Functionalism, Superduperfunctionalism, and Physicalism: Lessons From Supervenience

(forthcoming in *Synthese*)

Ronald Endicott
Department of Philosophy & Religious Studies
North Carolina State University
Campus Box 8103, Raleigh, NC 27695
ron_endicott@ncsu.edu

1. Introduction

Philosophers almost universally believe that concepts of supervenience fail to satisfy the standards for physicalism because they offer mere property correlations that are left unexplained. They are thus compatible with non-physicalist accounts of those relations. Moreover, many philosophers not only prefer some kind of functional-role theory as a physically acceptable account of mind-body and other inter-level relations, but they use it as a form of "superdupervenience" to explain supervenience in a physically acceptable way (Kim 1990, 1998, 2005; Horgan 1993; Loewer 1995; Melnyk 1994, 2003). [1] But I reject a central part of this common narrative. I argue that functional-role theories fail by the same standards for physicalism because they merely state without explaining how a physical property plays or occupies a functional role. They are thus compatible with non-physicalist accounts of that role-occupying relation.

I also argue that one cannot redeploy functional-role theory at a deeper level to explain role occupation, specifically by iterating the role-occupant scheme. Instead, one must use part-whole structural and mechanistic explanations that differ from functional-role theory in important ways. These explanations represent a form of "superduperfunctionalism" that stand to functional-role theory as concepts of

[1] Some argue that a theory of superdupervenience is superfluous inasmuch as the added concepts do all the explanatory work (Melnyk 1999; Wilson 2005). But as a matter of classification I count any theory that explains supervenience as a theory of superdupervenience.

superdupervenience stand to concepts of supervenience. I then close by suggesting a revision of the standards for physicalism that preserves the parallel between supervenience and functional-role theory in a different way. By the revised way of thinking, supervenience and functional-role theory both count as physically acceptable theories. Either way, the story of supervenience and functionalism should be told with the same basic plot.

2. The Complaint Against Supervenience

I will take the doctrine of physicalism to encompass both the reductive view whereby all entities are physical (physical identity) as well as the nonreductive view whereby some entities are nonphysical but they are explained by and ultimately depend upon physical entities (physical priority). [2] Now philosophers almost universally believe that supervenience relations fail to meet the demands of physicalism. To illustrate, consider Jaegwon Kim's notion of "strong supervenience." Where *A* is a set of supervening properties and *B* is a set of physical properties that serve as the subvenient base for *A*, Kim defines it thus: *(S1) necessarily, for any object x and any property F in A, if x has property F, then there exists some property G in B such that x has G, and (S2), necessarily, for any object y, if y has G then y has F* (Kim 1984, p.165). Clause (S1) rules out free-floating properties. All supervenient *A* properties are accompanied by a physical property in *B*. Clause (S2) then requires that the physical *B* properties determine the *A* properties. Thus, interpretations of (S2) entail a set of supervenience laws, meaning a set of modally strong property correlations over the same individual such as: it is *necessary that, for any object, if it has a neural property G then it has a mental property F*.

What is the problem? The basic idea is that *supervenience relations are consistent with objectionable non-physicalist positions, such as British emergentism and G.E. Moore's meta-ethical non-naturalism* (Schiffer 1987; Kim 1990, 1993; Horgan 1993;

---

[2] Some philosophers also speak of "materialism" and "naturalism." But I will speak uniformly of physicalism. One may also understand physical priority in terms of current or future physics, although there are problems with either option. For some recent discussion, see Melnyk (1997) and Wilson (2006).

Wilson 2005). Philosophers have emphasized different aspects of this problem, but the central point is about *explanation*, specifically, that certain facts about supervenience are left unexplained. For example, the supervenience law in clause (S2) could be true if *F* emerges from a physical *G* as an inexplicable brute fact. Terence Horgan explains this problem using a notion of "physical supervenience," which is to say, a preferred physicalist notion that restricts supervenience to possible situations consistent with actual physical laws:

> There are important lessons in the fact that the thesis of physical supervenience is consistent with the central doctrines of British emergentism, because those doctrines should be repudiated by anyone who advocates a broadly materialistic metaphysics … a materialistic position should assert that all supervenience facts are explainable – indeed, explainable in some materialistically acceptable way (Horgan 1993, p.560).

Or, again, when speaking about the inter-level connections of a part-whole-layered world, Kim expresses a similar worry:

> On the layered interpretation, mind-body supervenience is an instance of *mereological supervenience*, and this might seem like an advance, tempting us into thinking that we might try explaining mind-body supervenience in parallel with the way macrophysical properties are determined and explained by microphysical properties. But supervenience or determination is one thing, explanation quite another. We may know that *B* determines *A* (or *A* supervenes on *B*) without having any idea why this is so – why *A* should arise from *B*, not *C*, why *A*, rather than *D*, arises from *B* (1998, p.18).

So supervenience must be supplemented with a physically acceptable explanation for its property correlations so that the resulting view is no longer consistent with

emergentism. More generally, any physically acceptable theory should exclude non-physical positions by showing that its facts are physically explainable, save a fundamental physical theory whose facts are not subject to further explanation. Call this the "physical explanation" condition: *a non-fundamental theory must have the resources to show that its ontology is explainable in a physically acceptable way*. Of course there are other standards for physicalism that have been raised against supervenience. Some worry that supervenience is also consistent with non-physical extras such as souls or divine activities that do not violate physical laws (see Witmer 1999; Hawthorne 2002). Yet one might think that such non-physical items are problematic precisely because their activities are unexplainable. Some physicalists believe that explainability even trumps non-materiality, as illustrated by the acceptance of abstract objects in mathematics because they are allegedly indispensible to explanations in science (e.g., Quine 1976). Or again, some philosophers argue that the correlations of supervenience do not guarantee that all facts depend upon physical facts (Grimes 1988; Kim 1990, 1993). Yet the difference between dependence and mere correlation arguably turns on the fact that the former is an explanatory relation whereas the latter is not. Thus I think it is fair to focus on the physical explanation condition, since it is central to the overall complaint against supervenience.

3. Judging Functional-Role Theory by the Same Standard

I now turn to the family of functional-role theories. They divide into different kinds based on formal, causal, historical, social, and normative senses of functional roles (see Polger 2004). But I will present my argument in terms of the still popular causal-role functionalist view. This theory is identified by two propositions, schematically put: (F1) *x has functional property F = x has a physical property that occupies causal-role R,* and (F2) *x has a physical property G that occupies causal-role R* (Lewis 1966, p.17; Armstrong 1968, pp.90-91; Kim 2011, pp.171, 178). I add four points that are relevant to my argument. First, I stated (F1) in a way that is neutral between a *reductive physicalist version* whereby *F* is identical to the physical property that occupies role *R* versus a *nonreductive physicalist version* whereby *F* is a second-order property of having some first-order physical property occupy causal-role *R* (see Block 1980). This difference will

not affect my thesis, which concerns the facts of causal-role occupation described by (F2). Indeed, my argument applies even if $F = G$. Second, causal-role functionalism is a *single-subject* theory in the sense that the same object possesses both role and occupant properties (follow the variable for individuals in Block, 1980; Lewis, 1980; Shoemaker 1982; and see Kim 1998, 82). [3] This feature will provide part of the contrast with the multi-subject resources of a part-whole explanation for role occupation.

Third, functionalists understand *role playing* or *role occupation* as the possession of a property described by a lower-level realization theory standing in the causal relations specified by a higher-level functional theory. Using the well-known Ramsey-Lewis method (Lewis 1980; Shoemaker 1982; Rey 1997; Kim 2011), where "C_i ... C_n causes F causes E_1 ... E_n" is a conveniently abbreviated series of causal statements that constitute causal-role R described in (F1) by a functional theory, a statement of role occupation (F2) is then equivalent to "x has a physical property G and (C_i ... C_n causes G causes E_i ... E_n)," where "G" is a term from a more basic realization or implementation theory that has been substituted for "F." [4] A statement of role occupation thus expresses a modally strong

---

[3] This is the traditional picture of functional-role theory. Sydney Shoemaker (2003, 2007) has recently explored a multiple-subject view that incorporates a coincident microphysical state of affairs.

[4] Let me add three smaller points. One, this is "literal" role occupation, as Carl Gillett (2002) describes it, as opposed to the way specific part properties might account for F by standing in their own causal relations. Two, I will frequently speak of property G standing in the causal relations, although others may prefer the expanded "G enables its instances to stand in causal relations." Three, in light of Shoemaker's (1982) distinction between a "core" realization denoted by "G" versus a "total" realization denoted by the entire open predicate "x has G and (C_1 ... C_n causes G causes E_1 ... E_n)," I prefer to say that the core occupies the role insofar as the occupier is caused by C_1, and causes E_1, and so on (the total realization is not caused by C_1, it is the instantiation of the entire set of causal relations that *includes* C_1 causing G). Still, when one explains how the core occupier G is able to stand in the pertinent set of causal relations, one thereby explains

property correlation over the same individual, in the mind-brain case, such as: *it is causally necessary that, for any object, if it has a sensory input property C then it has a neural property G … and if it has a neural property G then it has a behavioral property E* (cf. the modally strong connections of strong supervenience). I will argue that the laws of causal-role occupation, and the singular facts that fall under them, bring in train the same type of concerns that were raised about supervenience.

Fourth, the facts of role occupation are typically *not fundamental facts of physics where explanations come to an end*. Consider the paradigm cases discussed in the literature (see Tye 1995). In a familiar mind-brain case, *being a system of neurons* occupies the role of information processing for a mind by transmitting signals to other areas of the brain. Yet being a system of neurons that transmits signals to other areas of the brain is not a fundamental fact of physics. Even on a reductive view, a system of neurons is a massive aggregate of fundamental entities, not a fundamental entity itself. Likewise, *being a lattice structure of carbon atoms* occupies the role of hardness in a diamond by doing things like resisting penetration from a macro object and passing a scratch test. Yet being a lattice structure of carbon atoms that resists penetration from a macro object and passes a scratch test is not a fundamental fact of physics either. Or again, *being H_2O* occupies the role of water by doing things like appearing clear in a glass or expanding at 0°C. Yet the fact that H_2O appears clear in a glass and expands at 0°C is not a fundamental fact of physics.

Parenthetically, this point about non-fundamentality makes perfect sense given the standard interpretation of occupier properties as complex *structural properties*, that is, properties whose instances have a structure that implies parts. [5] So *being a system of*

---

the total realization that consists of *G* standing in those causal relations. So no harm will result if I speak of the core as the occupier property.

[5] For a way to develop the notion of a structural property, see Pagés (2002); cf. the notion of a micro-based property in Kim (1998, p.84). Also, whereas it is generally true that a structural property is not a fundamental physical property, there are exceptions. E.g., quantum entanglement seems to be a fundamental physical state type that is both complex but not determined by the properties of the parts (see Maudlin 1998).

neurons implies the more basic units of individual neurons, *being a lattice structure of carbon atoms* implies the more basic units of individual carbon atoms, and *being H_2O* implies the more basic units of hydrogen and oxygen atoms. Indeed, one may view an occupier *G* like a function in the mathematical sense that takes lower-level parts as arguments and yields a complex higher-level whole as its value, generating a non-fundamental composite that exists at a higher mereological level appropriate for the possession of the property *F* as described by a relatively higher-level functional theory as well as the causal activity deemed relevant by that higher-level functional theory.

