makes it clear why such functions are ill-suited to figure into functional-cum-causal law explanations of events, as those explanations are typically understood. We further argue that a proper understanding of how functions enter into design explanations of capacities enables us to show why two prominent objections to functionalism in the philosophy mind—the argument from metaphysically necessary effects (Bennett, 2007; Rupert, 2006) and causal exclusion (Kim, 1993, 1998; Malcolm, 1968)—are misguided when interpreted as posing a threat to functional explanation in science across the board. If those arguments pose a threat at all, they pose it to instances of (i); however, a great number of the functional explanations we find in psychology—and the sciences generally—are instances of (ii).

Keywords: Causation; Explanation; Functionalism

1. Introduction

This paper considers two ways functions figure into scientific explanations: (i) v*ia laws*—events are causally explained by subsuming those events under functional laws; and (ii) v*ia designs*—capacities are explained by specifying the functional design of a system. We argue that a proper understanding of how functions figure into design explanations of capacities makes it clear why such functions are ill-suited to figure into functional-cum-causal law explanations of events, as those explanations are typically understood. We further argue that a proper understanding of how functions enter into design explanations of capacities enables us to show why two prominent objections to functionalism in the philosophy mind—the argument from metaphysically necessary effects (Bennett, 2007; Rupert, 2006) and causal exclusion (Kim, 1993, 1998; Malcolm,

Martin Roth is Assistant Professor of Philosophy at Drake University.
Robert Cummins is Emeritus Professor of Philosophy at University of California, Davis.
Correspondence to: Martin Roth, Department of Philosophy and Religion, Drake University, 2507 University Avenue, Des Moines, IA 50311, USA. Email: martin.roth@drake.edu

1968)—are misguided when interpreted as posing a threat to functional explanation in science across the board. If those arguments pose a threat at all, they pose it to instances of (i); however, a great number of the functional explanations we find in psychology—and the sciences generally—are instances of (ii).

2. Functionalism as Metaphysics of Mind

As a thesis in the philosophy of mind, functionalism is the doctrine that mental states can be specified in terms of functional roles. Here is a simple-minded example of the sort often used to introduce the idea in philosophy classes. Begin with some "laws" of folk psychology:

> Stabs and kicks cause pain.
> Stress causes anxiety.
> Anxiety causes elevated galvanic skin response (GSR).
> Pain causes anxiety, screams, and avoidance.

The same information can be put in a diagram (see Figure 1). Although the diagram contains the same information as the laws, the diagram makes it easier to see the basic idea of functionalism: S is in pain if and only if S is in a state that plays the role indicated by the oval with 'Pain' in it. This is the conception that enables a Ramsey sentence definition of pain (Lewis, 1972):

> Pain $=_{\mathrm{df}}$ the state that is (a) caused by stabs and kicks, and (b) causes screams, avoidance, and the state that both causes elevated GSR and is caused by stress.

Note the part played by causation here: it is in the "laws," represented by the arrows in the diagram and featured in the Ramsey sentence. Logically, the strategy works no matter what the arrows stand for: we could substitute any arbitrary relation. The reason for the choice of causality is that it is the (or a) choice that makes the "laws" interesting and, in the case of a real theory, possibly true.

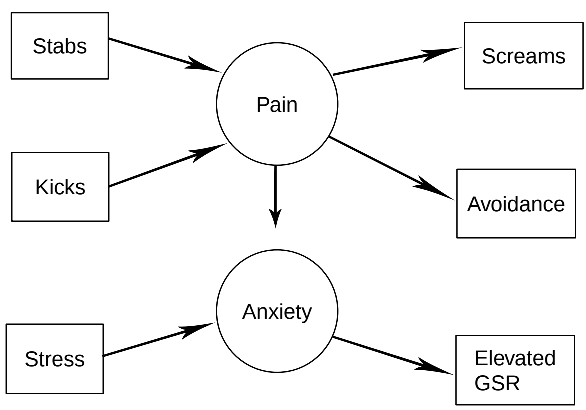

Figure 1 The pain role.

Most versions of functionalism hold that functional roles can be specified naturalistically, and since functional characterizations of mental states are neutral about the underlying details of the states that occupy functional roles, mental states appear to be multiply realizable. The significance for metaphysics is thought to be this: if mental states are multiply realizable, then mental states do not reduce to physical states, i.e., to states essential to fundamental physics. Put in the language of properties, functionalism is widely interpreted as the thesis that mental properties are "second-order properties," e.g., the property pain is the property of having a first-order property that satisfies the functional profile for pain. But properties like being a doorstop and being a catalyst seem to be second-order as well, so if doorstops and catalysts do not generate worries for physicalism, then neither should pains, beliefs, and other mental properties. Functionalism thus looks to provide a physicalist-friendly, yet non-reductive, theory of our mental life.

