Functional Reduction with a Third Step: a Larger and Less Reductive Picture [1]

Ronald Endicott
North Carolina State University

Forthcoming in *ProtoSociology: An International Journal of InterDisciplinary Research*, vol 39 (2023), Special Volume entitled *Physicalism "Near Enough,"* In Memory of Jaegwon Kim, editor Gerhard Preyer (http://www.protosociology.de/Volumes/Volume39.html).

1. Introduction.

In a volume dedicated to Jaegwon Kim, it is especially fitting to discuss his developed doctrine of functional reduction. It stands as the third and final pillar in Kim's overall metaphysics, placed alongside his early-period work on the nature of events and his middle-period work on concepts of supervenience. But my discussion will cover a broad spectrum of views about functional reduction, not just Kim's. Like many others, Kim operates with a notion of functional reduction that involves a single-subject theory whereby the same object *x* possesses the role and occupant properties. Also like many others, Kim's (1998) first version of functional reduction follows two familiar steps: a definition of a higher-level or special science property in terms of a functional role, then a statement describing a physical property that plays or occupies that role. But Kim (2005) subsequently adds a third step, namely, an explanation regarding how the physical property occupies the functional role.

I think Kim is correct. But how is the third step satisfied? Kim left the matter to the appropriate sciences. Yet an examination of the pertinent scientific explanations reveals that the

---

[1] I was fortunate to have studied under Jaegwon at Michigan many years ago, and he remained a kind and generous guide in the years that followed.

third step is best satisfied by a multiple-subject, part-whole explanation, which is to say, a decomposition of the occupier's causal capacities or relations. This is true even in cases wherein role and occupant properties are identical, for an occupier's causal capacities are always underwritten by a part-whole explanation. As a consequence, functional reduction is transformed into a larger picture that at bottom always contains multiple layers of distinct, nonidentical properties that divide between parts and their whole systems. I call it "Part-Based Functional Reduction." My aim is to develop this larger picture of reduction.

2. The Original Account of Functional Reduction.

To set the stage, functional reduction to be a form of explanation that reveals how a portion of ontology is simplified by means of a *role* and the *occupation of that role*. Yet philosophers also understand roles and occupation differently. The roles might be formal or causal or social in nature (Polger 2004), and the occupation might be a correspondence, or a semantic relation of satisfaction, or a metaphysical relation of standing in a set of causal relations (Endicott 2005, 2012). I will confine the discussion to theories with causal roles and a metaphysical relation of occupation. Thus understood, functional reduction involves two main steps first outlined by David Lewis (1966). Joseph Levine expresses them as follows: "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 a causal role by which we are seeking the underlying mechanisms. Stage 2 involves the empirical work of discovering just what those underlying mechanisms are" (Levine 1993, 132). More generally:

> [FR] A property *M*, that belongs to an object *x* and described by a theory T_1, is functionally reduced if and only if (1) *M* is possessed by an object *x* and defined functionally in terms of a causal role *R* described by a non-basic theory T_1; and (2) *P* is possessed by the same object *x* but described by a more basic theory T_2 and discovered to occupy the causal role *R*.

FR belongs to a single-subject theory whereby the same object *x* possesses role-defined and occupant properties. This reflects the functionalist literature in the philosophy of mind with few exceptions, save the occasional discussion of coincident objects like a statue and its clay (e.g., Shoemaker 2003). But explanations involving the same *x*, or even coincident objects *x* and *y*, differ from the explanations by smaller non-coincident parts that I will discuss later.

To further clarify, I do not require that the definition mentioned at (1) is known *a priori*, thus remaining neutral on analytic versus scientific functionalism. I also permit the theory T_1 that specifies the role for *M* to be either a folk theory or a scientific theory. But theory T_2 about *P* mentioned at step (2) is scientific in nature, befitting the fact that (2) represents an empirical discovery. In addition, although I focus upon the properties *M* and *P*, one may speak of the objects and events that possess them, for example, saying that a particular event occupies a causal role by possessing the occupying property and standing in the causal relations (see Endicott 2010 for translating from properties to instances with theories of realization).

As well, explanations with (1) and (2) involve unstated auxiliary assumptions about more general explanatory frameworks that address whether the functional reduction is cast in terms of non-representational ontic facts, or texts, or pragmatic acts of communication, or processes of understanding, or deductions involving them, and so on. Most functional-role theorists in the

philosophy of mind leave this matter of frameworks open, but I will return to the issue when relevant to the nature of the advertised third step. Moreover, whereas (1) and (2) constitute the *explanans*, a functional reduction may involve a range of *explananda* – not just the fact that an object has *M*, but a correlation between *M* and *P*, or even an identity *M* = *P* according to some versions of FR.