Now given that the facts of role occupation expressed by paradigm statements of (F2) are not fundamental facts of physics where explanations come to an end, it follows that they must be explained. This should not be controversial. For example, David Papineau accepts a version of functional-role theory, yet he observes that there are "role-filling explanations" even if the functional property *F* is identical to the occupier *G* and even if identities require no explanation: "Take the claim that water is H_2O. If we understand the term 'water' as in some sense a priori equivalent to 'the familiar liquid which is colourless, odourless and tasteless,' then we can sensibly ask why H_2O is water, and read this as a request for an explanation why H_2O is colourless, odourless and tasteless, a request that can in principle be answered by reference to the physical chemistry of H_2O" (1998, p.380). Or again, in a recent work Kim mentions the familiar two steps of functional-role explanation constituted by (F1) and (F2), but he now adds a third: "Step 3 [Developing an Explanatory Theory] Construct a theory that explains how the realizers of *F* perform task *R*" (2005, p.102, with a change in variables). Later Kim returns to the point: "The third step consists in developing an explanation at the lower, reductive level of how these mechanisms perform the assigned causal work" (2005, p.164). So one should explain how *G*, the "mechanism" or "realizer" of *F*, performs its causal task, which is to say that one should explain how the mechanism or realizer *G* occupies causal-role *R*. But this need to explain the facts of role occupation leads directly to my thesis.

Simply put, by itself, the conjunction of (F1) and (F2) leaves the explanation for role occupation wide open. Hence, given just this canonical formulation of functional-role theory, the fact that a physical *G* stands in causal relations *R* could be accounted for

in physically unacceptable ways, thus violating the previously discussed physical explanation condition. To illustrate with a mind-brain case, it has been the leading hypothesis for several decades that different types of neural systems play psychological roles by receiving signals from certain areas of the brain and sending signals to other areas of the brain. That is, neural systems occupy psychological roles by neurotransmission. So suppose (F1) *x* has functional property *F* (e.g., is a face-recognition device) = *x* has some type of neural system that is caused to be activated by signals from C_i ... C_n (e.g., face-like stimuli) and causes signals to be sent to E_i ... E_n (e.g., behavioral reactions to faces), and (F2) *x* has a type of neural system *G* that is caused to be activated by C_i ... C_n and causes signals to be sent to E_i ... E_n. [6] Now (F2) does not explain how the neural system *G* is able to receive and send signals. It simply states the fact that needs to be explained. But (F1) does not explain the fact in question either. It equates having the functional property *F* with having some neural property whose instances send and receives signals without explaining how any property enables its instances to perform the causal task in question. Consequently one must offer some additional propositions to explain the fact expressed by (F2).

So consider the physically unacceptable emergentist hypothesis whereby neurotransmission is a brute fact, determined but unexplainable by the activities of the sub-neural components and molecules and atoms that underlie the neural system. Indeed, emergentists maintained that biological phenomena arise from chemistry and physics in an unexplained way. Moreover, the emergentist hypothesis about neurotransmission might have appeared plausible until relatively recently when the properties of single ion channels within the neuron were finally understood (more on the role of ion channels within neurotransmission shortly). [7] But the important point is that the emergentist

---

[6] PET and MRI studies have shown that, among adult humans, the neural system *G* that recognizes faces is typically instantiated in the fusiform gyrus or Broadman area 37 (see Sergent, et. al.,1992; and Kanwisher et. al., 1997).

[7] Erwin Neher and Bert Sakmann (1976) made the discoveries in question, which earned them a Nobel Prize in 1991. Also, one might locate the emergentist threat in different lower-level places. E.g., Brian McLaughlin (1992) argues that British emergentism was

hypothesis is consistent with (F1) and (F2) – it would remain true that a system of neurons is caused to be activated by signals from face-like stimuli and causes signals to be sent to the appropriate centers for behavioral control even if the mechanism for neurotransmission were a complete mystery so that neural system had its causal capacities by brute emergence from the underlying chemistry.

This possibility is depicted below, where the top level represents the fact expressed by (F2) regarding causal connections described by a psycho-functional theory with a neural occupier *G* in the place formerly described by the functional term "*F*," the bottom level conveniently represents causal connections between several levels of sub-neural constituents, and the bold vertical arrow between the part-whole levels represents brute determination for the pertinent constitutive relations:

Causation between x's sensory inputs $C_i \ldots C_n$*, neural system G, and behavior* $E_i \ldots E_n$

Causation between parts of x's subneural, chemical, and physical properties $P_i \ldots P_n$

Illustration 1: An emergentist hypothesis that accounts for the neural occupation of a functionally described causal role.

I represent emergence as a purely inter-level matter, not an intra-level matter. That is important, for the claim that "$C_i \ldots C_n$ *causes G causes* $E_i \ldots E_n$ is emergent" might suggest two different things. It might suggest that the intra-level relation expressed by "causes" is an unexplained emergent relation between $C_i \ldots C_n$ and *G* and again between *G* and $E_i \ldots E_n$. Or it might suggest that the inter-level relation between the parts and the causal fact that $C_i \ldots C_n$ *causes G causes* $E_i \ldots E_n$ involving the whole is an unexplained

---

not rebutted until the advent of quantum chemistry, which means that one must block the threat from emergentism at the place where quantum mechanics interacts with chemical phenomena.

emergent relation. The first would be problematic on the assumption of causal-role functionalism, given a standard interpretation that does not treat causation as brute and unexplained. But I intend the second. *The emergence only concerns inter-level part-whole relations rather than intra-level causal relations*. Hence, as long as one does not conflate the two, the emergentist hypothesis is consistent with a robust interpretation of causation as an explanatorily relevant dependency as required by causal-role functionalism.

Indeed, assorted analyses of causation allow for the possibility in question. For example, it may still be true that, were there no face-like stimuli, the appropriate group of neurons would not be activated, as required by counterfactual and manipulability accounts of causation, even if both face-like stimuli and the face-recognizing neural system emerge from the underlying chemistry and physics in an unexplained way (manipulations carried out on face-like stimuli still "make a difference" for face-recognizing processing). Similarly, it may still be true that the presence of face-like stimuli raise the probability that the pertinent group of neurons would be activated, or that they transfer energy to the pertinent group of neurons, or are connected by some kind of process, as required by other accounts of causation, even if the neural system emerges from the underlying chemistry and physics in an unexplained way. [8]

---

[8] The preceding two paragraphs were meant to allay the worries of an anonymous reviewer that my argument might depend upon a fairly weak account of causation, such as correlations that would remain intact under various non-physical possibilities. The reviewer also added that, if the fact that *G* plays causal role *R* is accounted for in an emergentist way, then a functionalist might say it is not *G* that plays role *R* but rather *G* in combination with whatever emergent feature of reality linked *G* to *R*. But saying that an "emergent feature links *G* to *R*" is ambiguous in much the way I indicated in the text. It might mean that some emergent property *X* links *G* to *R* by standing in the intra-level relation between *G* and *R* (*C brings about X which brings about G*). Or it might mean that some lower-level feature of the parts accounts for why *G* stands in *R* by some inter-level emergent relation. Once again, whereas the first conflicts with the causal relations posited by functional-role theory, the second claim about constitutive relations does not. Let me

Of course it is correct to see a lack of explanatorily relevant dependence between *G* and the lower-level properties of the parts. Part-whole emergence is not an explanatory dependence relation. Yet that is exactly what the counterexample is supposed to show – for it is just another way of saying that the facts of role occupation are consistent with emergence from lower-level facts about the parts. Yet the compatibility of (F1) and (F2) with a non-physicalist emergence also shows that *the bare statement of causal-role functionalism does not satisfy the physical explanation condition*. Those propositions do not show that the facts of role occupation are explainable in a physically acceptable way. So causal-role functionalism, as defined by (F1) and (F2), fails by the same standard that was raised against supervenience. [9]

---

also add that, on a standard causal theory of properties, *G* is to be individuated by its intra-level causes and effects – forward and backward looking powers (Shoemaker 1998, 2007) – not upward and downward constitutive relations, meaning in the present case, not by its inter-level constitutive relations to the parts of the instances of *G*.

[9] Jessica Wilson (1999, p.40) was the first to argue in this way. Specifically, she argued that Horgan's kind of superdupervenience, where there is an explanation for a macro-feature like liquidity in terms of micro-properties *via* a functional definition, is not sufficient for physicalism because it is consistent with liquidity being *supercaused* by the micro-properties in an unexplained way (citing Stephen Yablo's 1992, pp.256-257, emergentist interpretation of supervenience). But whereas Wilson was focused on the supervenience relation, I focus on the fact of role occupation. I also think this creates a dialectical advantage. For it is not clear how the supervenience law $G \Rightarrow F$ remains emergent if it is subject to a functional-role explanation. By challenging instead the physical acceptability of the fact of role occupation $C_i \ldots C_n$ *causes* G causes $E_i \ldots E_n$, I thereby challenge a *premise* in the functional-role explanation for the supervenience law $G \Rightarrow F$. Moreover, I go well beyond Wilson's case by considering a number of objections (section 4), doing exhaustive search through the resources of functional-role theory (section 5), and criticizing iterations of the role-occupant scheme as a strategy for explaining role occupation (section 6).

Let me also underscore that the present argument is not a problem for nonreductive versions of causal-role functionalism alone. Assume that mental *F* is identical to neural *G*. Nevertheless, the question is not about the "function-to-realizer" or *F*-to-*G* relation, the answer to which may well be the proposed identity. The question is about the "realizer-to-role" or *G*-to-*R* relation. That is not an identity but a neural property standing in a number of causal relations to distinct types of sensory inputs and behavioral outputs. And my point is simply that, given just the canonical statement of causal-role functionalism provided by (F1) and (F2), this neural property might stand in the occupied causal relations by the most inexplicable means vis-à-vis the underlying chemistry and physics. To merely assert rather than explain causal-role occupation ensures that "physicalist functionalism" is compatible with objectionable emergentist positions at a deeper level where explanations for role occupation should apply. One may call the foregoing argument "the threat from below," since the threat to physicalism does not arise from above the occupier *vis-à-vis* the existence of a higher-level irreducible functional property. Rather, the threat to physicalism arises from below the occupier *vis-à-vis* deeper-level non-physicalist accounts for role occupation.

Let me also emphasize that the argument does not depend upon any particularities of the chosen case. The same argument can be made for any paradigm case of role occupation discussed in the literature. Consider the case of H_2O occupying the role of water, and to simplify, consider just one aspect of the causal role of water: (F1) water = a type of thing that is caused to expand by freezing temperature; and (F2) H_2O is caused to expand by freezing temperature. But this does not explain why freezing temperature causes H_2O to expand. It leaves the matter wide open. Hence it is consistent with these assertions that H_2O expands at freezing temperatures as an emergent fact from the underlying physics. Again, one needs a further set of propositions to explain this fact of role occupation and thus remove the threat from below.