The version of functionalism according to which mental properties are second-order properties is often called *role* functionalism (McLaughlin, 2006). *Realizer* functionalism, by contrast, is the view according to which mental properties are identical to the first-order properties that realize the relevant functional roles. Unlike role functionalism, realizer functionalism *denies* that mental properties are multiply realizable. To see the difference here, suppose we were to discover an alien species whose nervous system is physically quite different from ours. If there were states of the alien nervous system that realized the functional role of pain, then, according to role functionalism, humans and aliens would both instantiate the second-order property of being in pain. However, according to realizer functionalism, because the first-order properties that realize the functional role of pain differ in humans and aliens, there would not be a single property—being in pain—shared by humans and aliens. Rather, there would be *two* properties here: the first-order property that realizes the functional role of pain in humans and the first-order property that realizes pain in aliens.[1]

Those who favor realizer functionalism often do so because of alleged troubles role functionalism makes for mental causation. For example, Rupert (2006) argues that functionalism suffers from the problem of metaphysically necessary effects. Here is Rupert's statement of the problem:

> 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 *e* (under such-and-such conditions) causes *e* (under those conditions). 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 *e*; any state that does not cause *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. (2006, p. 256)

The problem becomes clear when we note that role functionalism permits us to substitute the Ramsey definition into the "laws." But when we do so, we do not get a contingent nomic generalization. We get something like this:

> (a) Stabs and kicks cause the state that is caused by stabs and kicks, and (b) screams, avoidance, and the state that both causes elevated GSR and is caused by stress are caused by the state that causes screams, avoidance, and the state that both causes elevated GSR and is caused by stress.

Making the "laws," and hence the diagram, more complete and more accurate will evidently not solve this problem.

Moreover, role functionalism seems to face the problem of causal exclusion (Kim, 1993, 1998). Here is a quick summary of the argument:

(P1) Each instantiation of a mental property is realized by an instantiation of some physical property.
(P2) The instantiation of physical properties is sufficient to produce whatever effects the instantiation of mental properties are thought to produce.
(P3) No effect can have more than one sufficient cause.
(C) Therefore, if mental properties are second-order properties, then the physical properties that realize mental properties causally exclude mental properties—all the causal "work" is done by physical properties, and mental properties are thus epiphenomenal

The thought here is that, if the laws represented in the diagram shown in Figure 1 are taken to be causal laws, then we must suppose that mental properties are causal properties, i.e., that pains cause anxiety and screams. But surely the states that realize pains do the causing. If so, then it cannot be the case that a pain causes a scream unless pain *just is* the realizer state, i.e., the first-order property that realizes the functional role of pain. Hence, the laws represented in the diagram are not causal laws if those laws feature a non-reduced, multiply realized property *pain*.

Neither the argument from metaphysically necessary effects nor the argument from causal exclusion has gone without detractors,[2] but regardless of how these debates in the metaphysics of mind get resolved, notice that, as formulated above, none of them speak *directly* to the science of psychology. This isn't to say that the parties to these debates have ignored the implications for psychology; as we will see, the alleged significance of the debate about functional properties to the science of psychology is often made explicit. However, we think many of the claims made about the implications of the metaphysical debates for the science of psychology rest on questionable assumptions about the way functions enter into the explanatory enterprise, and once we take a closer look at these assumptions, it becomes a lot less clear that the worries raised for role functionalism pose any threat to functional explanations in psychology—or in the sciences generally.

3. Functional Explanation as Causal Explanation (Tale I)

If successful, what implications would the arguments from metaphysically necessary effects and causal exclusion have for the science of psychology? According to Rupert, the cost of denying the contingency of causal generalizations is that the resulting generalizations threaten to render functional explanations vacuous:

> 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, aversion behavior exhibits aversion behavior. Such a "law" appears vacuous, stating no more than that states that cause *e* cause *e*. (2006, p. 258)

As this passage makes clear, empirical psychology is supposed to be in the business of causally explaining events by subsuming those events under laws. If the generalizations that role functionalism yields are not contingent, then the resulting laws—and thus explanations—are vacuous. Rupert is not alone in assuming the above connection between causation, laws, and scientific explanation. For instance, Kim writes:

> Causal powers involve laws, and laws are regularities that are projectible. Thus, if pain ... is not a kind over which inductive projections can be made, it cannot enter into laws, and therefore cannot qualify as a causal kind; and this disqualifies it as a scientific kind. (1993, p. 327)

And though Fodor and Kim disagree about the viability of functionalist psychology, Fodor and Kim agree that psychological explanation is (at least in large part) subsumption of events under causal law. For example, Fodor writes:

> We have, in practice, no alternative to the vocabulary of common-sense psychological explanation; we have no other way of describing our behaviors and their causes if we want our behaviors and their causes to be subsumed by any counterfactual-supporting generalizations that we know about. (1987, p. 8)

The position that emerges from these passages from Rupert, Kim, and Fodor is what Roth (forthcoming) calls the *Hempel-Hume* account of scientific explanation. According to the Hempel-Hume account, psychological explanations require generalizations because science is in the business of explaining events, and events are explained only if they are deducible from laws—understood as generalizations—and a statement of initial (or boundary) conditions (Hempel, 1966; Hempel & Oppenheim, 1948). The availability of generalizations is not enough to establish that psychological explanations are scientific, of course, for the generalizations we find in science appear to state contingent connections. However, the contingency of psychological generalizations looks to be secured by the thought that psychological explanations are causal explanations, and causal connections are contingent. So, for example, Fodor writes, "causal powers are a function of your *contingent* connections, not of your conceptual connections. As, indeed, Uncle Hume taught us" (1991, p. 19).

