

# Non-Linear Supervenience and the Mind

© 2005 by Damon Woolsey

## Introduction

There are two seemingly different domains in the universe: physical and functional. The physical domain is entirely described by physics, while the functional domain is somewhat abstract. And although it may seem that everything can, in principle, be described by physics, a great deal seems to be left out. That is, physics does not seem capable of describing the functional domain in a satisfying way. If this is so, then the functional domain must have some level of autonomy. But here we should seek simplicity in our explanations, for what we are dealing with is the very fabric of reality itself; and if reality is to be explained, it would be easier to understand why there is something rather than nothing if we could explain it simpliciter. This has much to do with the various sciences, and simplicity demands that all sciences be unified. It seems that the universe *should* operate on some fundamental level, and thus all the sciences should turn out to study different aspects of the same thing. One approach to this would be the reduction of all sciences to physics, while another involves the use of bridging laws. Still other views deny these, maintaining that some characteristics about our world supervene on, yet are non-reducible to, physics. I believe that none of these is correct, and here I describe yet another alternative. I believe that the universe does operate on some one fundamental level, and that the key to unlocking this secret is to be found within our own minds. It is the ultimate irony that the key to understanding the fundamental nature of the universe lies *within* us, yet we have no direct (i.e. introspective) access to it.

Physics cannot fully characterize the world as we experience it, but let us consider the general sentiment of type-physicalism as expressed by J. J. C. Smart:

There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. ...That everything should be explicable in terms of physics...except the occurrence of sensations seems to me to be frankly unbelievable (1959, 142).

To this we could add functional properties, but still arrive at the conclusion that all there is in the world is physical and can be described from an objective standpoint. Now, let us temper this claim with a quote from Thomas Nagel:

A great deal is essentially connected to a particular point of view, or type of point of view, and the attempt to give a complete account of the world in objective terms detached from these perspectives inevitably leads to false reductions or to outright denial that certain patently real phenomena exist at all.

This form of objective blindness is most conspicuous in the philosophy of mind, where one or another external theory of the mental, from physicalism to functionalism, is widely held (1986, 7).

The problem is that physics fails to explain how there can be a point of view in the world at all, while non-reductionism suffers from a flaw so serious that I regard it as untenable. What I shall argue is that most of the categories we observe in the world (including functional properties) supervene *by virtue of* our minds, and thus cannot serve as an explanation *of* our minds. Yet I shall not take minds as being fundamental constituents of reality, for they are far too complex,

and therefore must supervene on something else. However, even though physics fails to capture conscious experience in any satisfying way, I still maintain that everything should be reducible to some fundamental level. Therefore, I reject not only physicalism, but dualism as well. This necessitates something new, something better.

## I

I begin with the assumption that the physical domain is causally closed. In principle, it is possible to explain the universe in a Laplacean manner, so we should wonder why there is a functional domain at all. We could take the functional domain as being fundamental and epiphenomenal, but that is assuming too much. We could say that the functional domain emerges in its entirety from the physical domain, but it is not clear by what laws of physics that happens. Both of these views are essentially the same, and being that simplicity is my goal, I reject them outright. Yet another contender (hereafter referred to as hierarchism) involves a somewhat smooth transition from the physical to the functional, and it is this view that I shall consider at some length.

There are two types of taxonomies in the world. The first sort involves categorizing things that are physically identical (e.g. electrons), allowing of course for differences in location, time, and energy state. Collections of such things can also fall into larger categories (e.g. subatomic particles), such that certain hierarchies of categories can be characterized. These taxonomies (hereafter referred to as "physical taxonomies") are entirely described by physics, and as such form the physical domain. Another sort of taxonomy involves categorizing things by similarity of form and/or function, and it is this sort that is of interest. These latter sorts of taxonomies (hereafter referred to as "functional taxonomies") form the functional domain. Now,

being that these functional taxonomies are all, in general, composed of (or describe the operations of) physical stuff, the laws of physics still apply to them. We can refer, for example, to projectile motion, using such concepts as mass and initial velocity, although the category "projectile" could include just about anything at all. Likewise, physics can describe the operation of the device that launched the projectile, although the category "projectile launcher" could be realized in an almost infinite number of ways. My first goal, then, is to establish how things can be grouped together into categories, such that we may clarify the differences between the physical and functional types.