That being said, the generic idea of FR is familiar: why does an object *x* have *M*? Because it has the occupier *P* which meets the causal condition that is definitive of *M*. To illustrate, let *M* represent having an active theory of mind module or TOMM that enables one to think about the thoughts of others. A cognitive theory T_1 describes *M*'s causal relations, starting with perceptual processes such as seeing the direction of another person's eyes, which then feed into systems for joint attention and empathy, which then feed into TOMM, whose judgements are then sent to various downstream systems of language and behavior (Baron-Cohen 2005). So consider FR with a standard second-order definition at (1), where the bold letter "**P**" is a property variable and the stated causal relations are represented by collective types "*C*" and "*E*" linked together by a causal or counterfactual conditional "=>." The functional definition is, roughly: *x* has *M* = df. *x* has a second-order functional property of having some physical property **P** such that (*x* has *C* => *x* has **P** => *x* has *E*), and *x* has **P**.

Now suppose the scientific evidence suggests that *M*, or TOMM, correlates with high activity in the right temporal parietal junction (Shurz and Perner 2015). Let that be *P*, which is described in the vocabulary and concepts of neurobiology T_2. So next there is a statement of role occupation whereby *P* is caused by *C* and causes *E*, pace (2): for any *x*, if *x* has *P*, then (*x* has *C* => *x* has *P* => *x* has *E*), plus the proposition that some actual object has *P* (or *P* is instantiated). Taken together, (1) and (2) result in the desired explanation that *x* has TOMM (see a full

derivation in fn. 5). That is, *x* has TOMM because it has the occupying activity in the right temporal parietal junction which meets the causal condition that is definitive of TOMM.

Let me also address the two basic kinds that fall under the generic FR. There is functional reduction$_1$, or *reduction to the functional*, which is secured by a specific interpretation of (1) and then conjoined with (2) in a way that accepts determination by the physical occupier but denies the identity between the functional and the physical property. For example, on a standard interpretation of the above second-order definition, *M* is the second-order functional property of having some *other* property that stands in the place between *C* and *E*. So $M \neq P$. However, even though functional reduction$_1$ lacks a function-to-physical identity, it has a special-science-to-function or mental-to-function reduction by the definition at step (1). That is one way to simplify the ontology. [2] Next there is functional reduction$_2$, or *reduction to the physical*, which is secured by a different interpretation of (1) and then conjoined with (2) in order to draw the conclusion that the functional property is identical with the physical occupier property. For example, on a standard first-order view, the functional *M* is defined as the one and only property associated with the place between *C* and *E*, from which it follows that $M = P$. [3] Each is a reduction, for each simplifies ontology, although functional reduction$_2$ provides an additional simplification.

---

[2] It is not unusual for philosophers to classify a position as a form of reduction simply by virtue of the functional definition in (1). As Kim states: "When the first step has been carried out and the property targeted for reduction has been functionalized, in an important sense the property has been shown to be 'reducible'" (Kim 2005, 164).

[3] Raphael Van Riel (2014) also describes functional reduction in a way that encompasses what I call reduction$_1$ and reduction$_2$.

Parenthetically, the foregoing distinction is not quite the same as Ned Block's (1980) well-known division between a *second-order, functional state identity* theory versus a *first-order functional specifier* theory. Specifically, functional reduction$_1$ ≠ a second-order, functional state identity theory, and functional reduction$_2$ ≠ a first-order specifier theory. Sydney Shoemaker (2007) promotes functional reduction$_1$, but he employs a first-order, functional specifier scheme (see McLaughlin 2007 for a discussion of Shoemaker on this point). Kim promotes functional reduction$_2$, but he employs the second-order language of a functional state identity theory, albeit with a nonstandard, deflationary interpretation according to which "the property of having property *Q*" picks out the same thing as "the property *Q*," thus enabling him to claim that *M* = *P* (Kim 1998, 99). In any case, my argument will apply to both functional reduction$_1$ and functional reduction$_2$ cases, for my thesis does not turn on differences in functional definitions at step (1). My thesis concerns (2), the fact of role occupation, and how it should be explained.

3. The Third Step.

In a later work Kim mentions the two steps of functional reduction, but he adds a third. Here is the full passage:

> To reduce a property, say being a gene, we must first "functionalize" it; that is, we must define, or redefine, it in terms of the causal task to perform. Thus, being a gene may be defined as being a mechanism that encodes and transmits genetic information. That is the first step. Next, we must find the "realizers" of the functionally defined property – that is, properties in the reduction base domain that perform the specified causal task. It turns out that DNA molecules are the

> mechanisms that perform the task of coding and transmitting genetic information – at least in terrestrial organisms. Third, we must have an explanatory theory that explains just how the realizers of the property being reduced manage to perform the causal task. In the case of the gene and DNA molecules, presumably molecular biology is in charge of providing the desired explanations" (Kim 2005, 101).

Curiously, Kim introduces the third step with little fanfare, and he does not spend much time discussing it. Rather, after a schematic summary, [4] he immediately moves to other issues, for example, about a general constraint on reductive explanation that it should not to refer to the target phenomenon in its premises (Kim 2005, 103-108), and whether functional definitions involve conceptual connections (Kim 2005, 108-112).