If science is primarily in the business of explaining events and explanation is subsumption under contingent causal law, then it is no wonder that the arguments from metaphysically necessary effects and causal exclusion, if successful, look to spell trouble for functionalist psychology. It is typically assumed that the predicates used to specify laws denote properties (Loewer, 2007), and viewed through the lens of Hempel-Hume, functionalist psychology requires that the properties denoted by psychological predicates (the predicates used to specify the laws of psychology) are functional properties. But if psychological predicates pick out functional properties, and functional properties are not identical to the properties that realize functional roles, then functionalist psychology entails role functionalism. Thus, if role functionalism is false, then functionalist psychology is bankrupt.

There are two general lines of response available here. The first, and more familiar, is to argue that functional explanation is ubiquitous in science, and is causal through and through. Moreover, it is implausible to think that the laws figuring in such explanations will reduce to laws of physics. For example, consider the circuit diagram given in Figure 2. Resistors (represented by the jagged lines) are anything across which there is a drop in potential: resistors cause potential drops. A power supply is something across which there is a rise in potential: power supplies cause potential increases. Now suppose the switch is closed, and we take a reading at V (potential in volts). What caused the change represented by that reading? Closing the switch, obviously. But, of course, one would like to know why closing the switch causes just those changes. A parody of Rupert's argument might go like this:

> Ohm's law allows us to calculate the current at A, given the potential across the power supply and the resistance of the two resistors in sequence. But a resistor is just something across which there is a potential drop. So it is not a contingent fact that applying a current across the resistors causes a drop in potential. If this did not happen, the things represented by the jagged lines would not be resistors. So the explanation by appeal to functionally specified components is empirically vacuous.

And now the causal exclusion argument:

> What actually causes the drop in potential is the microphysics of whatever implements the resistors. So, the drop in potential is not caused by the resistor in virtue of its being a resistor, but in virtue of its microphysical constitution. So the functional description of the circuit diagram does not reveal causal structure.

Put these two together, and we get the result that explanations that appeal to functional properties are either vacuous or non-causal. Obviously, something has gone seriously

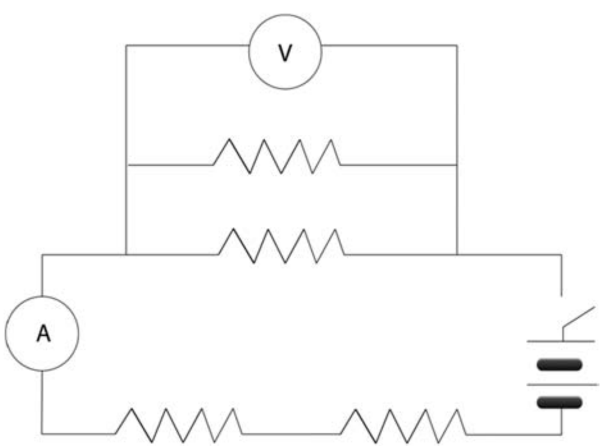

Figure 2 A circuit diagram.

wrong. The diagram abstracts away from implementation details, but appealing to Ohm's law and a circuit diagram to explain the occurrence of a potential drop is surely an example of scientifically legitimate causal explanation.[3]

The second, and less familiar, line of response is to challenge the assumption that functional explanations provide causal-subsumptivist explanations of individual events. In the next section, we will argue that, while functional explanations are ubiquitous in science—including psychology—a great number of the functional explanations we find are not in the business of subsuming events under functional laws. And once these functional explanations are properly understood, it becomes clear that the arguments from metaphysically necessary effects and causal exclusion are irrelevant to much of the functional explanation we find in psychology—and the sciences generally.

4. Functional Explanation as Design Explanation (Tale II)

A good deal of science is concerned with explaining the properties of complex systems—especially their characteristic effects—by analysis, i.e., by the analysis of a systemic property into organized interaction among other simpler systemic properties or properties of component subsystems (Cummins, 1975, 1983). This explanation-by-analysis is *functional* analysis because it operates at a level of abstraction that identifies analyzing properties in terms of what they do or contribute, rather than their intrinsic constitutions. As we are thinking of them here, effects are not individual events, but dispositional properties, which can be specified as regularities. Examples include the photoelectric effect, the Stroop effect, the McGurk effect, Thorndike's law of effect, and Emmert's law. To distinguish the two senses of 'effect', we will prepend an asterisk when we mean the regularity, e.g., the Stroop *effect. Most *effects in psychology do not have names, but they are what almost all psychology papers report (Cummins, 2000).[4] The strategy of explaining *effects—the properties of complex systems—by functional analysis is ubiquitous in science and engineering, and by no means special to psychology.[5]

Cognitive folk psychology, while often portrayed as a set of laws or truisms, is sometimes portrayed as a functional analysis of intelligent or adaptive behavior; in particular, the ability to act so as to satisfy goals or desires. Stephen Stich (personal communication) once referred to this as the game of life: get contents from the desire box into the belief box by forming an intention that generates a plan to change the environment, which, in turn, changes belief. As an illustration, consider the following (toy) analysis of the ability to act so as to satisfy one's desires, yielding a regularity in behavior that we might call the Planning *effect (see Figure 3). Though this diagram reveals information about causes, the explanatory targets of this sort of analysis are not particular points in state space or particular trajectories through it. Rather, the aim of this kind of analysis is to appeal to a system's design to explain why one finds the trajectories one does find and not others. The design provides a model of the state space and constrains the possible paths through it, thereby explaining the system's characteristic *effects.[6]