So, to summarize the overall dialectic thus far, strong supervenience, in the form of (S1) and (S2), was offered as an account of mind-brain and similar relations. But philosophers objected that the laws expressed by (S2), and the singular facts that fall under them, are compatible with objectionable non-physical positions like emergentism and must be explained in a physically acceptable way. Likewise, causal-role

functionalism, in the form of (F1) and (F2), was offered as an account of mind-brain and similar relations. But I have pointed out that the laws expressed by (F2), and the singular facts that fall under them, are likewise compatible with non-physical positions like emergentism and must be explained in a physically acceptable way. Given that functionalists were quite vocal in their complaints about supervenience, this result should count as one of the greater ironies in contemporary philosophy.

4. Initial Objections and Replies

I will now address five objections that help clarify my position. *Objection 1*. Someone might reject my argument, claiming that role occupation can be explained by (F1) and (F2) on a reductive physicalist version of the theory on one traditional account of explanation. That is, one can *deduce* (F2) from a first-order reading of (F1) and the identity of *F* and *G*. If *F* = the property that stands in *R,* and if *F* = *G*, then it follows that *G* stands in *R*. But, in response, deduction is no guarantee of explanation, otherwise one should retract the complaint against supervenience – mind-brain connections can be deduced from mind-brain supervenience. Moreover, it is obvious that something is missing. The conclusion only states that *F/G* stands in the said role, without presenting any information that explains how or why it does so. It thus remains compatible with the aforementioned non-physical or emergentist threat from below.

Objection 2. Someone might find a doctrine of emergence to be plausible, or even true (see defenders and critics in Bedau and Humphreys 2008). Indeed, some believe that causation is a brute or primitive relation (see Carroll 1994). So one might wonder why the relation of physical occupation could not be brute as well. [10] Yet, in response, the present question is not whether emergence is plausible, or true, or justifiably held on the basis of other brute relations. Perhaps intra-level causation is brute. Perhaps inter-level constitutive relations are brute as well. Rather the question is this – do certain standards for physicalism that are contrary to brute emergence only count against the physical acceptability of supervenience or do they also count against the physical acceptability of

[10] I thank an anonymous referee for raising this issue.

functional-role theory? I have answered in the affirmative even if those standards are mistaken or even if physicalism is false.

Objection 3. Someone might attempt to sidestep my argument by recommending that occupiers and roles be specified in terms of fundamental physics, locating them in fundamental physical facts for which there is no further explanation. That is, if occupier properties are conceived to be fundamental physical properties, and the pertinent causal roles are likewise conceived to be fundamental physical relations, then there can be no emergentist threat from below for the simple reason that nothing is more basic – there is nothing further to explain how *G* occupies role *R* in a physically *un*acceptable way. This would be part of a larger reductive physicalist program. Of course the emergentist threat I have described would still apply to nonreductive versions of functional-role theory. But, as I suggested earlier, I believe my argument also applies to reductive versions of functional-role theory.

To begin, there is every reason to doubt that an occupier property is a fundamental property of physics rather than a non-fundamental property of physics. For example, one identifies a mind with a massive non-fundamental aggregate of fundamental entities, and one identifies a property that occupies a mental role with a complex non-fundamental physical property built out of or explained by fundamental physical properties. Moreover, there remains the need to explain how these massive aggregates and their non-fundamental physical properties occupy the roles described in higher-level theories. My view, pace much of the recent work on mechanistic explanation, is that part-whole theories which utilize resources beyond functional-role theory provide a necessary link between the basic reducing theory of fundamental physics and the items targeted for an explanatory reduction in higher-level theories (see, e.g., Machamer, Darden, and Craver 2000; Craver 2007; and Gillett 2007).

To illustrate, return to a reductive case already mentioned, the case of water and H_2O. Assume the latter is reducible to fundamental physics. Even so, as Papineau pointed out (1998, p.380), one must explain how H_2O occupies the water role, meaning that one must explain various aspects of the water role, for example, why freezing a body of H_2O causes it to expand. And the explanation is that a perfectly bonded H_2O molecule has a V-shaped H-O-H angle with an open space between the hydrogen atoms at the one end.

At temperatures above 0°C there is more thermal energy to break the hydrogen bonds and shake the hydrogen atoms out of position, partially collapsing the structure. But at 0°C the molecule becomes completely hydrogen bonded due to less thermal energy. Consequently there is more open space between the hydrogen atoms in its solid state, in contrast to its liquid state, which thus creates an increase in volume for the entire body of H_2O.

Now I think three things are plausible. First, such explanations are necessary for the program of reductive physicalism. Water = H_2O, and so on down to fundamental physics. But without explanations like the one above it would be a complete mystery how H_2O stands in the causal role of water. Indeed, such explanations serve to justify the claim of identity, and without them the scientific community might have reasonably kept looking for a more complicated state associated with H_2O to identify with water, one whose behavior under different conditions would provide the needed explanation for water's expansion when frozen as well as the many other aspects of the water role. Second, such explanations are not couched in the language of fundamental physics, contrary to the suggestion presently under consideration. Indeed, if the target for reduction is an item described in a higher-level psychological theory, many of the relevant parts cited in the explanation would be neurobiological (see the example provided in the next section). [11] Put differently, such explanations provide intermediate-level explanatory links between fundamental physics and the targeted higher-level theories. Accordingly, recent accounts of mechanistic explanation emphasize this point.

---

[11] Thus there are familiar reasons why no one offers, say, quantum conditions for statements of role occupation regarding higher-level theories. They are practically impossible to formulate, given the sheer number and complexity of fundamental entities involved. As well, their description would lose any serious explanatory connection with the higher-level phenomena targeted for explanation. Even Stuart Hameroff and Roger Penrose's (1996) controversial quantum theory of consciousness does not appeal directly to basic physics, maintaining instead that quantum events effect the microtubules in a neuron, which in turn effect neurotransmission in a way that is relevant to alleged non-computable aspects of consciousness.

Thus Peter Machamer, Lindley Darden, and Carl Craver refer to multiple part-whole levels in a mechanistic explanation as "nested hierarchies," which they illustrate with the aforementioned case of neurotransmission: "the activation of the sodium channel is a component of the mechanism of depolarization, which is a component of the mechanism of chemical neurotransmission, which is a component of most higher-level mechanisms in the central nervous system" (2000, p.13). They also rightly observe that such mechanistic explanations typically "bottom out" in the lowest level of interest for a given scientist, research group, or field, noting explicitly that, in molecular biology and neuroscience, such explanations "do not typically regress to the quantum level" (loc. cit.).

Third, and finally, such explanations take one beyond the resources of functional-role theory as defined by (F1) and (F2). This is the burden of section 5 which follows. But, to briefly introduce my position, functional-role theory requires that an occupier property described by (F2) stand in the very causal relations defined for a functional property *F* defined by (F1). But, to cite just one difference, the properties cited in the above explanation regarding how *G* stands in role *R* are *part properties that do not stand in role R.* Thus an individual hydrogen atom does not expand at 0°C, only the H_2O molecules by virtue of the positions of two hydrogen atoms within each molecule. In general, the properties of the parts stand in different causal relations than the properties of their wholes. Yet explanations of this kind turn on statements about the parts taken individually, in the above example, statements about the position of individual atoms. Thus I will argue more completely in the next section that these and other features of a part-whole explanation for role occupation constitute a quite different theory than what is described by (F1) and (F2).

Objection 4. Someone might think there is a significant disanalogy between the problem for supervenience and the problem I raise for functional-role theory, based upon the goals of the theories and the assumptions that are permitted under those goals. Specifically, one might maintain that functional-role theorists simply assumed the physical acceptability of the occupier *G*, and that they were correct to do so, given that their goal was only to show the physical acceptability of *F*. That is, (F1) and (F2) of causal-role functionalism ensure that *F* is physically acceptable, on the assumption that *G* is physically acceptable. In contrast, the modal correlations (S1) and (S2) of

supervenience do not ensure that *F* is physically acceptable, on the same on the assumption that *G* is physically acceptable. One might then conclude that my focus on the physical acceptability of the occupier *G* is misplaced, or at least not the kind of worry that had originally troubled philosophers about supervenience. [12]

Nonetheless, there is more to the history and my thesis remains intact. To begin, functional-role theorists wanted to show the physical acceptability of *F*, the role property. But many were also explicitly concerned to explain mind-brain correlations between *G* and *F*. Here is a passage from Kim:

> The mental supervenes on the physical because mental properties are second-order functional properties with physical realizers (and no nonphysical realizers). And we have an explanation of mental-physical correlations. Why is it that whenever *P* is realized in a system *s*, it instantiates mental property *M*? The answer is that by definition, having *M* is having a property with causal specification *D*, and in systems like *s*, *P* is the property (or one of the properties) meeting specification *D* (Kim 1998, p.24).

So the target was not just some solitary and physically questionable *F*, leaving *G* out of the picture and hence removing *G* from consideration regarding its physical acceptability. Rather, the target was also the correlation between *G* and *F*, which brings *G* directly into the *explanandum* and which makes the physical acceptability of *G* a relevant issue. Yes there was a concern about *F's* over-and-aboveness *vis-à-vis G*. But there was also a concern about the over-and-aboveness of *any non-fundamental relation between F and G* (this is related to a familiar concern in the literature on scientific reduction over the status of "bridge laws" used in reductive explanations).

Also, putting the target *explanandum* to one side, the idea that functional-role theorists were permitted to assume the physical acceptability of *G* misses the fact that the

---

[12] I thank an anonymous referee for this very perceptive observation, as well as the next objection I discuss.

standards for physicalism discussed by the likes of Horgan and Kim require that *the resources of the theory used to explain F must be shown to be physically acceptable as well*. The standards were meant to provide a check on the *resources of the explanans*, not just the targeted *explanandum*. Supervenience was supposed to be good *explanans* to account for things like mental properties as well as mind-body correlations and similar relations, and yet bare supervenience was found to be physically unacceptable by the standards for physicalism put in play. Likewise, I argue that functional-role theory is supposed to be a good *explanans* to account for things like mental properties, mind-body correlations, and even mind-body supervenience, and yet bare functional-role theory is also found to be physically unacceptable by the main standard for physicalism that was put in play. Specifically, I focused on a "physical explanation" condition, which I formulated in a general way whereby *a non-fundamental theory must have the resources to show that its ontology is explainable in a physically acceptable way.* It is clear that causal-role functionalism does not satisfy that condition, because a crucial part of its ontology is the fact of role occupation for *G* expressed by (F2), which is left unexplained by (F1) and (F2), making the theory compatible with emergence (just like the fact expressed by (S2) was left unexplained by (S1) and (S2), making the theory compatible with emergence). [13]

[13] If someone objects to my general way of expressing the physical explanation condition, urging a more specific condition that requires only that the *G* => *F* connection must be explained in a physically acceptable way, that would be special pleading in the extreme. Granted, discussions in the literature were often framed in terms of supervenience laws. E.g., Jessica Wilson puts Horgan's complaint in terms of supervenience: "Any genuinely physicalist metaphysics should countenance ontological inter-level supervenience relations only if they are robustly explainable in a physicalistically acceptable way" (2002, p.55). See also Kim (2002, pp.36-37). This is understandable, since the subject was supervenience and physicalism. But, again, there is no good reason to exclude the facts of role occupation from worries over physical acceptability, if in fact one is concerned about the physical acceptability of the theory used in an explanation.