Perhaps the lack of discussion is understandable, given that a few philosophers had already mentioned something like the third step which Kim describes. For example, David Papineau accepted a version of functional-role theory, yet he observed that there are also "role-filling explanations" even if the functional property is identical to the occupier property. Thus,

---

[4] Kim summarizes: "Step 1 [Functionalization of the Target Property] … Having M = $_{\text{def.}}$ having some property or other P (in the reduction base domain) such that P performs causal task C. Step 2 [Identification of the Realizers of M] Find the properties (or mechanisms) in the reduction base that perform the causal task C. Step 3 [Developing an Explanatory Theory] Construct a theory that explains how the realizers of M perform task C" (Kim 2005, 101-102). Kim only briefly returns to the third step later, restating the basic idea (Kim 2005, 164).

where water is the function and H_2O is the occupier, Papineau said: "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" (Papineau 1998, 380; for another brief mention of an explanation for role occupation, see Michael Tye 1995, 47). I will return to Papineau's example shortly.

Still, more discussion would have been desirable, especially if I am correct that the third step leads to a quite different picture of functional reduction. Indeed, the proposed third step raises several issues. Let me address one preliminary issue now, and then turn to exactly how one should understand these explanations for role occupation. Thus, one might have doubts that an explanation for role occupation should count as a part of the functional reduction itself rather than stand as an independent explanation on grounds that steps (1) and (2) provide good explanations by themselves. Consider one of Kim's illustrations:

> (i) To be in pain, by definition, is to be in a state which is caused by tissue damage and it in turn causes winces and groans.
>
> (ii) C-fiber stimulation (in humans) is caused by tissue damage and which in turn causes winces and groans.
>
> (iii) Jones' C-fibers are stimulated at *t*.

(iv) Therefore, Jones in pain at *t* (Kim 2005, 28, with numbering added and the order of the premises changed). [5]

One could maintain that this sample functional reduction counts as a good explanation for why an object or system like Jones is in pain even though it lacks any explanation for the role occupation stated at (ii)/(2) precisely because the premises provide a reason why an object has *M*. The proposed third step then provides something else, namely, *a reason for that reason* mentioned at (ii)/(2) (Bradford Skow 2016 calls this a "second-level reason").

Even so, I think Kim is correct that a third step is needed, whether one classifies it with FR or counts it as a separate explanation. One reason is that, without the third step, one would not have a *physically acceptable* functional reduction. A commitment to a minimal kind of physicalism requires that there are no brute facts in the special or non-basic sciences, which is to say that non-basic facts should not be emergent but explainable in a physically acceptable way (see McLaughlin 1992; Horgan 1993). Yet the above inferences are consistent with the fact

---

[5] Dropping Kim's variable for times, the deduction with supplementary premises is this:

(i)	$(\forall x)\ [Mx <=> \exists \mathbf{P}\ (((Cx => \mathbf{P}x)\ \&\ (\mathbf{P}x => Ex))\ \&\ \mathbf{P}x)]$	[assum]
(ii)	$(\forall x)\ [Px => ((Cx => Px)\ \&\ (Px => Ex))]$	[assum]
(s_a)	$Fa <=> \exists \mathbf{P}\ (((Ca => \mathbf{P}a)\ \&\ (\mathbf{P}a => Ea))\ \&\ \mathbf{P}x)$	[(i), UI]
(s_b)	$Pa => ((Ca => Pa)\ \&\ (Pa => Ea))$	[(ii), UI]
(iii)	Pa	[assum]
(s_c)	$(Ca => Pa)\ \&\ (Pa => Ea)$	[(s_b), (iii), MP]
(s_d)	$Pa\ \&\ ((Ca => Pa)\ \&\ (Pa => Ea))$	[(iii), (s_c), conj]
(s_e)	$((Ca => Pa)\ \&\ (Pa => Ea))\ \&\ Pa$	[(s_d), assoc]
(s_f)	$\exists \mathbf{P}\ (((Ca => \mathbf{P}a)\ \&\ (\mathbf{P}a => Ea))\ \&\ \mathbf{P}a)$	[(s_e), EG]
(iv)	Ma	[(s_a), (s_f), equiv].

expressed by (2) being physically unacceptable, say, a non-basic but unexplainable fact regarding how or why the occupier *P* stands in the pertinent causal relations (imagine someone claiming that it is a brute fact that DNA transmits genetic information, or a brute fact that H_2O appears colorless). As I stated elsewhere (Endicott 2016a), just as supervenience requires supplementation by a theory of superdupervenience that explains what are otherwise brute inter-level laws, functional-role theory requires supplementation by a theory of superduperfunctionalism that explains what are otherwise brute facts of role occupation.

Another reason for a third step is that *explanations for role occupation are part of the empirical discovery that allows one to justifiably assert (2).* For example, the discovery that H_2O occupies the role of water involved numerous discoveries, including why it is colorless in a glass, or how it evaporates when heated, or the process by which it expands when frozen. I will address the precise nature of these explanations in the next section. So I assume the third step is important. But my main concern is about how one should understand the account of *reduction* that results from the added third step. To that topic I now turn.

4. Part-Whole Explanation & Part-Based Functional Reduction.

Exactly how does an explanation for role occupation proceed? Let me start with some examples. Recall Papineau's case. Water is the functional and H_2O is the physical. But he said that there is an explanation in principle from chemistry for why H_2O is "colourless, odourless and tasteless" (Papineau 1998, 380). This is an explanation regarding why it occupies the water role. Unfortunately, the cited features raise complicated issues about whether one needs to invoke external relations (water is clear in a glass but blue in the ocean) or even perceptual faculties (the

minds for whom water has no color, odor, or taste). So consider a more straightforward aspect of water's causal role.