More generally, the explanandum of a functional analysis is a property, e.g., the capacity for adaptive behavior, and the strategy is to understand the properties of a complex system by exhibiting the abstract functional design of that system—to show,

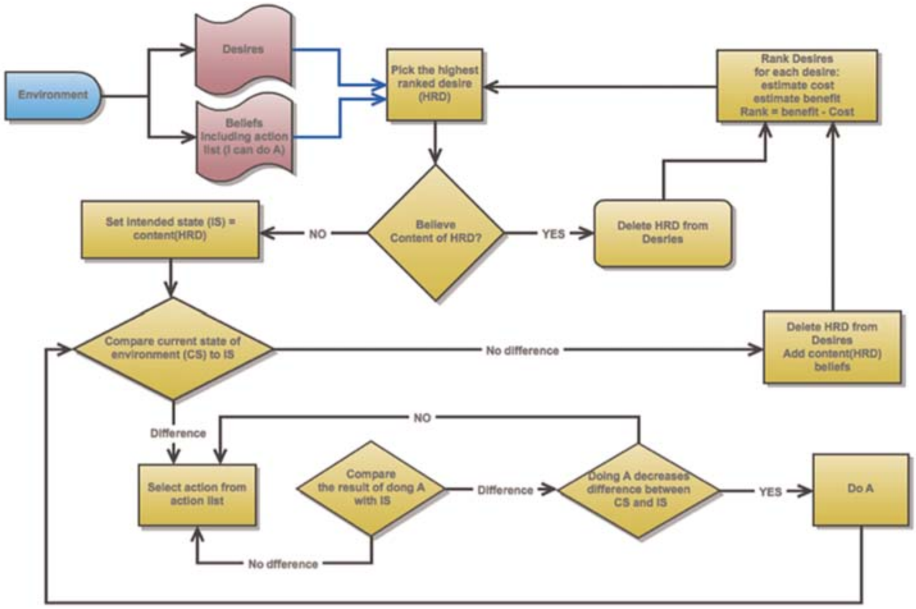

Figure 3 A toy analysis of the Planning *effect.

in short, that a system with a certain design is bound to have the property in question. Block provides a nice (toy) illustration of the idea:

> Suppose one wants to explain how we understand language. Part of the system will recognize individual words. This word-recognizer might be composed of three components, one of which has the task of fetching each incoming word, one at a time, and passing it to a second component. The second component includes a dictionary, i.e., a list of all the words in the vocabulary, together with syntactic and semantic information about each word. This second component compares the target word with words in the vocabulary (perhaps executing many such comparisons simultaneously) until it gets a match. When it finds a match, it sends a signal to a third component whose job it is to retrieve the syntactic and semantic information stored in the dictionary. This speculation about how a model of language understanding works is supposed to illustrate how a cognitive competence can be explained by appeal to simpler cognitive competences, in this case, the simple mechanical operations of fetching and matching. (1995, p. 385)

The target of this explanation is a capacity, and the explanation works by specifying a functional design, i.e., the way components—specified in terms of the *jobs* they perform—are organized. When components are specified functionally, some causal consequences of components are highlighted and others are ignored. In the context of a successful design explanation, the causal consequences highlighted by functional specifications are precisely those consequences that enable us to understand why a system with a particular organization of components has the capacity or *effect we are trying to explain. In this way, functional specifications provide *causal relevance filters*: by selecting from the myriad causal consequences of a system's components those that are relevant to the target *effect, functional specification makes it transparent *why* anything that possesses a certain organization of components is bound to have the *effect in question. Without this filtering, we are simply left with a welter of noisy detail with no indication of what is relevant and what is a mere by-product of this or that implementation. Furthermore, by abstracting from the implementation details that are irrelevant to the achievement of the targeted *effect, causal relevance filtering enables us to understand why implementations that differ in those details but retain the design all exhibit the targeted *effect. In short, causal relevance filtering gives you multiple realizability for free.

It is old news that functional properties are multiply realizable, of course, but the significance of multiple realizability has tended to be obscured by the underlying assumption that functional explanation is the subsumption of events under contingent causal law. For example, Fodor claims that functionalists need "to explain why we should (why we do) prefer higher-level nondisjunctive laws (*pain leads to avoidance behavior*) to lower-level laws with open disjunctions (*states that are* R1 ∨ R2 ∨ ... *lead to avoidance*), all else being equal" (1998, p. 20), and his explanation is that

> this policy complies with a dictate that inductive practice obeys quite generally: *Prefer the strongest claim compatible with the evidence, all else being equal*. Quantification over instances is one aspect of rational compliance with this injunction; reification of high level kinds is another. (1998, p. 20)

Fodor's talk of inductive practice and preferring the strongest claim compatible with the evidence suggests that the relevant generalizations are contingent and thus need empirical support to establish them. This is not surprising, of course; as we've seen, Fodor assumes a Hempel-Hume model of functional explanation. For Fodor, multiple realizability is licensed by sound scientific practice.