Moreover, any attempt to drive a wedge between questions about the physical acceptability of the occupier *G* and the physical acceptability of supervenience laws *G* => *F* ignores their similarity. The fact of role occupation is not *G* in isolation, but *G* standing in relation to causes and effects described by a functional theory. And, again, paradigm facts of role occupation are *just as non-fundamental* and *equally in need of explanation* as the facts of supervenience. Indeed, I add that, like the supervenience connection *G* => *F*, a fact of role occupation such as *C* => *G* => *E* is also *inter-theoretic* in nature (they are also equivalent on the matter of levels when both supervenience and functional-role theory are interpreted in the traditional "flat" way whereby *F* and *G* belong to the same object at the same mereological level – they are inter-theoretic statements about entities at the same mereological level). This inter-theoretic nature follows directly from the fact that property *F* and role *R* are specified in (F1) by a functional theory, whereas the occupier *G* is specified in (F2) by the appropriate realization or implementation theory.

Consequently, given this status as non-fundamental, inter-theoretic links between functional and realization theories, then saying that a neural *G* stands in a metaphysically determinative relation *vis-à-vis* an *F* specified by a psychological theory, and saying that a neural *G* stands in a causally determinative relation *vis-à-vis* distinct causes and effects specified in a psychological theory, ought to be equally questionable from a physicalist point of view. To put this claim in terms of Horgan's "standpoint question" (1993, p.578), when one asks what sort of facts, over and above the fundamental physical facts, could combine to yield physically kosher explanations for a mental *F* or a brain-to-mind correlation between *G* and *F*, a paradigm fact of role occupation is not one of the fundamental facts of physics. Rather, like the supervenience law *G* => *F*, it is an interesting inter-theoretic fact that raises the very same question about physical acceptability as the targeted *F* or the targeted correlation between *G* and *F*.

But what is the last and most important point, even granting that functional-role theorists were correct to assume the physical acceptability of *G*, given the standards for physicalism that were put in play, the other question is *whether that assumption carries one beyond the resources of functional-role theory*. I will argue that it does inasmuch as it requires a part-whole explanation that differs from the bare statement of functional-role

theory in important ways. Indeed, an affirmative answer to this question secures my thesis. To assume that *G* standing in role *R* is physically acceptable requires an explanation for role occupation that takes one beyond the bare statement of functional-role theory. Hence one needs a theory of superduperfunctionalism to supplement functionalism, just like one needs a theory of superdupervenience to supplement supervenience.

Objection 5. Finally, someone might think that my argument overlooks the resources that are available to functional-role theorists and which might explain role occupation. In the case of supervenience, its explanatory resources are exhausted by property correlations with different modal strength that hold for individuals, regions, or worlds. But causal-role functionalism offers additional explanatory resources, including obviously (i) the concept of a role and occupant, also (ii) facts about causation included in role *R* that define *F* and are occupied by *G*, (iii) reference to the occupying structural property *G*, and perhaps (iv) additional causal relations *R*′ that define the essence of *G*, either by an iteration of the role-occupant scheme that gives a functional specification for *G* in terms of *R*′ or by a causal theory of properties applied to *G*, all of which might provide the desired explanation for *G* occupying role *R*. [14] Yet, in response, whereas I agree that causal-role functionalism has more conceptual resources than supervenience (otherwise it could not explain supervenience), I deny that these resources provide an adequate explanation for role occupation that satisfies the standards for physicalism that

[14] Of course the appeal to structural properties and a causal theory of properties is not the prerogative of functional-role theory alone. One may also supplement supervenience with the same general metaphysical ideas. E.g., regarding Kim's definition of strong supervenience, let the supervenience base *B* contain structural properties that imply parts for their instances, and let the physical properties in *B* be individuated by a causal theory of properties. One might then argue, in a similar vein, that these additional resources supply an explanation why a subvenient *G* correlates with a supervenient *F*. But, in point of fact, I do not believe that the additional facts explain either supervenience or role occupation.

were raised against supervenience. Take this as a promissory note, to be fully cashed in over the next two sections where I highlight the difference between functional-role theory and the part-whole explanations required to explain role occupation. For now I will just briefly present my case.

To begin, I have already argued that resource (i) concerning the concept of a role and an occupant does not explain the fact of role occupation in the case of (F1) and (F2). (F1) merely equates having the functional property *F* with having some property that stands in those causal relations *R* without explaining how any property is able to do so, and (F2) expresses the target *explanandum* that *G* stands in causal relations *R*. In fact, the only way to explain (F2) using resource (i) would be to iterate the role-occupant scheme at a deeper level, creating a separate *explanans* that is constituted by the additional proposition (F1′) that defines *G* in terms of its own causal role *R′* as well as the additional proposition (F2′) that a still lower-level physical property *P* occupies that *R′*. But I will argue in two sections hence that this is a bad explanation for role occupation. So consider both resource (ii) concerning the facts about causation included in *R* and resource (iii) concerning reference to an occupying structural property *G*. Yet taken together this is just the *explanandum* (F2), not a candidate *explanans*. One does not have a good explanation by simply repeating what must be explained.

Granted, one might envision a familiar part-whole *analysis* of the structural property *G* that in effect takes apart its instances in order to understand the causal capacities that *G* bestows upon those instances (e.g., Cummins 1975, 1983; Craver 2001). But one must distinguish between *the explanandum* (F2), which is a statement that some object or system *x* has a structural property *G* that stands in causal relations *R*, versus the *explanans* provided by the aforementioned part-whole analysis, which is another set of statements about the parts of *x* and their part properties $P_i \ldots P_n$ and how they behave under various conditions that are relevant to understanding the causal capacities that *G* bestows upon *x*. This kind of explanation, as I will labor to show in the next section, differs from causal-role functionalism in important ways.

That leaves resource (iv) concerning certain additional causal relations *R*′ that provide the essence of *G*. Yet, similarly, one must distinguish between the causal relations that individuate the complex structural property *G* from the causal relations of

the component part properties P_i ... P_n implied by *G* (as well as other part properties utilized in a part-whole explanation for *G's* causal capacities). In the previous mind-brain case, one must distinguish between the causal relations that individuate *the entire neural system G* which occupies the role of face recognition versus the causal relations for the properties of the parts, like *being a neuron* that transmits signals to another neuron within the same system, or *being a sub-neural ion channel* that opens to allow positively charged atoms to pass, or *being an ion atom* whose entry into the cell body is crucial for depolarization and thus the neuron's capacity to signal to another cell (to think otherwise is to commit a part-whole fallacy). [15] For example, the entire neural system that recognizes faces is not caused to enter a cell body, and an ion atom does not send a signal to another area of the brain. But the cited facts about the parts make it clear that the kind of part-whole explanation in question offers information well beyond the causal relations of the target property *G*, and well beyond the resources of functional-role theory more generally. To that topic I now turn.

5. The Difference Between Part-Whole Explanations and Functional-Role Theory

I now want to show that the accepted scientific explanations for role occupation utilize ideas and information that are not contained in the basic statements of functional-role theory patterned after (F1) and (F2). The explanations take a familiar form. They are either *part-whole structural explanations* or *part-whole mechanistic explanations*. [16] I

---

[15] Put in a different way, on a standard causal theory of properties, one individuates *G* by its intra-level causal relations, not its inter-level realization relations that connect *G* to still lower-level properties. I developed an alternative theory that individuates properties by their total nomic relations, including inter-level realization relations (see Endicott 2007).

[16] John Haugeland (1978, p.216) called them "morphological" and "systematic" explanations. I add two points. One, the division is a popular one of convenience, for I think it is more accurate to view the relevant explanations along a continuum, where those that target relatively static structures are located at one end (the lattice structure of carbon atoms in a diamond), those that target more fluid structures are located in the

begin by developing the example of neurotransmission. The basic explanation is that neural system *G* has the ability to receive and transmit signals because each neuron in system *G* is such that, once it receives neurotransmitter molecules, this causes porous ion channels in a neuron's membrane to open, which causes positively charged ion atoms to enter its cell body, which causes the cell to depolarize, which thus sends a chemical signal to the next neuron (for more scientific details on neurotransmission, see Doyle, et. al. 1998; Jensen, et. al. 2012). According to Machamer, Darden, and Craver's (2000) influential analysis, this is a special kind of multiple-level, part-whole explanation that involves a systematic process, specifically, how the parts of the neural system work from a start-up condition whereby a pre-synaptic neuron releases neurotransmitter molecules, an intermediate stage whereby a post-synaptic cell receives the neurotransmitters, to an end-state condition whereby the post-synaptic neuron depolarizes and thus transmits a signal.

Some mechanistic explanations might involve more than multiple levels within a systematic process. For example, William Bechtel (2011) believes that many mechanisms in biology are not ordered in a simple sequential way from a start-up condition to an end-state condition. Rather, they have a cyclic organization with positive and negative feedback loops. Also, the kind of levels discussed by Machamer, Darden, and Craver are not simple part-whole levels, but part-whole levels constrained by the interests of scientists in those parts that serve a particular mechanism (2000, 13). But my goal is simply to distinguish the relevant kinds of part-whole explanations from the role-occupant scheme defined by (F1) and (F2), and three points about the foregoing part-

---

middle (H_2O in its different forms), and those that target mechanistic processes are located at the opposite end (the human brain with its systematic processes for cognition). Two, one might wonder how facts about particulars – the parts and wholes – could explain facts about properties and hence the targeted facts about occupier properties. But one can understand how a property occupies role by understanding how instances of that property stand in the relations constitutive of the pertinent role, and one can understand how instances of the property stand in relations constitutive of that role by means of the proffered part-whole explanations.

whole mechanistic explanation are enough to accomplish this goal. First, the explanation utilizes a *multiple-subject, part-whole theory*. There is reference to a neural system *x* that receives and sends signals vis-à-vis different areas of the brain by the property of neurotransmission *G*, and there is also reference to parts $y_i \ldots y_n$ that are distinct from *x*, with their part properties $P_i \ldots P_n$, for example, statements about ion channels.

Second, the explanation spans *multiple part-whole levels* rather than a single part-whole pair. Taken bottom up, there are details about positively charged atoms (atomic level), neurotransmitter molecules (molecular level), ion channels in the cell membrane (molecular and sub-cellular level), individual neurons (cellular level), and the system of neural cells that exhibit the targeted neurotransmission activity (system cellular level), all in the same explanation for the phenomenon of neurotransmission. [17] So there is reference to the system x, its parts $y_i \ldots y_n$ and subparts $z_i \ldots z_n$ that are distinct from *x*, with their part and subpart properties $P_i \ldots P_n$ and $Q_i \ldots Q_n$.