Water expands when frozen. It is not a basic fact, so there should be an explanation regarding how or why freezing a body of H_2O causes it to expand. [6] Moreover, the explanation is about the hydrogen bond network studied in the special field of soft condensed matter within physical chemistry. Very roughly, at temperatures above 0°C there is more thermal energy to break the bonds between the hydrogen atoms in an H_2O molecule and shake them out of position, which partially collapses the structure. But at 0°C the molecule are completely bonded in a V-shape with an open space between the two hydrogen atoms because there is less thermal energy to disrupt the bonds. Consequently, there is more open space between the hydrogen atoms in its frozen solid state, in contrast to its liquid state, which creates the expansion for the entire body of H_2O. [7]

---

[6] Notice the "how or why." Kim describes the third step in terms of a how-question, e.g.: "we must have an explanatory theory that explains just how the realizers of the property being reduced manage to perform the causal task" (Kim 2005, 101). But Papineau employed a why-question: "an explanation why H_2O is colourless, odourless and tasteless" (Papineau 1998, 380). That is acceptable linguistically, since the use of "how" and "why" overlap in many cases (for some examples, see Richardson 2018). But how-questions are especially appropriate for part-whole explanations, for typically they indicate the means, or manner, or method by which a system is able to operate by the processes of its parts.

[7] Physical chemists Iwao Ohmine and Hideki Tanaka described the process in the other direction, from a solid to a liquid state: "At low temperature, water exists in one of the ice crystalline

This is a familiar *part-whole explanation*, otherwise known as *a decomposition of a causal capacity* (Cummins 1975; Craver 2001). Facts about individual molecules and atoms y_i … y_n and their properties Q_i … Q_n such as chemical bonds and behaviors at different temperatures, explain the pertinent system property regarding a causal capacity or relation for the whole body x of water/H_2O – why it expands when frozen (for present purposes one may take "causal capacity" and "causal relation" interchangeably). Of course there are many different kinds of part-whole explanations. Some involve simple aggregates, others involve more complicated mechanisms, and others involve still more complicated computational systems (Cummins 1983; Bechtel and Richardson 1993; Endicott 2016b). But part-whole explanations are widespread in the sciences.

Return to the case of having an active theory of mind module *M* and having high activity in the right temporal parietal junction *P.* Pace the third step, one should explain how the neural *P* occupies the role of TOMM, say, receiving signals from the joint attention mechanism and then sending signals to language and motor control. This capacity for signal transmission is embodied in the brain as *neurotransmission*, and its explanation involves several smaller processes. Roughly, neurotransmitter molecules bind to ligand-gated receptors in the dendrites, which causes them to open and allow sodium (Na+) ions to rush in the cell body and change its polarity

---

structures. When the temperature increases to the melting point (0°C), water absorbs the latent heat of about 80 cal/g, or 1.44 kcal/mol, hydrogen bonds are partially broken and the system becomes 'frustrated.' The system then undergoes facile hydrogen-bond network rearrangement involving collective motions of many water molecules accompanied by large energy fluctuations" (Ohmine and Tanaka 1993, 2545-2546).

from negative to positive. Then nearby voltage-gated ion channels open, allowing more sodium ions to enter, which causes a cascade that runs down to the axon terminal where vesicles filled with neurotransmitters are released, which triggers the same process all over again at the next neuron (see Doyle, et. al. 1998; framed in terms of a specific account of part-whole mechanistic explanation, see Machamer, Darden, and Craver 2000).

So the explanation turns on facts about individual parts $y_i \ldots y_n$ and their properties $Q_i \ldots Q_n$ and the processes they undergo, which through an orderly sequence bring about the transmission of neuro-chemical signals for a system of neurons x. Of course there may be competing explanations. I will consider a few in the next section. But part-whole explanations are not like the explanations provided by a single-subject, functional-role theory (Gillett 2002; Endicott 2016a). They are polar opposites. [8]

The single-subject theory posits a large physical structure for a complex structural property in step (2), like having activity in the right temporal parietal junction, so that it occupies the same place within a pattern of causes and effects as the system described by a higher-level cognitive theory at step (1), like having a theory of mind module. But the multiple-subject theory decomposes the large structure and its system property into smaller parts and their properties, such as having neurons, being an ion channel, and possessing positive and negative charges in

[8] The difference is obscured by the fact that "functionalism" stands for quite different things in philosophy, both a single-subject, functional-role theory (Lewis, Kim, Stalnaker) and a Cummins-style, part-whole functional analysis (Cummins 1975; Craver 2001) that is tied to the analytic method of decomposing something into its parts in order to understand it.

order to explain how the activity in the right temporal parietal junction is caused by inputs from other systems and in turn causes activity in other systems.