However, though we agree with Fodor that sound scientific practice licenses multiple realizability, we think his defense of it is misguided. As we discuss in the next section, Rupert is right that the laws you get when you reify functional kinds are metaphysically necessary, and though this may violate Humean strictures concerning causal relations between events, this isn't, pace Fodor, something that those attempting to provide functional analyses should care about. Sound scientific practice licenses multiple realization because sound scientific practice licenses abstraction across implementation details, and sound scientific practice licenses abstraction across implementation details because the functional terms that figure in functional analyses are causal relevance filters.

5. Metaphysics and the Two Tales

From the point of view of functional analysis, functional properties are dispositional properties, and dispositional properties are like Moliere's dormitive virtues: you cannot explain why opium puts people to sleep by appeal to its dormitive virtue, and you cannot explain why a clock keeps time by appeal to its chronographic virtue.[7] But, returning to the circuit diagram, no one should think that the circuit diagram generates empirically vacuous explanations simply because "resistors cause potential drops" is not contingent. Similarly, while it is true that potential drops, when they occur, are mediated by whatever happens to implement the resistors in question, no one should think that this undermines the explanatory value of the circuit diagrams. We'll take up these points in turn.

Something may cause a potential drop in one circuit but not in another, or under some conditions of load, but not others (hence the importance of super cooling to magnetic levitation). It *is* contingent that *this* resistor causes a potential drop in *this* circuit under *these* conditions. You might say:

> *r* is a resistor, but isn't functioning as one *here/now*.

or

> *Here/now*, *r* isn't a resistor.

When you do get a potential drop because of *r*, you get it because, in that context, *r* functions as a resistor, i.e., because, in that context, *r* will cause potential drops. You know this empirically: that kind of resistor works under those conditions. You get no science for free, here. If you want to insist on a parse that wears its contingency on its sleeve, you can say: "this resistor is a silicone semiconductor that is known to function as a resistor in this context." It is similarly contingent that there was a resistor in this circuit, and that its presence, rather than, say, a battery problem, was the cause. Along the same lines, while it is not a contingent fact that a system that executes an addition algorithm adds,[8] or that a planner plans, it is a contingent fact that some particular system executes an addition algorithm or is a planner, and hence it is a contingent fact that some system can add or plan. It is, therefore, also a contingent fact that some system outputs 2 given $1 + 1$, or plans to rob a painting. These points seem to us enough to make enlightened descendants of Uncle Hume happy.[9]

Turning to the causal explanation of events, we will grant for the sake of argument that the realizer properties, not the functional properties realized, do all the causal work, e.g., we will grant that the thing that implements a resistor causes potential drops to occur—not the resistor qua resistor.[10] But admitting that realizer properties do all the causal work in no way threatens the explanations provided by functional analysis. The causal exclusion argument imagines a causal competition between functional properties and the properties that realize functional properties, but from the point of view of functional analysis, the imagined competition does not exist. It does not exist because the target explananda are different: *effects versus effects. Again, the connection between functional characterization and causation is that functional terms are causal relevance filters. These filters do not introduce new competitors; rather, they select from among the myriad causal consequences of a state, process, or event, and this is precisely what is needed in order for them to do their work in functional analysis.

To illustrate this last point, consider the following explanation of how a logic circuit works. Figure 4 is a diagram of a CMOS (complementary metal oxide semiconductor) circuit that implements XOR: letting positive voltage stand for *true* and negative voltage stand for *false*, the output voltage is positive if and only if there is positive voltage applied to just one of the inputs (A or B). To explain the "XOR *effect," we need to

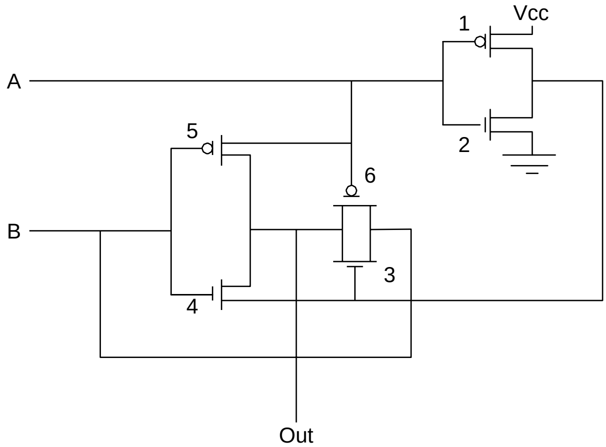

Figure 4 An XOR circuit diagram.