Third, the part properties introduced in the *explanans* are *smaller* relative to the target function of the whole in the *explanandum*. One may take this way of speaking about properties as a metaphorical extension of the way the particulars are commonly described, since the common sense notion of a concrete part denotes an object that occupies a smaller spatio-temporal region than the concrete whole of which it is a part. [18] But I intend to stipulate a less metaphorical and technical meaning that is directly relevant to the contrast with causal-role functionalism. I will say that *a property B is larger than or equal to a property A iff either B stands in all the causal relations R that*

---

[17] I mentioned earlier that Machamer, Darden, and Craver refer to this feature of multiple part-whole levels in terms of "nested hierarchies" (2000, 13). See also William Bechtel and Adele Abrahamsen (2005), Lindley Darden (2005), and Maureen O'Malley, et. al. (2014) for emphasis upon multiple levels in mechanistic explanation.

[18] This common notion of a part differs from technical and philosophical notions that allow a part to occupy the same spatio-temporal region as its whole, e.g., the notion of an improper part that allows for identity, or the reflexive notion of a part whereby everything is as a part of itself, or the notion that a non-identical but coincident object is a part of the whole.

define A as well as additional causal relations R′ that define B or A and B stand in the same causal relations, otherwise B is smaller than A. I will discuss how the occupiers of causal-role functionalism count as larger than or equal to the functional properties shortly. But the individual part properties cited in a mechanistic explanation do not stand in all the causal relations of the targeted property of the whole, and hence they count as smaller by this definition.

For example, being an ion atom does not recognize face-like stimuli, only a larger type of neural system; being a carbon atom does not pass a scratch test, only a larger type of lattice structure that contains many carbon atoms; and being an oxygen atom does not expand at freezing temperatures, only H_2O (recall that the expansion occurs because of the position of two hydrogen atoms within a perfectly bonded H_2O molecule). Granted, the *conjunction* of all the pertinent part properties might constitute a larger property that stands in all the causal relations of the target occupier *G* (then again it might not, depending upon whether the parts and properties selected as explanatorily relevant constitute a sufficient condition for the target property). [19] But the mechanistic explanation does not simply cite such a complex conjunctive property. To the contrary, it cites lesser components and their lesser properties individually, how these parts and properties behave in this particular area of the system at this stage in the process, and how those parts and properties behave in that area of the system at that stage in the process.

These same points apply to part-whole explanations that target structures whose parts do not exhibit either the kind or degree of systematic processes illustrated by paradigm mechanisms. So consider again the case of H_2O occupying the causal role of water, and recall the explanation discussed earlier regarding why freezing a body of H_2O causes it to expand (at 0°C and below there is a perfectly bonded H_2O molecule with a V-shaped H-O-H angle has an open space between the hydrogen atoms at the one end, but there is more thermal energy at temperatures above 0°C to break the hydrogen bonds and

[19] E.g., Carl Craver (2007, 160) maintains that the explanatorily relevant parts and properties cited in a mechanistic explanation do not necessarily enable one to derive the phenomenon targeted for explanation.

shake the hydrogen atoms out of position). This explanation also utilizes a *multiple-subject, part-whole theory*. There is reference to a body of water/H_2O *x* that exhibits the property of being H_2O (*G*) and which expands under freezing temperatures (*R*), and there is also reference to parts y_i ... y_n that are distinct from *x*, with their part properties P_i ... P_n, for example, statements about the positions of individual hydrogen atoms. The explanation also spans *multiple part-whole levels* with the body of water *x*, its molecular parts y_i ... y_n, and their subpart atoms z_i ... z_n. Finally, the properties introduced in the *explanans* are *smaller* relative to the target function in the *explanandum* in the sense that the individual part properties do not stand in all the causal relations of the targeted property of the whole. Thus an individual hydrogen atom does not expand at 0°C, only the H_2O molecules by virtue of the positions of two hydrogen atoms within each molecule. So, to summarize thus far, structural and mechanistic explanations are (I) multiple-subject, part-whole theories that (II) describe multiple part-whole levels wherein (III) the parts possess smaller properties in the sense that they do not stand in the same causal relations as the properties of their wholes. For convenience, call any theory like this a (PW) explanation.

Let me now return functional-role theory. I mentioned earlier that causal-role functionalism is a *single-subject* theory in the sense that (F1) and (F2) jointly describe the same object *x* that possesses role and occupant properties *F* and *G*. Moreover, the object is a complex structure by virtue of the fact that the occupier property *G* is a structural property. Therefore the first point of difference is this:

> I. (F1) and (F2) offer a single subject theory that attributes properties *F* and *G* to the same complex object or system *x*, but (PW) is a multiple-subject, part-whole theory that explains property *F/G* of that complex system *x* by describing certain parts y_i ... y_n and their part properties P_i ... P_n and how they behave under various conditions.

The distinction between single-subject and multiple-subject theories is common in the literature on supervenience (Kim 1993; McLaughlin 1995), and it should be respected in the case of functional-role theory defined by (F1) and (F2) versus a (PW) explanation.

Of course I do not deny that (F1) and (F2) express some part-whole structure by virtue of the fact that the occupier *G* is a structural property. But, to repeat an earlier point, one must distinguish between a statement that some complex object or system *x* has a structural property *G* that stands in its causal relations versus the quite different set of statements in a (PW) explanation about the parts and subparts of *x*, their part and subpart properties, and how they behave under various conditions that are relevant to understanding the fact that *G* stands in its causal relations. Moreover, the (PW) explanation supplies a much greater amount of information about the parts than what is expressed by a statement like (F2). In the case of H_2O occupying the role of water, for example, (F2) implies only that the occupier has instances with two hydrogen atoms and one oxygen atom. But the part-whole explanation regarding how H_2O is able to occupy the water role provides much more information even for the one aspect of the water role that concerns the expansion of water, including how H_2O behaves under differing amounts of thermal energy at different temperatures, how that affects perfect versus imperfect bonds at those different temperatures, how that changes the space between the hydrogen atoms, and how that subsequently changes the space occupied by the whole molecules at those temperatures. The structural information supplied by a typical statement of role occupation is impoverished compared to the rich information provided by a multiple-subject, part-whole explanation. [20]

---

[20] I mean "information" in a descriptive sense, since one does not possess the pertinent information merely by credit of the concept of H_2O on a purely causal or reliabilist or otherwise externalist theory of meaning. One possesses the scientific information by means of encoded theories about H_2O that were articulated by experts and transmitted *via* descriptions, charts, graphs, and other representational items. Also, one could in priniciple laboriously pack all the needed information into an incredibly lengthy structural predicate that applies to the same *x*. Indeed, one could in principle pack all the information about the entire universe into an incredibly lengthy relational predicate, pace Leibniz, that expresses the "complete concept" of an individual *x*. But, again, a (PW) explanation enters to break down the property picked out by that predicate, thus constituting an acceptable explanation.

Next, and a related point, it is also true that causal-role functionalism presents a more limited picture when compared to the multiple part-whole levels in a (PW) explanation. Some claim that (F1) and (F2) describe a single mereological level, which is to say that causal-role functionalism is metaphysically "flat" (see Gillett 2002, 2003). [21] *F* and *G* are instantiated by the same complex *x*, as opposed to *F* being instantiated by *x* and *G* being instantiated by a part of *x*. Moreover, *G* could not occupy the causal role of *F*, or stand in all the causal relations of *F*, if *F* were a property of a whole system *x* and *G* were a property of a part of *x* that is instantiated by a part of *x* that exists at a lower mereological level. Again, a single neuron does not recognize faces, only a larger system of neurons.

But the flat claim is contentious because the occupier *G* is a structural property that implies parts for its instances. So I will include what is implied by the description of the structural property, namely, that causal-role functionalism, defined by (F1) and (F2), spans not one but two levels of a single part-whole pair. In the previous mind-brain case of face recognition, *being a system of neurons* is instantiated at one mereological level, whereas the part property that is implied by that description, namely, *being a neuron*, is instantiated at one mereological level below. Likewise, *being* H_2O is instantiated by *x*, and it implies the single lower level of hydrogen and oxygen atoms. And again, *being a lattice structure of carbon atoms* is instantiated by *x*, and it implies the single lower level of individual carbon atoms. Certainly philosophers have understood functional-role theory in a restricted mereological way, otherwise one could not make sense of the idea that roles and occupants must be iterated down the many mereological levels of nature.

---

[21] One might reject the flat claim for the wrong reasons. E.g., given a nonreductive interpretation of (F1), one might fail to see the difference between the two property "orders" of *F* and *G* versus the present issue about mereological "levels" for the particulars that instantiate *F* and *G* (see Kim 1998, pp.80-83). Or one might confuse the fact that causal-role functionalism is "inter-theoretic" in nature by having a functional theory specify "*F*" and a realization theory specify "*G*" with the different issue of being "inter-level" in nature by a mereological criterion.

Consider how William Lycan applies the scheme of roles and occupants to the many mereological levels of nature:

> See Nature as hierarchically organized in this way, and the "function"/ "structure" distinction *goes relative*: something is a role, as opposed to an occupant, a functional state as opposed to a realizer, or vice versa, only *modulo* a designated level of nature ... Physiology and microphysiology abound with examples: *Cells* – to take a conspicuously functional term (!) – are constituted of cooperating teams of smaller items including membrane, nucleus, mitochondria, and the like: these items are themselves *systems* of yet smaller, still cooperating constituents (1987, p.38).

Lycan re-applies the role-occupant distinction *ad seriatum* or one mereological level deeper at a time, from (relative functional) *being a cell* to (relative occupant) properties such as *being membranes and nuclei and mitochondria*, then again from (relative functional) *being membranes and nuclei and mitochondria* to (relative occupant) smaller part properties such as *being DNA*. Or again, after describing his *homuncular* view, Lycan says explicitly: "the psychologist will first explain the behavior and behavioral capacities of the whole person in terms of the joint behavior and capacities of the person's *immediately* subpersonal departments, and if deeper and more detailed explanation is desired, the psychologist will explain the behavior of the departments in terms of the joint behavior and capacities of their joint components, and so on down as far as anyone might care to go" (1988, pp.5-6, italics mine). I think there is also a plausible explanation for this step-by-step application by functional-role theorists. Namely, philosophers have developed functionalism as a metaphysical picture of the world. The iterations of functional-role theory thus follow the metaphysical levels of the world, one layer at a time, with each layer metaphysically sufficient for the one above. But a (PW) explanation takes a deeper view at each application, encompassing as many part-whole levels as are sufficient to explain the workings of the targeted mechanism. Hence the second point of difference is this:

> II. (F1) and (F2) describe either a single mereological level (*F* and *G* possessed by the same complex *x*) or the two levels of a single part-whole pair (the complex *x* that possesses the structural property *G* and the single level of parts with their properties implied by *G*), but (PW) spans multiple part-whole levels.