Physical categories consist of things considered to be fundamental (e.g. quarks), or, in the case of hierarchical concepts, things defined by sets of necessary and sufficient conditions. For example, a neutron consists of one "up" and two "down" quarks, whereas a proton is composed of two "up" and one "down" quarks. This gets a bit trickier as we move up the scientific ladder to chemistry. A hydrogen atom consists of one proton surrounded by one electron. But, there are also hydrogen ions (+ charge) and anions (- charge). A hydrogen ion is simply a single proton, while a hydrogen anion is a proton with more than one electron. This is still quite easy to understand in physical terms, so long as we keep track of the charges. Moving further up the ladder, the situation quickly becomes intractable. Biology is essentially chemistry, which is essentially physics, but the wholesale reduction of biology *to* physics simply cannot be understood by the human mind. This is not to say, however, that biology is not essentially physical; it is obvious that we are dealing with "physical" stuff. What is different here, and what grants biology its relative autonomy from physics, is the need for such a science to be tractable. This requirement necessitates the use of functional taxonomies. How much more so when we continue on from biology to psychology! This smooth transition from physics to psychology at

first seems quite satisfactory: everything is physical, and yet we can achieve a greater understanding of our world through the use of functional taxonomies. But there is a fatal flaw here that goes almost unnoticed. In fact, it is something we have so far taken for granted; yet to characterize it is to topple this house of cards that we have built.

Essentially, hierarchism takes the seemingly large gap between the physical and functional domains and breaks it down into a series of smaller steps (e.g. physics  $\rightarrow$  chemistry  $\rightarrow$  biology  $\rightarrow$  psychology), each of which remains just as inexplicable. A taxonomy either is or is not a physical taxonomy. And no matter how small a step we take, if we move from a physical taxonomy to a functional taxonomy, we are faced with an explanatory gap. If particular arrangements of physical-type categories yield functional categories, then functional categories are really just physical categories. If not, then we must accept the inexplicable emergence of additional non-physical properties, which is unacceptable. It is like saying  $2 + 2 = 5$ , or  $A + B = (A + B) + \textit{something else}$ . But if everything in the universe can, in principle, be explained by physics, then there is no "something else", unless it is some property of the individual constituents. For if  $A + B = (A + B) + \textit{something else}$ , then A and B each better contain some part of this "something else"; and if they do, then physics does not describe everything. Now it could be that this "something else" (let's call it Z) is not constituted by some attributes of A and B, but that it is brought about by the *interaction* of A and B. On this view, by considering A and B together as a unified thing (let's call it C), C has the additional property of Z. So far this explanation seems satisfactory, but this is an illusion. Since Z can only be considered as a property of some unification of physical-type constituents (e.g. A and B), then the functional categories must actually be physical categories. This would necessitate type-physicalism, but here I am assuming that the functional taxonomies are not physical taxonomies. The point here is

that if physics were capable of explaining functional taxonomies, then they would actually *be* physical taxonomies.

Now, if the functional taxonomies really are not physical taxonomies, then we are faced with an explanatory gap; for we cannot explain the functional taxonomies using functional concepts due to vicious circularity, and physical explanations have been ruled out. The conclusion is that physicalism is false, and the only place to look for an explanation for the functional taxonomies is in the minds of those who observe them. The reason for these functional taxonomies is purely for the purpose of understanding, and as such they have *no existence* outside of observation by a conscious mind. As such, using these functional taxonomies to climb the scientific ladder from physics to psychology, and then attempting to use psychology to explain how there can be minds, is entirely circular. That is, the properties of mind must either exactly match the properties of matter, or we must accept inexplicable emergent properties and parallelism/epiphenomenalism, or we must look elsewhere; I prefer the latter option. My aim is to provide a suitable alternative, so I shall now characterize how functional taxonomies supervene on the mind.

## II

One interesting aspect of functional taxonomies is that they are characterized by their function as much as by their physical form. In many cases, form follows function, such that the defining function of the category places physical constraints on what can count as such. For example, a river is not a projectile, since its physical form does not allow for instantiation of the desired function. But in other cases, function follows form, such that the ability to perform certain tasks is constrained by the physical form of the object. For example, a cannon ball cannot

flow like a river, as its physical form does not allow for the defining function. What is common in both these cases is that the categories are identified by *similarity* of form and/or function, rather than being physically identical. As such, I refer to these taxonomies jointly as "functional", as the same principle of similarity applies to both. This concept of definition by similarity presents some difficulty in understanding exactly how these taxonomies are characterized, as similarity is a somewhat vague concept. We should like to have something better, in order to fully understand the functional domain.