understand how the transistors open and close gates (labeled by the numbers) and how the orchestration of gate openings and closings affects the output voltage. One of the advantages of this CMOS circuit is that is does not heat up nearly as much as other circuits (in fact, reducing heat was one of the motivations that led to the invention of CMOS circuits), but notice that a complete explanation of why the circuit exhibits the XOR *effect can be given without mentioning anything about how much the circuit heats up. The functional analysis that explains why the system exhibits the XOR *effect abstracts away from heat because the heat produced in a system satisfying this design is irrelevant to understanding the XOR *effect. If you tried to explain how the circuit works in terms of the properties that realize transistors, no advantage would be gained as far as the explanation of the XOR *effect goes. In fact, the resulting explanation would be disadvantageous. Heat can vary across systems that exhibit the XOR *effect, and since the description used in the lower-level explanation does not single out the causal consequences relevant to explaining the *effect, the heat produced by the system will be on all fours with the output voltage, as far as the lower-level description is concerned. So, to get the kind of understanding that the explanation is supposed to provide, we will have to be told to ignore the heat. *But this is what the diagrams do*. The description in terms of realizer properties *undermines* the transparency provided by causal relevance filtering and thereby undermines the understanding provided by the design. Moreover, even if we could *identify* the property of being a transistor with some realizer property, P, replacing ‘transistor’ with ‘P’ in this context would mask the information needed to understand why the system exhibits the XOR *effect.[11]

Functional analyses specify designs, and because a design creates a state space and constrains the possible paths through it, designs can explain a system’s *effects. But systems exhibit a multitude of *effects, and in order for a functional analysis to explain an *effect, it has to isolate those causal consequences that are relevant to the *effect in question. This is what functional concepts provide. When viewed this way, it should be obvious why worries about metaphysically necessary effects and causal exclusion do not threaten the legitimacy of functional explanations. Functional terms are causal relevance filters, but the explanatory power they provide has nothing to do with the subsumption of cause-effect pairs under functional laws. From the perspective of functional analysis, the arguments for abandoning role functionalism in favor of realizer functionalism thus make little sense.

With the above points in mind, let's return to the diagrams in Figures 1 and 3. The trouble for role functionalism starts when we (a) notice that the diagrams can be read as alternate representations of a set of "laws"—stabs cause pains, beliefs and desires cause intentions—and then (b) take the point of the explanatory content of the diagrams to be the same as the explanatory content of the laws, as construed along Hempel-Hume lines: subsumption under causal law of individual pains; anxiety attacks; avoidance behaviors; increases in GSR; or intentions to go to the refrigerator. One will see beliefs interacting with desires to cause plans, but at the same time one will see beliefs and desires defined, in part, by the very behaviors they are said to cause—hence the problem of metaphysically necessary effects. One will also see causal relations among events occurring in virtue of functional properties rather than in virtue of realizer properties—hence the problem of causal exclusion. Just as *abandoning* role functionalism makes no sense from the point of view of functional analysis, *adopting* role functionalism makes no sense from the point of view of Hempel-Hume.

We are now in a position to see that the culprit here is not role functionalism. Nor is it causal explanation. The culprit is a largely unchallenged but dubious assumption about how role functionalism is supposed to enter into the explanatory enterprise. If we read the diagrams as *functional analyses*, it should be clear that embracing role functionalism does not undermine the explanatory power of such analyses. In the first case, we have a functional analysis of (one kind of) pain *effect—what we might call the Scream and Sweat *effect. In the second case, we have a functional analysis of (one kind of) adaptive behavior—what we have been calling a Planning *effect. To repeat: functional analysis does not explain events by subsuming events under functional laws. Functional analysis explains *effects: dispositional properties of complex systems. In particular, it explains how a complex system works, and, consequently, why it has the property that is the target explanandum. It does not do this by identifying the causes of the system's acquisition of that property, but by specifying the abstract design of the system, a design the having of which amounts to having the target property.

With that said, we think there *is* something suspicious about trying to turn the pain diagram into a functional analysis.[12] The worry is not about the metaphysics, however. Given what's in the diagram, the Scream and Sweat *effect comes to the regularity that kicks and stabs are each followed by screams, avoidance, and elevated GSR. If the "analysis" of this regularity is given in terms of pain, but pain is then *defined* as the state that mediates the connection between kicks and stabs, on the one side, and screams, avoidance, and elevated GSR, on the other, then the analysis amounts to little more than a dormitive virtue explanation of the *effect. This trivialization is inevitable if you specify properties functionally and then take the point of functional explanation to consist in subsuming events under laws, where the event types subsumed are part of the functional specification of the property that figures in the law. This is why the pain diagram generates a trivial explanation of the Scream and Sweat *effect, while the circuit diagram in Figure 4 does not generate a trivial explanation of the XOR *effect. The XOR *effect is not explained by sole appeal to voltage levels across A and B and the "law" that transistors are switches (open and close gates). The "law" is relevant to the explanation of why the circuit exhibits the XOR *effect only when we add the *design* of the circuit, i.e., only when we specify the *number* and *locations* of the transistors. Specifying those components as *transistors* is what enables the "law," when coupled with information about the number and locations of those components, to be explanatory. But again, the explanandum here is the XOR *effect (and perhaps, indirectly, manifestations of the XOR *effect), not particular gate openings and closings.

There are two lessons we want to draw from all of this. The first is that, for those interested in the implications of the metaphysics of mind for science, the evaluation of realizer and role functionalism should be sensitive to the distinction between Hempel-Hume and functional analysis. If you are after causal-subsumptivist explanations of potential drops and screams—individual events—then realizer functionalism about resistors and pains is the way to go. If you are after design explanations of *effects, then role functionalism is the way to go. The failure to honor this distinction will inevitably tempt us to divorce role functionalism from functional analysis and/or wed role functionalism to Hempel-Hume, but as we've shown, each makes little sense.