Finally, recall the technical notion that a property *B* is larger than or equal to a property *A* iff either *B* stands in all the causal relations *R* that define *A* as well as additional causal relations *R′* that define *B* or *A* and *B* stand in the same causal relations, otherwise *B* is smaller than *A*. On a nonreductive version of causal-role functionalism, an occupier *G* is larger than the functional property *F* because it stands in all the causal relations *R* that define *F* along with additional causal relations *R′* that define *G*, either by an iteration of the role-occupant scheme that gives a functional specification for *G* in terms of its own role *R′* or by a causal theory of properties that defines the essence *G*. On the reductive version of causal-role functionalism *F* = *G*, and hence *F/G* stands in the very same causal relations *R* and *R′*. [22] But, as I have already discussed, none of this is true for the properties described in a (PW) explanation. A part property does not stand in all the causal relations of the whole property whose causal capacities it serves to explain. So whereas properties become larger or remain equal as one moves down the property orders of a role-occupant scheme, the properties become smaller as one moves down the mereological levels described in a (PW) explanation. Hence the third point of difference is this:

---

[22] Functional-role theory is often supplemented with a subset view of realization (see Shoemaker 2007; Wilson 1999). So the point can be put alternatively by saying that an occupier *G* is larger than or equal to the functional property *F* because the causal powers of *F* are a subset of the causal powers of *G* (a proper subset for the nonreductive view). This, again, is not true for the part properties $P_i \ldots P_n$ vis-à-vis the function *F/G* of the whole they serve to explain.

III. (F1) and (F2) describe larger or equal properties at lower orders, but (PW) describes smaller properties at deeper levels.

I conclude that a (PW) explanation differs from causal-role functionalism as defined by (F1) and (F2) by at least three measures, which is to say that its conceptual resources differ from the conceptual resources of causal-role functionalism. But a (PW) explanation also provides the correct scientific explanation for the fact of role occupation described by (F2). So I will call the conjunction of (F1), (F2), and (PW) a theory of "superduperfunctionalism." As such, (PW) provides something that (F1) and (F2) do not. Specifically, whereas (F1) states *what F* is by means of an essence specifying causal role, and (F2) states *that G* stands in the relations constitutive of that role, (PW) enables one to understand *how G* stands in those relations. [23] Let me also reinforce the parallel with supervenience. The problem with supervenience was generated by the bare statement of supervenience rather than the inclusion of additional propositions that constitute a theory of superdupervenience. That is why philosophers used modifiers and qualifiers to express their complaints about "mere" supervenience (Horgan 1993, p.565), or "bare" supervenience (Horgan 1993, p.566), or supervenience "in itself" (Kim 1998, p.12). Likewise, the present problem is generated by the bare statement of causal-role functionalism rather than the inclusion of additional propositions that constitute a theory of superduperfunctionalism.

Let me also emphasize that accepting a (PW) explanation for role occupation is not new. Jerry Fodor accepted a decompositional analysis of mental functions (1968) and the role-occupant scheme (1981), and presumably that synthesis was achieved by letting the decompositional analysis target the physical occupiers of the roles associated with mental functions, as I have indicated. Similarly, Robert Cummins (1983, p.21) described how his general part-whole property instantiation theory and a causal transition theory fit

[23] I make a parallel point with respect to a synthesis of flat functional-role theories of realization with part-whole dimensioned theories of realization (Endicott 2011), though my discussion does not concern supervenience or the standards for physicalism.

together. [24] Yet his property instantiation theory is the direct ancestor of recent mechanistic theories that I include in a (PW) explanation (see Craver 2001 for the connection). Likewise, William Lycan (1987) included both he idea of functional roles and occupants along with a decompositional analysis as part of his homuncular functionalism (although his notion of function was teleological, not purely causal). But I have suggested a new way of looking at some old facts, one that makes supervenience and its supposed physical guarantor, functional-role theory, appear equally unable to preserve the doctrine of physicalism on their own without help from the resources of a different kind of theory.

Finally, as an intriguing "trailer" for a concluding observation, I have purposely not claimed that all (PW) explanations will by themselves satisfy the standards for physicalism in question. To be sure, the sample part-whole structural and mechanistic explanations are grounded in physics – each bottom out in facts about atoms and their properties – neurotransmission is based on the distribution of positive versus negatively charged atoms within neural cells; H_2O's expansion is based on the space between hydrogen atoms. But deeper non-physical hypotheses are possible unless the proffered explanations bottom out in *fundamental* physics. I will return to this point later. My aim in the present section was only to show that the facts of role occupation have good scientific explanations that utilize conceptual resources beyond functional-role theory proper.

---

[24] Someone might worry that the causal dispositions to which Cummins refers are not captured by the laws of causal-role functionalism, at least on a nomic regularity interpretation of those laws (see Martin 1994). I think these worries can be allayed (e.g., see Choi 2006, 2008). Or one might reinterpret causal-role functionalism in terms of dispositions by mapping the inputs, internal states, and outputs onto the triggering conditions, dispositions, and their manifestations. However that may be, Cummins maintains that psychological laws are often the data to be explained (2000), which is perfectly consistent with using his functional analysis as a (PW) explanation in the way suggested here.

6. Why Role-Occupant Iterations Are Bad Explanations for Role Occupation

In spite of the accepted scientific explanations for role occupation just discussed, some philosophers might want to explain role occupation in a different way that utilizes only the conceptual resources of the basic functional-role theory, specifically, by *iterating* the role-occupant scheme. Lycan (1987) was the first to explicitly present an iterated role-occupant scheme, although the idea was arguably implicit in earlier discussions that extend causal-role functionalism to areas outside the mind and brain (as in Fodor 1974). Still, one could maintain that the world displays a repeating pattern of roles and occupants without maintaining that one explains the other, as I will discuss shortly. So, as an example of the explanatory proposal, consider Michael Tye's (1995) discussion about realization and explanatory mechanisms.

According to Tye, realization is a form of synchronic inter-level determination that is mediated by an implementing mechanism (1995, pp.41-42). Moreover, Tye understands this picture of realization by a general model akin to the second-order version of causal-role functionalism. He expresses the model both in terms of dispositions (1995, p.47) and higher-order functional properties (1995, p.48). To use Tye's example, one understands the mechanism by which a diamond's hardness is generated by knowing that (F1) *x* has a disposition or functional property *F* (hardness) whose essence requires that *x* has a constitutional property that disposes *x* to *V* (resist penetration), or whose essence requires that a constitutional property stand in the relations constitutive of role *R* (resist penetration, pass a scratch test, and so on), and (F2) *x* has a physical constitutional property *G* (a lattice structure of carbon atoms) that disposes *x* to *V* or that stands in the relations constitutive of *R*.

But Tye is aware that (F2) must be explained, and thus he adds: "Of course, the particular law appealed to here, namely, that objects having the lower-level property are disposed to *V*, itself demands an explanation if it is not microphysical. Further mechanisms and still-lower-level laws will be relevant to *this* explanation" (1995, p.47). The law that *objects having the lower-level property G are disposed to V* (on the dispositional version), or the law that *objects having a lower-level property G play role R* (on the functional version), is the occupying fact (F2) stated with nomological necessity. And Tye's remark that one can explain this law of role occupation by "further

mechanisms and still-lower-level laws" seems to suggest the idea that one can explain occupying facts by postulating *further mechanisms and laws of the kind just described*, that is, by mechanisms and laws that conform to the general model of (F1) and (F2), thus iterating the role-occupant scheme. So I will interpret Tye's remarks to mean that one can explain (F2) *x has a physical property G that occupies causal-role R* (e.g., where *R* is a set of causal relations described by a psycho-functional theory) by citing an essence specifying definition for *G*, *(F1′) x has functional property G = x has a physical property that occupies causal-role R′* (e.g., where *R′* is a set of causal relations described by neuroscience), along with the fact that there is an occupier for *G*'s role, *(F2′) x has a still-lower-level physical property P that occupies causal-role R′* (e.g., where *P* is described in chemistry or physics).

Before I criticize the position, let me make three preliminary points. First, one might think the iterative strategy is plausible on grounds that if a functional-role explanation in the form of (F1) and (F2) is a good explanation for some other target *explanandum*, then the strategy is a good explanation when it is redeployed to target a fact of role occupation (F2) by (F1′) and (F2′). But one might reject the antecedent, given that the mere assertion of (F1) and (F2) does not satisfy the standards for physicalism that were raised against supervenience (I will return to this point in the final section). Or one might reject the inference. An iteration of a good thing is not always a good thing – repeating a good meal for days on end, or reproducing more children on a limited budget, or reusing an evasive tactic under the watchful eye of a predator are examples.

Second, the issue is not simply whether the role-occupant distinction can be iterated. That can be true in a world with the structure required by the proffered part-whole explanations (PW). So recall that Lycan relativized the role-occupant distinction to the many mereological levels of nature. Consequently it is certainly possible – I argue more plausible – to offer a part-whole explanation gleaned from the mereological structure of the world even if there are iterations of roles and occupants. On the view I suggest, the iterations only *describe* different levels of nature but do not *explain* them (cf. the repeating pattern of colors that run down a North American Coral Snake – the iterated pattern is true of the snake, but one iteration does not explain another). Call the

explanatory interpretation of role-occupant iterations "ex-iterations." I reject the ex-iterations, not the iterations.

Third, I assume that a functional-role theorist is not permitted to stipulate that an occupier referred to within a proposed role-occupant iteration is a multiple part-whole level property picked out by a description that is extensionally equivalent to a (PW) explanation. That would betray the logic of an iterated role-occupant scheme, which applies *ad seriatum* or one additional level at a time, as Lycan described. In fact, the stipulation in question would make the iterative strategy pointless, since, if legitimate, the functional-role theorist could have postulated a multiple part-whole level occupier *G* at the outset, thus removing the need to explain how a single level mechanism *G* is able to perform its causal task by appealing to iterations of still lower-level facts about the parts of the mechanism that enable one to understand how that mechanism works.

Now for the argument. I offer the following refutation by analogy to show that ex-iterations are bad explanations if they are not supplemented by a (PW) explanation. Suppose I am able to "stand in" for my son in some capacity, say, I am able to be the guarantor for his bank loan. Suppose further that I go with my son to the bank and he makes a request for that loan. Naturally the bank officer will ask for proof of repayment, and at this point my son could give two answers, one good and the other bad. The good answer would be for my son to supply the bank officer with a financial analysis or "breakdown" of my assets versus my debts, showing that the former add up to a larger sum than the latter, that the difference is large enough to pass a reasonable threshold for repayment, and so on. The bad answer would be for my son to say that his father is himself a son whose mother will stand in for his debt, or again, that the grandmother has someone else to stand in for the debt. Unless my son is offering a mere scam, he has never answered the loan officer's request for an explanation regarding how he or how anyone else can repay the loan until he stops repeating the debtor-guarantor structure and gives a plain reckoning of his or mine or someone else's financial status. The good answer is analogous to a part-whole analysis of a capacity. The bad answer is analogous to a role-occupant ex-iteration.

I think the analogy is strong: (a) *a causal role = a capacity to repay a debt;* (b) *the role-occupant distinction allows an item to occupy a role that was defined for another*

= the debtor-guarantor system allows someone to assume a debt that was incurred by another; (c) *a part-whole explanation of a causal capacity = a financial analysis of the capacity to repay a debt;* and (d) *appealing to role-occupant iterations without part-whole explanations = appealing to a series of debtor-guarantors without an analysis of any one individual's capacity to repay the debt*. Notice too that each person in the named series of people may have a larger financial capacity – deeper pockets to mimic the larger causal capacities that are presumably associated with properties at deeper levels of reality in a nonreductive role-occupant ex-iteration. But claims about larger capacities, and the larger ones standing in for the smaller ones, must be backed by an analysis of those capacities, otherwise, to switch the analogy, it is just smaller turtles standing on larger turtles all the way down. [25]

Let me also point out that, although the foregoing analogy fits a nonreductive physicalist version of causal-role functionalism (son, father, and grandmother are numerically distinct), the iterated pattern can be easily adapted to a reductive theory. Suppose my son goes to the bank and requests a loan, as before. He applies under the name "Ethan Alexander." The loan officer asks how he is able to pay for the loan, but now my son answers by saying that he goes by another name, "Fernando Alexandro," and under that name he is recognized to have a larger financial capacity and is thus able to back the loan. This may be an unusual answer, but regardless the loan officer will certainly press for a financial analysis of my son's capacity to repay the loan under his more recognized name, and hopefully my son would not repeat the same strategy again. That is, my son must provide a plain reckoning of his capacity to pay the debt that is not forestalled by new names and additional role-occupant claims.