Moreover, the single-subject theory excludes properties from being occupiers that do not stand in the same causal relations as a role-defined property, including the individual part properties mentioned in the sample explanations. Thus, having an ion channel does not stand in the same causal relations as the system property concerning activity in the right temporal parietal junction. The property of being a hydrogen bond does not stand in the same causal relations as the system property of being body of H_2O. But these lesser properties help to explain the relevant occupational fact in a part-whole fashion. Granted, one may take all of the parts and their properties collectively as a large object with a complex conjunction of all those part properties. But that is an occupier, not its decomposition. So the picture that begins to emerge is that whereas steps (1) and (2) involve a single-subject theory, step (3) involves a part-whole theory.

Furthermore, the part-whole explanation regarding H_2O is relevant to a case of functional reduction$_2$, meaning that a property identity between the functional (water) and the physical (H_2O) does not remove the need for a part-whole explanation of the physical occupier's causal capacities. As well, there are part-whole explanations for role occupation in functional reduction$_1$ cases. Perhaps the case of TOMM and its occupation by the right temporal parietal junction is a good example, if the former can be realized by alternate physical objects and properties. Thus there are two kinds of expanded functional reduction:

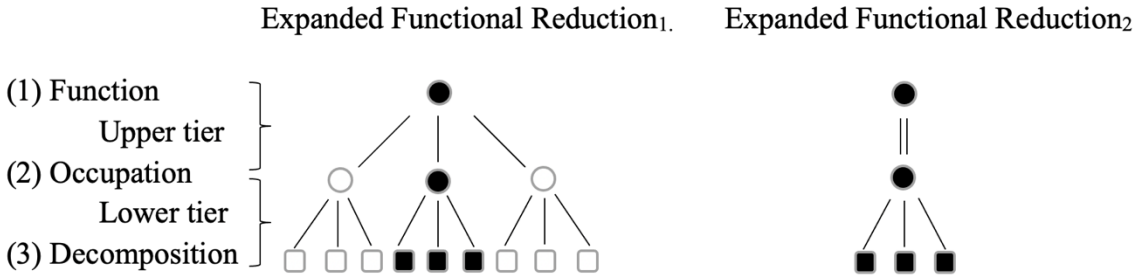

Diagram A: L, expanded functional reduction$_1$ with alternative occupiers plus underlying part properties; R, expanded functional reduction$_2$ with an identity between function and occupier plus underlying part properties.

The function and occupying properties possessed by the same *x* are represented in Diagram A by circles, whereas part properties possessed by *x*'s parts are represented by squares. The alternate occupiers of multiple realization in reduction$_1$ cases are represented by unfilled dots and squares. But both reduction$_1$ and reduction$_2$ cases share an analogous structure in each case, here represented by the filled in figures, which I call "the Pearl Tower" (named after the remarkably similar shaped Oriental Pearl Radio and Television Tower in Shanghai).

Now I believe that both types of reduction mirror real facts in the world. All is not like water, Thales not withstanding. But the important point is that the same moral holds true for both reductive$_1$ and reductive$_2$ cases: the explanation of role occupation results in a larger picture of functional reduction wherein the roles and occupants at steps (1) and (2) are joined with a deeper level of part-whole relations at (3). I call it "Part-Based, Functional Reduction," or PFR for short. I define the general idea as follows:

> (PFR) Property *M*, that belongs to an object *x* and described by a theory T_1 is part-based, functionally reduced if and only if (1) *M* is possessed by an object *x* and defined in terms of a causal role *R* described by a non-basic theory T_1; (2) *P* is possessed by the same *x*, described by a more basic theory T_2, and discovered to occupy the causal role *R*; and (3) *P's* occupying role *R* is explained by means of *x*'s parts $y_i \ldots y_n$ and their properties $Q_i \ldots Q_n$ as described by either T_2 or still more basic theories T_n.

PFR presents a three-step *explanans*, which may again have different *explananda* (why *x* has *M*, why there is a correlation between *M* and *P*, why *M* = *P*). Also, PFR presents reduction as a *three-place relation* between properties and objects of theories: *Mx* of T_1, *Px* of T_2, and $Q_iy_i \ldots Q_ny_n$ of $T_{2\text{-}n}$. Accordingly, one may view its notion of explanation as a three-place relation: *M*, that is possessed by *x* and described by T_1, is part-based, reductively explained by occupier *P*, possessed by *x* and described by T_2 in virtue of *x*'s parts $y_i \ldots y_n$ and their properties $Q_i \ldots Q_n$ described by $T_{2\text{-}n}$. Moreover, functional reduction is *full* if it applies to all of the properties in the domain of T_1, and partial otherwise. Kim (2005) thus promotes a partial reduction of mind because of his qualms about qualia. Furthermore, "explained" in the final clause may take on different meanings, depending upon the kind of part-whole explanation appropriate for the objects in question (aggregate explanation, mechanistic explanation, etc.).

As well, PFR displays an interplay between two notions of levels (for a good discussion of levels, see Craver 2014). On the one hand, there are *part-whole or mechanistic levels*. Accordingly, object *x* is a whole with a deeper "level" of parts that creates a division between (2) and (3). On the other hand, there is also a notion of *levels via descriptions or representations by*

less basic versus more basic scientific theories. Accordingly, there is a different division between (1) and (2), and probably again within (3). For example, *Mx* is described in a cognitive theory for TOMM T_1; *Px* is described by the resources of a more basic neurobiological theory T_2 for the right temporal parietal junction; and the parts Q_iy_i … Q_ny_n are described by neurobiological and physical theories $T_{2\text{-}n}$ (a full part-whole explanation often draws upon several specialized areas in the sciences dedicated to several levels of parts; cf. Machamer, Darden, and Craver 2000, 13, regarding "nested hierarchies" for mechanisms; and notice the reference to quantum properties for the hydrogen bonds that explain various behaviors of frozen water in Huang, et. al. 2015).