The second lesson is this: discussions of mental causation have had a prominent place in the functionalism literature, and as a review of that literature reveals, those discussions tend to assume the Hempel-Hume account of scientific explanation, i.e., that functional explanation is the subsumption of events under functional-cum-causal laws. This assumption can and should be challenged—not only as an assumption about psychological explanation, but scientific explanation generally. A close examination of the science reveals functional analysis—and the design explanations it provides—to be ubiquitous, and so when we look at the metaphysics with an eye towards science, the worries about metaphysically necessary effects and causal exclusion should start to evaporate.

6. Conclusion

Functionalism has typically been illustrated with the help of folk-psychological "laws," or by explaining a function as something defined in terms of its role in a causal network. This, in turn, has made it look like functionalism is committed to causal laws couched in functional terms, laws that can explain events. This, in turn, has made role functionalism look vulnerable to the problem of metaphysically necessary effects and causal exclusion. But if our concern is to salvage functional explanation as much of it actually occurs in science, we can embrace role functionalism—even if that requires us to accept that functional "laws" are not contingent and that causal powers lie in realizer properties, not realized properties. Role functionalism has its home in design explanations of *effects—explanations that are ubiquitous in everyday life, engineering, and the sciences generally—not in causal explanations of events.

Notes

[1] The realizer functionalist can allow that there are second-order properties shared by humans and aliens, but the realizer functionalist will deny that these second-order properties are mental properties. As we discuss in the text, the denial is motivated by concerns about mental causation.

[2] For example, according to the causal theory of properties, the causal features of a property are essential to that property (Shoemaker, 1998). Because causal theorists think all causal connections are metaphysically necessary, they won't see metaphysically necessary effects as a problem for functionalism. And for those who accept a dependence account of causation (Lewis, 1983), there seems to be no principled reason why we can't say that functional properties and realizer properties are each causes.

[3] The worry that the causal exclusion argument would render all macro-level causation illusory (i.e., all causation except causation in fundamental physics) has been much discussed in the metaphysics of mind literature. (Block, 2003 provides a particularly nice treatment of the issue.) But we think that causation in fundamental physics is far more mysterious and problematic than ordinary cases of putative macro-level causation, e.g., we think that "the blowout caused the accident" is on way firmer ground than singular causal claims in fundamental physics. There are genuinely *serious* problems about causation in quantum mechanics. See, for example, Berkovitz (1998) and Elby (1992). "Blowout" may not be functional, but blowouts are certainly multiply realizable, so if multiple realizability is supposed to be the problem, it is a problem that has to do with *abstraction* across different realizations, functional characterization being just one flavor. One response to this might be to argue that while not all abstraction leads to exclusion, functional abstraction does. This possibility is worth exploring further, but we will not do so here.

[4] Because functional language is frequently used in the specification of *effects themselves, many *effects are called "functional laws." But if *effects are primarily explananda, the prevalence of "functional laws" doesn't automatically entail the prevalence of, or the necessity for, explanation by subsumption under functional law. While we allow for the latter, we think that it is relatively shallow—and not the primary role of functional characterization in science. For our present purposes, however, it is enough that a good number of the functional explanations we find in science are not properly construed along subsumptivist lines.

[5] Well-known historical examples include Einstein's analysis of the photoelectric effect, Mendel's analysis of inheritance patterns in plants, Helmholtz's analysis of Emmert's Law, and Newell and Simon's analysis of problem solving (to name just a few). In engineering, design specification is almost always functional analysis (e.g., circuit diagrams in electrical engineering), and if you check the Wikipedia article for just about any *effect in any science, you will find (explicitly or implicitly) a functional analysis.

[6] The kind of explanation provided by the design should not be confused with what Dretske (1988) calls a structuring cause. As Dretske characterizes them, structuring causes are events or processes that shape or structure another process. Dretske illustrates this with the example of conditioning. If a dog salivates, we might explain this by citing the ringing of a bell. The ringing of the bell is what Dretske calls the triggering cause of the salivating. But if we want to know why the dog salivates *instead of*, say, jumping up and down, then we will appeal to the conditioning process that produced the bell-to-salivating connection. The conditioning process is the structuring cause. Designs are not events or processes, however, and while designs may have structuring causes, designs themselves cannot be comfortably assimilated to structuring causes.

[7] McLaughlin (2006) makes the point that role functionalism provides a nice account of dispositions. We obviously agree, but we think the significance of this to scientific explanation—functional analysis, and the causal relevance filtering it provides—has not been sufficiently appreciated.

[8] Except in the trivial sense that it might have broken down. But a broken adding machine is not a machine that executes an addition algorithm either.