Finally, I think there is a plausible diagnosis of the problem. The role-occupant scheme requires complex structural properties for its occupier properties, otherwise it could not be iterated over the mereological levels of the world. Yet these metaphysical complexes require a part-whole analysis or explanation for their instances, which is

---

[25] A (PW) explanation is not turtles upon turtles, but turtles composed of organs, composed of cells, composed of molecules, composed of atoms, until the fundamental level of physics is reached. One is silly. The other is science.

precisely what role-occupant ex-iterations never deliver. This complaint is similar to the familiar point that a series of postulated functional homunculi must be "discharged" by appeal to physical mechanisms (Dennett 1978, pp.123-124), only I make no assumption that properties defined by causal or other types of roles must be tied to systems that are treated like intentional agents. One does not appeal to a part-whole structural or mechanistic explanation because the system is treated as if it were an intentional agent. One appeals to a part-whole structural or mechanistic explanation because the system is a physical structure or mechanism whose properties are understood in terms of its parts.

7. Concluding Observations

I have argued that the bare statement of functional-role theory is compatible with unwanted non-physical views because it leaves the facts of role occupation unexplained, thus violating the physical explanation condition that was raised against supervenience. I have also shown that sample scientific explanations for role occupation utilize theoretical resources beyond functional-role theory proper, and I have argued that it is far less plausible to explain the facts of role occupation by iterating the role-occupant scheme. Barring other suggestions, the net result is that one cannot justifiably remain within the confines of functional-role theory and satisfy the kind of standards for physicalism that were raised against supervenience.

I want to close by addressing three things: the scope and limits of my argument, why philosophers failed to draw the parallel with supervenience, and whether the standards raised against supervenience are in fact correct. First, I have only argued against bare functional-role theory as defined by (F1) and (F2). But there are versions of functional-role theory that contain additional ideas beyond (F1) and (F2). For example, Terence Horgan and Mark Timmons (1992) articulate a species of "semantic constraint satisfaction explanations" that cites semantic principles about how the extensions of functional terms are fixed across counterfactual possibilities, physical facts that contribute to their interpretation, and systems of lower-level laws. Moreover, there are theories that purport to explain supervenience that are distinct from functional-role theory, such as Jessica Wilson's (2002) subset theory of realization that gives central place to fundamental forces (see also Wilson 2011). Certainly the subset theory is similar

to functional-role theory since it is also a single-subject theory that treats *G* as a larger property than *F* by containing its powers (see again fn. 22). But I do not have the space to consider these views here. Such theories should be examined on a case-by-case basis in order to determine whether their additional theoretical resources provide an explanation for role occupation.

Second, it is worth considering why philosophers failed to draw the parallel with supervenience. I think there are several possible reasons. For example, some philosophers might have thought that role-occupant ex-iterations are a viable kind of explanation. Also, some philosophers might have focused more on what functional-role theory can explain (the *explanandum*) rather than functional-role theory itself (the *explanans*). Specifically, some philosophers might have focused more on the *F-to-G* relation as it pertains to a target supervenience relation rather than the *G-to-R* relation as it pertains to functional-role theory itself. Accordingly, whereas they were rightly concerned to rule out the unexplained emergence of a supervenient *F* from a subvenient *G*, they failed to notice that their assumption regarding the physical acceptability of *G* standing in role *R* depends upon a theory other than functional-role theory, that is, a theory of superduperfunctionalism that is not (F1) and (F2), or iterations thereon.

What is a related point, some philosophers might have been so impressed with the contrast between supervenience and functional-role theory on the point of *explanation* that they lost sight of the goal to provide a *physically acceptable explanation*. Consider Kim's claim that "supervenience itself is not an explanatory theory" (1998, p.14). Kim says this because he believes that supervenience merely records property correlations but does not explain them. Of course one may also say that functional-role theory merely records the facts of role occupation but does not explain them. But let us grant, for the sake of argument, that there are many contexts of inquiry such that *functional-role theory provides an explanation but supervenience does not*. Even so, for present purposes this is the wrong contrast. When the question of physicalism has been raised, and when specific standards for physicalism like the physical explanation condition have been put into play, functional-role theory must provide a physically acceptable explanation by those

standards, not just an explanation *per se* that is better than supervenience. [26] That, I have argued, it fails to do.

Another reason why philosophers might have failed to draw the parallel with supervenience is that some conceive of role occupation in a more expanded way that combines (F2) with a (PW) explanation. Consider Joseph Levine's remarks about functional-role explanation, which he calls "explanatory reduction":

> Note that on this view explanatory reduction is, in a way, a two-stage process. Stage 1 involves the (relatively? quasi?) *a priori* process of working the concept of the property to be reduced 'into shape' for reduction by identifying the causal role for which we are seeking the underlying mechanisms. Stage 2 involves the empirical work of discovering just what those mechanisms are (1993, p.132).

This *seems* like the standard two steps of a functional-role explanation designated earlier as (F1) and (F2). But Levine intends something more than the mere statement of role occupation for the empirical process of "discovery" at stage 2, which is shown by his example just prior: "We justify the claim that water is H_2O by tracing the causal responsibility for, and the explicability of, the various superficial properties by which we identify water – its liquidity at room temperature, its freezing and boiling points, etc. – to H_2O" (1993, p.131). Levine includes the *evidence* for the identity that is both causally responsible for and explains the properties by which one identifies water. But these properties are explained by the appropriate set of (PW) explanations, exactly as I illustrated with the expansion of water at 0°C. In other words, Levine's stage 2 includes

---

[26] To use a well-known example from Michael Scriven (1962), in most everyday contexts it might be perfectly acceptable to explain why there is a ink spot on the carpet by saying merely that the ink well was spilled. But in contexts where one has raised the issue of physicalism, Scriven's answer will not suffice since it does not indicate that the ink well was spilled in a physically acceptable way rather than by telekinesis or some divine intervention or as some inexplicable emergent fact.

the statement of role occupation (F2) along with (PW). Consequently, by viewing role occupation in conjunction with the evidence and explanations involved in its discovery, the bare statement of causal-role functionalism represented by (F1) and (F2) is not isolated for evaluation by the standards for physicalism, as supervenience was isolated for evaluation by the standards for physicalism.

Third, and finally, let me make some observations about the standards for physicalism. Recall again that the most influential criticisms of supervenience are based upon its failure to meet the physical explanation condition whereby a non-fundamental theory must have the resources to show that its ontology is explainable in a physically acceptable way. Now I think there is something right about linking physicalism to concerns over explainability. But this is not to say that standards about explanation express this concerns in the best way. Notice that the standard in question is applied *locally* inasmuch as each non-fundamental theory, taken individually, must have the resources to show that its ontology is explainable by physical facts alone. Thus Horgan says that "a materialistic position should assert that all supervenience facts are explainable" (1993, p.560), the implication being that the bare statement of supervenience does not say this. Also, the standards are motivated by the belief that each physically acceptable non-fundamental theory, taken individually, should be *inconsistent* with non-physicalist positions. So Horgan complains that "physical supervenience is consistent with the central doctrines of British emergentism" (1993, p.560), a consistency that is removed by adding the explanations provided by a theory of superdupervenience (1993, p.566). [27]

But consider again to the accepted scientific explanations for role occupation discussed earlier. Suppose one provides what seems to be a physically acceptable mechanistic explanation regarding how neurons occupy the role of information processing associated with mental properties by citing the fact that neurotransmitters cause ion channels to open within the cell membrane of a neuron, which allows positively

[27] The idea is also implicit in Kim's (1990, 1993, 1998) discussions of physical dependence, since a non-fundamental theory guarantees an explanatorily relevant form of physical dependence only if it excludes non-physicalist views.

charged atoms to enter, which causes the neuron to depolarize, which thus sends a signal like an electronic circuit. Yet this explanation is silent about the happenings at the basic level of quantum mechanics. Indeed, whereas one might cite facts about the distribution of protons and electrons in an explanation of depolarization, one typically does not cite facts about quantum events that ground this behavior (See again Machamer, Darden, and Craver 2000, p.13). *A fortiori* such explanations are compatible with absolutely crazy things at the quantum level, including objectionable non-physical things – as was proposed by the Nobel Prize winning physicist Eugene Wigner, who suggested that "observations" as literal macro-level acts of awareness remove quantum indeterminacy and thus yield the determinate values for the things cited in non-fundamental explanations (Wigner 1967; see also Sklar 1992, chap.4).

The problem is perfectly general. Any non-fundamental theory has a limited domain consisting of, say, an *x* that causes *y* that causes *z*. But, by virtue of the fact that it is a non-fundamental theory, it gives no account of the more basic entities that explain *x*, *y*, and *z*. Consequently, it is consistent with an objectionable brute but non-fundamental facts. Yet surely *some* non-fundamental theories are physically acceptable. So I conclude that something is wrong with the pertinent standards for physicalism. I thus recommend the weaker requirement that every non-fundamental theory, taken individually, should *not imply* non-physical positions (rather than be inconsistent with or exclude non-physical positions). The mechanistic explanation for information processing does not imply non-physical positions, since it only describes physically acceptable things like protons and electrons, neurotransmitter molecules, and neural cells, remaining moot on all else. Likewise, bare supervenience and functional-role theory do not imply non-physical positions either. In the case of bare supervenience, for example, one must add additional claims about higher-level inexplicability to yield the position of British emergentism, such that there are higher-level irreducible laws, that they cannot be predicted from fundamental facts, and that they are unexplainable (see Kim 2006).

I also recommend, in tandem, that the condition for physical explainability be applied *globally* to one's total theory of the world rather than locally to each non-fundamental theory. That is, *a total theory of the world must have the resources to show that the ontology of all non-fundamental theories is explainable in a physically*

acceptable way. So bare supervenience, functional-role theory, and the accepted structural and mechanistic explanations are all physically acceptable because they do not imply non-physical positions, and because they are part of a total picture of the world that is inconsistent with non-physical positions. [28]

To summarize, then, my main negative thesis has been that functional-role theory fails by the standards for physicalism that were raised against supervenience because it leaves the facts of role occupation unexplained. But I also proposed a positive thesis that the facts of role occupation are best explained by part-whole structural and mechanistic explanations that function as a kind of superduperfunctionalism. Yet in order to block a similar threat to their physical acceptability, I then revised the standards for physicalism, which results in a more complicated conclusion. By the original standards, supervenience and functional-role theory count as physically unacceptable theories. By the revised standards, supervenience and functional-role theory count as physically acceptable theories. Either way, the parallel between supervenience and functional-role theory is preserved.