As such, PFR is a special kind of part-whole explanation, namely, one whose *explanandum* is the occupation of a functional role associated with *M* as described by T_1. PFR thus explicitly encompasses a larger portion of the sciences, or the theoretical vocabularies and concepts employed within those sciences, since a standard part-whole theory need only specify its system *Px* in its relation to its parts Q_iy_i … Q_ny_n specified by theories $T_{2\text{-}n}$. For example, a standard part-whole explanation for neurotransmission may be framed entirely within the vocabulary of concepts of neurobiology without mentioning any terms of a higher-level scientific theory – nothing about a cognitive theory for TOMM, and nothing about a more abstract theory of data transmission in computer science for signals between senders and receivers, even though the higher-level-theory items may map onto the electro-chemical signals transmitted between neurons.

Finally, the notion of reduction supplied by PFR is *mixed* and *modest*. It is mixed because, while there is always a property reduction or identity between non-basic and functional properties at (1), there is only a property reduction or identity with physical properties from (1)

and (2) and the necessary auxiliary premises in reduction$_2$ cases. Yet PFR reduction is always *modest*, even in reduction$_2$ cases, for there is always a part-whole reduction at (3), not a property reduction or identity. That is the less reductive part of PFR. So ends my exposition of the advertised larger but less reductive picture of functional reduction. [9]

5. Alternate Explanations and Conclusions.

I have suggested that the third step in a functional reduction is best satisfied by a part-whole explanation, which results in the expanded model of reduction PFR. I believe the initial case for the expansion is strong, for the pertinent scientific explanations for role occupation are naturally interpreted as part-whole explanations. Indeed, philosophers often interpret the realizer or occupier in functional-role theory as a "mechanism" (e.g., Levine 1993, 132; Kim 2005, 164). But mechanisms are part-whole systems, and their operations are explained by decomposition. Nevertheless, I have not addressed competing explanations. So with the space remaining let me

---

[9] I briefly suggested a similar fusion of flat and dimensioned theories of realization (Endicott 2011), and I had all the pieces in place with my suggestion that a functional-role theory would not guarantee an acceptable form of physicalism without a part-whole explanation for role occupation (Endicott 2016a). But this is the first time I have fully articulated the expanded picture of functional reduction. I add that Carl Gillett's (2002) original dimensioned theory of realization was a different kind of combination, i.e., a simple disjunction of flat and dimensioned ideas (see esp. Gillett 2002, 322; also the discussion in Endicott 2016b). In contrast, I present a fusion with each component making a contribution in every case: the single-subject theory for steps (1) and (2), the multiple-subject theory for step (3).

briefly consider the most dialectically interesting challenges, namely, possible explanations that draw upon the resources of the original single-subject FR. If they are acceptable, perhaps functional reduction does not require supplementation by a part-whole theory after all.

So the first way one might explain role occupation by the resources of FR and without recourse to a part-whole explanation is to *iterate steps (1) and (2)*. Step (3) then becomes another application of the single-subject (1) followed by (2). The rough idea is that *P* stands in the designated causal relations because of some further occupation of *P*'s role. Specifically, *P* itself is defined in terms of some property that stands in its causal relations, and property *U* stands in those relations, a property described by a still more basic theory T_u (for iterations of the role-occupant scheme, see Lycan 1987). [10] Is this a good explanation for *P*'s causal capacities, or why it is able to stand in the designated causal relations? No. The existence of a repeating metaphysical structure of roles and occupants is one thing, that it supplies a good explanation for causal capacities and causal relations is quite another (there are repeating patterns of notes printed on sheets of music, but what explains them is something beyond those patterns, i.e., the

---

[10] Thus William Lycan said: "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" (Lycan 1987, 38). Of course one might interpret the cooperating teams distributively rather than collectively. But the iterative scheme I have in mind requires the latter.

printer, or the composer's ideas and intentions, or the history of similar patterns that influenced the composer, depending upon a number of issues such as the contrast one intends to make or more generally the explanatory framework one adopts).