[9] While functional analysis does not causally explain events by subsuming events under functional laws, e.g., while individual potential drops are not explained by subsuming them under a law-like generalization relating resistors and potential drops, a design specified by a functional analysis can generate Humean-friendly causal subsumptivist explanations of events. For example, you can explain why a system S exhibited effect *E on a particular occasion (an instance of event type e2) by citing the fact that some instance of event type e1 acted on system S and the fact that S has design D. It is contingent that S has D, it is contingent that an instance of event type e1 occurred, and it is contingent that D operated as specified on the occasion in question. All of this is testable in the obvious ways. Furthermore, on the hypothesis that S has design D, you can generalize from the particular occasion in which an instance of event type e1 acting on S was followed by an instance of event type e2. With such a generalization in hand, you can explain, in a rather shallow but nevertheless Hume-certified way, an individual event by appeal to the following generalization: events of type e1 acting on system S are followed by events of type e2.

[10] Granting this is problematic because the function-implementation relation iterates: what is function at one level of abstraction is implementation when viewed from a higher level of abstraction, and what is implementation at one level of abstraction is function viewed from a lower level of abstraction (Lycan, 1987). Thus, a resistor is anything in a circuit across which there is a drop in potential. This might be a semiconductor, or a motor, or a light bulb: each of these might implement the resistance function. But we have just identified the implementations *themselves* in familiar functional terms, and there are many ways to implement semiconductors, light bulbs, and motors. There are at least two ways to respond to this. The first is to say that you don't get causal explanation of events until you reach fundamental physics. One problem with this response is that causation in fundamental physics is itself problematic (see note 3). Another problem is that, arguably, we find functional characterization even in fundamental physics. The second response is to relativize the causal explanation of events to particular levels in the function-implementation hierarchy, e.g., relative to a functional analysis in which resistor figures, a motor (or light bulb, etc.) counts as the realizer that does the causal work. While this would make causal explanation of events perspectival, some have suggested that this kind of perspectivalism is everywhere in science (e.g., Craver, 2012; Giere, 2006).

[11] We think Putnam is making this point when he writes: "even if it were not physically possible to realize human psychology in a creature made of anything but the usual protoplasm, DNA, etc., it would still not be correct to say that psychological states are identical with their physical realizations. For, as will be argued below, such an identification has no explanatory value in psychology" (1975, p. 293). We also think that the importance of Putnam's remark has been largely ignored or misunderstood.

[12] In the context of introducing functionalism in the philosophy of mind, the pain diagram is never used as a toy illustration of how functional analysis explains *effects. It is used, rather, as a toy illustration of how something can be identified by its place in a causal network (e.g., Armstrong, 1968). The latter got called functionalism, but it certainly isn't proto-functional analysis of the kind you find in cognitive science, or the sciences generally. For example, a resistor is not identified by its causal role in a circuit. Rather, it is identified by a dispositional property that can play a role in a *effect producing circuit design. Of course, you can try to turn the pain diagram into functional analysis by inventing an *effect (as we

do with the Scream and Sweat *effect). However, for reasons we discuss in the text, it is pretty clear that the diagram is not up to this task.

References

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge.

Bennett, K. (2007). Mental causation. *Philosophy Compass*, *2*(2), 316–337.

Berkovitz, J. (1998). Aspects of quantum non-locality II: Superluminal causation and relativity. *Studies in History and Philosophy of Modern Physics*, *29B*(4), 509–545.

Block, N. (1995). The mind as the software of the brain. In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith, & S. Sternberg (Eds.), *An invitation to cognitive science* (Vol. 3, pp. 377–426). Cambridge, MA: MIT Press.

Block, N. (2003). Do causal powers drain away? *Philosophy and Phenomenological Research*, *67*(1), 133–150.

Craver, C. F. (2012). Functions and mechanisms: A perspectivalist account. In P. Huneman (Ed.), *Functions: Selection and mechanisms* (pp. 133–158). Dordrecht: Springer.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*, 741–765.

Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press.

Cummins, R. (2000). "How does it work?" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (pp. 117–144). Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.

Elby, A. (1992). Should we explain the EPR correlations causally? *Philosophy of Science*, *59*(1), 16–25.

Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. (1991). A modal argument for narrow content. *The Journal of Philosophy*, *88*, 5–26.

Fodor, J. (1998). Special sciences: Still autonomous after all these years. In his *In critical condition* (pp. 9–24). Cambridge, MA: MIT Press.

Giere, R. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.

Hempel, C. (1966). *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice Hall.

Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135–175.

Kim, J. (1993). *Supervenience and mind*. New York: Cambridge University Press.

Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.

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

Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, *61*, 343–377.

Loewer, B. (2007). Mental causation, or something near enough. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in the philosophy of mind* (pp. 243–264). Boston, MA: Blackwell.

Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press.

Malcolm, N. (1968). The conceivability of mechanism. *Philosophical Review*, *77*, 45–72.

McLaughlin, B. (2006). Is role-functionalism committed to epiphenomenalism? *Consciousness Studies*, *13*(1–2), 39–66.

Putnam, H. (1975). Philosophy and our mental life. *Mind, language, and reality* (Vol. 2, pp. 291–303). New York: Cambridge University Press.

Roth, M. (forthcoming). Folk psychology as science. *Synthese*. DOI: 10.1007/s1122901202406

Rupert, R. (2006). Functionalism, mental causation, and the problem of metaphysically necessary effects. *Nous*, *40*(2), 256–283.

Shoemaker, S. (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly*, *79*, 59–77.