Acknowledgements

I thank two anonymous referees for some insightful comments. I also thank Terry Horgan for discussing an earlier draft of this paper, and Tom Polger for discussing the topic of role-occupant iterations covered in section 6.

[28] An anonymous referee worried that my conditions might be too weak, since one should want some guarantee that a given claim of realization or functional role satisfaction is physically acceptable, not merely the weaker claim, that if one is lucky and certain contingent possibilities turn out to be actual, a realized or functionally characterized property will be physically acceptable. But, in response, I assume that the pertinent theories which supply explanations for role occupation, from psychology down to quantum mechanics, are good scientific probabilities and not just contingent possibilities. Consequently, if one is a scientific functionalist, and if all but the fundamental level is thereby explained, that should be sufficient for physically acceptability.

References

Armstrong, D. (1968). *A materialist theory of the mind*. NY: Humanities Press.

Bedau, M., and Humphries, C. (2008). *Emergence: contemporary readings in philosophy and science*. Cambridge MA: MIT Press.

Bechtel, W. and Abrahamsen, A. (2005). Explanation: a mechanistic alternative, *Studies in History and Philosophy of the Biological and Biomedical Sciences*, 36 (2), 421-41.

Bechtel, W. (2011). Mechanism and biological explanation, *Philosophy of Science* 78 (4), 533-57.

Block, N. (1980). What is functionalism?, in N. Block, ed., *Readings in Philosophy of Psychology* 1 (pp.171-84). Cambridge MA: Harvard University Press.

Carroll, J. (1994). *Laws of nature*. Cambridge UK: Cambridge University Press.

Choi, S. (2006). The simple vs. reformed conditional analysis of dispositions, *Synthese* 148 (2), 369-79.

_____. (2008). Dispositional properties and counterfactual conditionals', *Mind* 117 (468), 795-841.

Craver, C. (2001). Role functions, mechanisms, and hierarchy, *Philosophy of Science* 68 (1), 53-74.

_____. 2007. *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. New York, NY: Oxford University Press.

Cummins, R. (1975). Functional analysis, *Journal of Philosophy* 72 (20), 741-65.

_____. (1983). *The Nature of Psychological Explanation*. Cambridge MA: MIT Press.

_____. (2000). 'How does it work?' vs. 'what are the laws?' Two conceptions of psychological explanation, in F. Keil and R.Wilson, eds., *Explanation and cognition* (pp.117-44). Cambridge MA: MIT Press.

Darden, L. (2005). Relations among fields: mendelian, cytological and molecular mechanisms, *Studies in History and Philosophy of Biological and Biomedical Sciences* 36, 357-71.

Dennett, D. (1978). *Brainstorms: philosophical essays on mind and psychology*. Cambridge MA: Bradford Books.

Dolye, D., Cabral, J., Pfuetzner, R., Kuo, A., Gulbis, J., Cohen, S., Chait, B., McKinnon,

R. (1998). The structure of the potassium channel: molecular basis of K+ conduction and selectivity," *Science* 280 (69), 69-77.

Endicott, R. (2007). Nomic-role nonreductionism: identifying properties by total nomic roles, *Philosophical Topics* 35 (nos.1&2), 217-40.

_____. (2011). Flat versus dimensioned: the what and how of functional realization, *Journal of Philosophical Research* 36, 191-208.

Fodor, J. (1968). The appeal to tacit knowledge in psychological explanation, *Journal of Philosophy*, 65 (20), 627-40.

_____. (1974). Special sciences: or the disunity of sciences as a working hypothesis, *Synthese* 28 (2), 97-115.

_____. (1981). Something of the state of the art, in *RePresentations: philosophical essays on the foundations of cognitive science* (pp.1-31). Cambridge MA: MIT Press.

Gillett, C. (2002). The dimensions of realization: a critique of the standard view, *Analysis* 62 (4), 316-22.

_____. (2003). The metaphysics of realization, multiple realizability, and the special sciences, *Journal of Philosophy* 100 (11), 591-603.

_____. (2007). Understanding the new reductionism: the metaphysics of science and compositional reduction, *Journal of Philosophy* 104 (4), 193-216.

Grimes, T. (1988). The myth of supervenience, *Pacific Philosophical Quarterly* 69 (June), 152-60.

Hameroff, S. and Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness, in S. R. Hameroff, A. W. Kaszniak & A. C. Scott, eds., *Toward a Science of Consciousness I* (pp.507-40). Cambridge MA: MIT Press.

Haugeland, J. (1978). The nature and plausibility of cognitivism, *Behavioral and Brain Sciences* 1 (2), 215-26.

Hawthorne, J. (2002). Blocking definitions of materialism, *Philosophical Studies* 110 (2), 103-13.

Horgan, T. (1993). From supervenience to superdupervenience: meeting the demands of a material world, *Mind* 102 (408), 555-86.

Horgan, T. and Timmons, M. (1992). Troubles on moral twin-earth: moral queerness

revived, *Synthese* 92 (2), 221-60.

Jensen, M., Jogini, V., Borhani, D., Leffler, A., Dror, R., and Shaw, D. (2012). Mechanism of voltage gating in potassium channels, *Science* 336 (229), 229-33.

Kanwisher, N., McDermott, J., and Chun, M. (1997) The fusiform face area: a module in human extrastriate cortex specialized for the perception of faces. *Journal of Neuroscience* 17 (11), 4302-11.

Kim, J. (1984). Concepts of supervenience, *Philosophy and Phenomenological Research* 45 (2), 153-76.

_____. (1990). Supervenience as a philosophical concept, *Metaphilosophy* 21 (1), 1-27.

_____. (1993). Postcripts on supervenience, in *Supervenience and mind: selected philosophical essays* (pp.161-71). Cambridge UK: Cambridge University Press.

_____. (1998). *Mind in a physical world*. Cambridge MA: MIT Press.

_____. (2002). Horgan's naturalistic metaphysics of mind, *Grazer Philosophische Studien* 63 (1), 27-52.

_____. (2005). *Physicalism, or something near enough*. Princeton NJ: Princeton University Press.

_____. (2006). Emergence: core ideas and issues, *Synthese* 151 (3), 547-59.

_____. (2011). *Philosophy of Mind*, third edition. Boulder CO: Westview Press.

Levine, J. (1993). On leaving out what it's like, in M. Davies and G. Humphreys, eds., *Consciousness: psychological and philosophical essays* (pp.121-36). Oxford: Blackwell.

Lewis, D. (1966). An argument for the identity theory, *Journal of Philosophy* 63 (2), 17-25.

_____. (1980). Psychophysical and theoretical identifications, rpt. in *Readings in Philosophy of Psychology* 1 (pp. 207-15).

Loewer, B. (1995). An argument for strong supervenience, in E. Savellos and Ü. Yalçin, eds., *Supervenience: New Essays* (pp.218-25). Cambridge UK: Cambridge University Press.

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

_____. (1988). Toward a homuncular theory of believing, in *Judgment and justification* (pp.3-24). New York, NY: Cambridge University Press.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms, *Philosophy of Science* 67 (1), 1-25.

Martin, C.B. (1994). Dispositions and conditionals, *The Philosophical Quarterly* 44 (174), 1-8.

Maudlin, T. (1998). Part and whole in quantum mechanics, in Elena Casttellani, ed., *Interpreting Bodies* (pp.46-60). NJ: Princeton University Press.

McLaughlin, B. (1992). The rise and fall of British emergentism, in A. Beckermann, H. Flohr, and J. Kim, eds., Energence or reduction? essays on the prospects of nonreductive physicalism (pp.49-93). Berlin: De Gruyter.

_____. (1995). Varieties of supervenience, in *Supervenience: New Essays*, (pp.16-59).

Melnyk, A. (1994). Being a physicalist: how and (more importantly) why, *Philosophical Studies* 74 (2), 221-41.

_____. (1997). How to keep the 'physical' in physicalism, *Journal of Philosophy* 94 (12), 622-37.

_____. (1999). Supercalifragilisticexpialidocious, *Nous* 33 (1):144-54.

_____. (2003). *A physicalist manifesto: thoroughly modern materialism*. Cambridge UK: Cambridge University Press.

Neher, E., and Sakmann, B. (1976). Single channel currents recorded from membrane of denervated frog muscle fibers, Nature 260 (April), 799-802.

O'Malley, M., Brigandt, I., Love A., Crawford J., Gilbert J., Knight, R., Mitchell S., and Rohwer, F. (2014). Multilevel research strategies and biological systems. *Philosophy of Science* 81 (5): 811-28.

Pagés, J. (2002). Structural universals and formal relations, *Synthese* 131 (2), 215-21.

Papineau, D. (1998). Mind the gap, *Philosophical Perspectives* 12: 373-389.

Polger, T. (2004). *Natural minds*. Cambridge MA: The MIT Press.

Quine, W. (1976). Carnap and logical truth, rpt. in *The ways of paradox and other essays*, revised edition (pp.107-32). Cambridge, MA: Harvard University Press.

Rey, G. (1997). *Contemporary philosophy of mind*. Boston MA: Blackwell.

Schiffer, S. (1987). *Remnants of meaning*. Cambridge MA: MIT Press.

Scriven, M. (1962). Explanations, predictions, and laws, *Minnesota Studies in the*

Philosophy of Science 3: 170-230.

Sergent, J., Ohta, S., MacDonald, B. (1992). Functional neuroanatomy of face and object processing, a positron emission tomography study, *Brain* 115 (1):15-36.

Shoemaker, S. (1982). Some varieties of functionalism, in J. Biro and R. Shahan, eds., *Mind, brain, and function* (pp. 93-119). Norman, OK: University of Oklahoma Press.

_____. (1998). Causal and metaphysical necessity, *Pacific Philosophical Quarterly* 79: 59-77.

_____. (2003). Realization, micro-realization, and coincidence, *Journal of Philosophy and Phenomenological Research* 67 (1), 1-23.

_____. (2007). *Physical realization*. Oxford: Oxford University Press.

Sklar, L. (1992). *Philosophy of physics*. Boulder CO: Westview Press.

Tye, M. (1995). *Ten problems of consciousness*. Cambridge MA: MIT Press.

Wigner, E. (1967). *Symmetries and reflections: scientific essays*. Bloomington IN: Indiana University Press.

Wilson, J. (1999). How superduper does a physicalist supervenience need to be? *Philosophical Quarterly* 49 (194), 33-52.

_____. (2002). Causal powers, forces, and superdupervenience, *Grazer Philosophische Studien* 63 (1), 53-78.

_____. (2005). Supervenience-based formulations of physicalism, *Nous* 39 (3), 426-59.

_____. (2006). On characterizing the physical, *Philosophical Studies* 131 (1), 61-91.

_____. (2011). Non-reductive realization and the powers-based subset strategy, *Monist* 94 (1), 121-54.

Witmer, G. (1999). Supervenience physicalism and the problem of extras, *Southern Journal of Philosophy* 37 (2), 315-31.

Yablo, S. (1992). Mental causation. *The Philosophical Review* 101 (2), 245-80.