The basic problem is that an iterated role-occupant scheme offers the same causal capacities without ever explaining them. Consider the following analogy. Suppose one defines *M* as a machine that has a machine *P* inside that performs a causal task, say, moving its robotic arms. To explain *P*'s role occupation, then, is to explain how it move those arms, and the proposed answer is that *P* is able to move those arms because it too can be defined in terms of a role that some machine *U* occupies, and that machine *U* moves the arms. The iteration may continue. How does *U* move the arms? *U* is able to move those arms because it too can be defined in terms of a role that some machine *X* occupies, and *X* moves the arms ... But these are not good explanations, for one does not yet know how *any* of these machines move the arms, *M* or *P* or *U* or *X*. Does a machine perform the causal task by sending signals to motor control? Is it by spring loaded movements? Is it a brute fact? Instead of presenting information about how the arms move, one is presented with something like an automated Russian doll whose inner workings are never revealed. [11]

---

[11] I gave a different refutation by analogy elsewhere (Endicott 2016a). Suppose I want to stand in for my son as the guarantor for his bank loan. The bank officer will ask for proof of my ability to repay. A good answer will supply a financial analysis or breakdown of my assets versus my debts, etc. A bad answer will repeat the debtor-guarantor scheme, i.e., saying that I am myself a son whose parent will stand in for my debt, and so on for however long the loan officer could

This argument is similar to Daniel Dennett's point that functional homunculi must be "discharged" by appeal to physical mechanisms (Dennett 1978, 123-124). But it differs in at least three ways. First, I do not assume that properties defined by causal roles must be tied to systems that are treated like intentional agents. Second, and more importantly, the point is that the occupying physical mechanisms themselves must be decomposed in order to understand their causal capacities, not just that they must be posited in order to discharge convenient intentional assumptions. Third, and most importantly, Dennett did not say that iterating levels of functional homunculi is a bad explanation for causal capacities. He implied only that it would be a bad explanation if its intentional assumptions were not ultimately discharged. My argument, instead, focuses upon the iteration of functional roles and occupants as an explanation of a causal capacity, and it rejects the iterations as an explanation for ~~a~~ that causal capacity.

Now the second way one might explain role occupation by the resources of FR and without recourse to a part-whole explanation is to *utilize a conclusion from (1) and (2), specifically, the identity in functional reduction*$_2$ *cases*. In fact, I believe that explanations by identities are plausible in some cases and within some contexts. For example, Chris Hill (1991) and Brian McLaughlin (2010) have argued, plausibly in my view, that the identity $M = P$ provides a good explanation for the correlation between M and P. So, extending this idea, perhaps the identity $M = P$ explains other facts about M and P, including role occupation. Consider: how is P able to stand in M's causal relations? Because P *is* M.

---

tolerate. The good answer is analogous to a part-whole explanation. The bad answer is a role-occupant iteration.

So let me just say the following. I am happy to grant that this appeal to an identity is a good explanation, but not by the pure resources of FR. Functionalists like Lewis and Kim accept the identity $M = P$. But they employ steps (1) and (2) of FR in order to explain that identity, specifically, by deriving it from those premises. [12] Therefore it would be circular to then use that conclusion to explain one of the premises, namely, the fact of role occupation at (2). Indeed, it would be preferable to dispense with FR as an explanation for the identity and treat it instead as an independent postulate. That was Hill's (1991) position, namely, that the identity theory explains things independently of functional-role theory.

Parenthetically, although FR is my focus, and not explanations by independently postulated identities, let me point out that at least some explanations by identities might well depend upon or assume a part-whole explanation in the manner shown by PFR in reduction$_2$ cases. Suppose one has adopted a pragmatic framework for explanation, meaning that a good explanation answers a question posed by an audience in conformity with pragmatic principles governing rational communication. If one asks: "Why did the bullet fail to penetrate Clark Kent?" The answer "Because Clark Kent is Superman" seems perfectly acceptable. But pragmatically apt answers are given under assumptions that are shared by the speaker and the audience, what Stalnaker (2002) calls the "common ground," and this particular answer has a part-whole explanation in that common ground – that Superman cannot be penetrated by a bullet because his body is made from different materials (if the audience believed instead that

---

[12] Specifically, from premises (1), (2), and the earlier-mentioned auxiliary assumption that "the property of having property Q" picks out the same thing as "the property Q," Kim deduces the conclusion $M = P$.

"Superman" was a normal human with a normal human constitution – the "super" of their building – then saying that the bullet failed to penetrate Clark Kent "because Clark Kent is that super man" would *not* be a good explanation).

Indeed, the example shows that at least some explanations by identities rest upon part-whole explanations about the constitution of a body in a way that resembles how explanations by the two steps of FR rest upon part-whole explanations. So, returning to the previous example, water = H2O. But that inter-level identity only forestalls the inevitable decomposition when the relata are non-basic and the basic objects are small. Generally speaking, and strange non-local quantum behavior aside, every inter-level identity involving a composite ultimately requires a decomposition into smaller parts and lesser properties at a more basic level of reality, as illustrated by The Pearl Tower structure of a part-whole functional reduction.

So the upshot is that the resources of FR alone will not suffice for an explanation of role occupation. Steps (1) and (2) must be supplemented with some other explanatory resource that satisfies step (3) – hence the expanded, part-based account PFR. Of course further work remains to be done. For example, I have not compared PFR to other accounts of reduction in the literature, only a standard two-step functional reduction and an unattached part-whole explanation. Even so, the synthesis of functional reduction and part-whole explanation yields a promising and comparatively modest account of reduction. It holds true for versions of functional reduction$_1$ with no function-to-physical identities, and it holds true for versions of functional reduction$_2$ with function-to-physical identities. But in every case the occupation of a causal role is explained by layers of distinct, nonidentical properties that divide between parts and their whole systems. Any hope for a universal scheme of property reduction gives way to a part-whole reduction. The world is painted with same and different.

References

Baron-Cohen, S. 2005. “The Empathizing System: A Revision of the 1994 Model of the Mindreading System, in *Origins of the Social Mind: Evolutionary Psychology and Child Development*, edited by B. J. Ellis and D. F. Bjorklund, 468–92. Guilford Press.

Bechtel, W. and Richardson, R. 1993. *Discovering Complexity*. Cambridge MA: MIT Press.

Block, N. 1980. “What is Functionalism?,” in *Readings in Philosophy of Psychology* 1, edited by N. Block, 171-84. Cambridge MA: Harvard University Press.

Craver, C. 2001. “Role Functions, Mechanisms, and Hierarchy,” *Philosophy of Science* 68: 53-74.

____. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York, NY: Oxford University Press.

____. 2014. “Levels,” in *Open MIND* 8, edited by T. Metzinger and J. M. Windt, Frankfurt and Main: MIND Group. https://open-mind.net/papers/levels.

Cummins, R. 1975. “Functional Analysis,” *Journal of Philosophy* 72: 741-65.

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

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 no. 69: 69-77.

Endicott, R. 2005. “Multiple Realizability,” in *The Encyclopedia of Philosophy*, vol.6, 2nd Edition, edited by D. Borchert, 427-32. US: Thomson Gale, Macmillan Reference.

_____. 2010. “Realization, Reductios, and Category Inclusion,” *Journal of Philosophy*, vol. 107,

no. 4: 213-19.

____. 2011. “Flat versus Dimensioned: The What and How of Functional Realization,” *Journal of Philosophical Research* 36: 191-208.

____. 2012 “Resolving Arguments by Different Conceptual Traditions of Realization,” *Philosophical Studies*, vol. 159, no.1: 41-59.

_____. 2016a. “Functionalism, Superduperfunctionalism, and Physicalism: Lessons from Supervenience,” *Synthese*, vol. 193 no.7: 2205-35.

____. 2016b. “Developing the Explanatory Dimensions of Part-Whole Realization,” *Philosophical Studies*, vol.173 no.12: 3347-68.

Gillett, C. 2002. “The Dimensions of Realization: A Critique of the Standard View,” *Analysis* 62 no.4: 316-22.

Hill, C. 1991. *Sensations: A Defense of Type Physicalism*. Cambridge UK: Cambridge University Press.

Horgan, T. 1993. “From Supervenience to Superdupervenience: Meeting the Demands of a Material World,” *Mind* 102 no. 408: 555-86.

Huang, Y., Zhang, X., Ma, Z., Zhou, Y., Zheng, W., Zhou, J., Sun, C. 2015. “Hydrogen-Bond Relaxation Dynamics: Resolving Mysteries of Water Ice,” *Coordination Chemistry Reviews* 285: 109-65.

Kim, J. 1998. *Mind in a Physical World*. Cambridge MA: MIT Press.

_____. 2005. *Physicalism, or Something Near Enough*. Princeton NJ: Princeton University Press.

Levine, J. 1993. “On Leaving Out What It’s Like,” in *Consciousness: Psychological and Philosophical Essays,* edited by M. Davies and G. Humphreys, 121-36. Oxford: Blackwell.

Lewis, D. 1966. “An Argument for the Identity Theory,” *Journal of Philosophy* 63 no. 2: 17-25.

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

Machamer, P., Darden, L., and Craver, C. 2000. "Thinking about Mechanisms," *Philosophy of Science* 67: 1–25.

McLaughlin, B. 1992. "The Rise and Fall of British Emergentism," in *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, edited by in A. Beckermann, H. Flohr, and J. Kim, 49-93. Berlin: De Gruyter.

____. 2007. "Mental Causation and Shoemaker-Realization," *Erkenntnis* 67: 149-72.

____. 2010. "Consciousness, Type Physicalism, and Inference to the Best Explanation," *Philosophical Issues* 20: 266-304.

Ohmine, I. and Tanaka, H. 1993. "Fluctuation, Relaxations, and Hydration in Liquid Water: Hydrogen-Bond Rearrangement Dynamics," *Chemistry Review* 93: 2545-66.

Papineau, D. 1998. "Mind the Gap," *Philosophical Perspectives* 12: 373-89.

Richardson, K. 2020. "Grounding Pluralism: Why and How," *Erkenntnis* 85 no.6: 1399–1415.

Skow, B. 2016. *Reasons Why*. UK: Oxford University Press.

Shoemaker, S. 2003. "Realization, Micro-Realization, and Coincidence," *Journal of Philosophy and Phenomenological Research* 67 no.1: 1-23.

_____. 2007. *Physical Realization*. Oxford: Oxford University Press.

Shurz, M., and Perner, J. 2015. "An Evaluation of Neurocognitive Models of Theory of Mind," *Frontiers in Psychology* 6, article 1610.

Stalnaker, R. 2002. "Common Ground," *Linguistics and Philosophy* 25: 701–21.

Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge MA: MIT Press.

Van Riel, R. 2014. *The Concept of Reduction*. Springer International Publishing Switzerland.