The problem with trying to nail down similarity is best illustrated by the Sorites Paradox (a.k.a. The Problem of the Heap). A heap can consist of many different items (e.g. grains of sand, trash, cars, etc...), such that the only condition for something being a heap seems to be "quantity of things in a particular arrangement". This results in a degree-of-fit, rather than a clear demarcation between heap/not-heap. Other functional categories have multiple conditions, each of them having the same degree-of-fit criterion. Take, for example, "ball". Balls exhibit the properties of "roundness", "bounces", and "size" (relative to ability to be carried/thrown). This seems to capture the essence of "ballness" in a very satisfying way, but note that these properties do not constitute a set of necessary and sufficient conditions. For example, footballs are only round about one axis, and bowling balls don't bounce. And while most balls are of a convenient size, we could imagine a ball being as big as the sun. This property of "ballness" is, then, to be had in varying degrees by a variety of different physical objects. Also note that each of these three conditions has its own degree-of-fit, just like the condition of something being a "heap". For example, footballs bounce, are about the right size (for throwing), and are round about an axis of symmetry, while being somewhat elliptical about the other two axes. Tires also bounce (although not very well). But while they are round about an axis of symmetry, they are flat about

the other two axes. They are also too large and heavy to throw or run with. Tires, while having some characteristics of "ballness", are not a good match. Bowling balls don't bounce very well, but they are round about all three axes, and are about the right size. The prototypical example of "ball", however, would be something like basketballs; they seem to fit all three categories perfectly. The degree to which an object matches a given category is the degree to which it exhibits the properties of that category. This degree-of-fit criterion is a suitable account of "similarity".

Due to the potential variance of individual conditions, things belonging to the same category can exhibit great variation in their properties. Interactions between objects from different categories (or even from the same category) will vary accordingly. When making predictions in the sciences, the greater the requirement for accuracy, the more specific the taxonomy must be. However, since members of the same category will have much in common, one can give *general* predictions regarding such interactions. In cases where belonging to a category is a degree-of-fit situation, one must settle for generalized predictions, or else calculate the specific details on a case-by-case basis. This is precisely the reason for what Fodor refers to as "ceteris paribus clauses" in the "special sciences" (1987, 4). This should not trouble us too much; it is the price we must pay to understand the world using our very limited mental faculties. The point here is that our understanding of the world through functional taxonomies is not perfect, because the world just doesn't work that way. Functional taxonomies are the way in which our minds *understand* the world, *not* the way the world works. We can gain an understanding of how our minds work through functional taxonomies, but that is only for purposes of understanding. As such, we cannot explain the operations of our minds by reference to functional taxonomies.

Here, I assume, the reader is thinking that I have pulled some sort of trick; that there is some sophistry in the argument thus far. It is easy to make the assumption that the functional categories in the world are objective, and stand alone as things-in-themselves. In reference to human behavior, Daniel Dennett has characterized them as "something perfectly objective" and "real pattern[s] in the world" (1991, 233). Likewise, Fodor states that, "If you want to know where my physical body will be next Thursday, mechanics—our best science of middle-sized objects after all, and reputed to be pretty good in its field—is *no use to you at all*. Far the best way to find out (usually, in practice, the *only* way to find out) is: *ask me!*" (1987, 6). Fodor and Dennett would agree that physics does in fact determine where you will be next Thursday. Furthermore, physics determines the machinations of your brain, and the expulsion of air from your lungs combined with the vibrations of your vocal cords. Physics also describes how the sound waves thus produced travel through the atmosphere to my ears, where they are translated into neural impulses that are then interpreted by my brain. But even though it is well accepted that physics can in principle describe everything, the functional categories are still thought to be an objective part of the world. The key to exposing the naivety of such an assumption is to understand that physics describes all of this; but *our understanding* of the situation is quite different. It is different, of course, *by virtue of* our minds. This is what is to be explained: how our minds accomplish this feat. To attempt to use this neat trick of our minds, to explain how we have minds, is entirely circular.

Having previously concluded that physics cannot explain the mind, and having shown that functionalist explanations of the mind are circular, we should seek out something better. We could simply take the mind as being fundamental, and thus convert to Berkeleyan idealism, but here again to do so would be to give up. The mind is far too complex to be a fundamental

characteristic of the universe; we seek something simpler. We should ask ourselves just how the mind accomplishes this feat of observing the world through functional taxonomies.

Computing the degree-of-fit of an object to a given category can be messy. A serial process can accomplish this by calculating one condition at a time, and comparing the result to a prototype of that condition. Parallel (non-linear) processes, however, can do this simultaneously. Minds are implemented by such non-linear processes, and so functional taxonomies supervene on something more fundamental by virtue of our minds. I refer to this as *non-linear supervenience*. But here again there is some difficulty; for if inclusion in a given category is a situation of degree-of-fit to a prototype, what are prototypes "made of"? We cannot reference any sort of functionalist account without vicious circularity, but we also cannot reference physics, as brute physics does not allow for intentionality. It is obvious then that intentionality and prototypes should have something in common, and that the explanation for them must be found by virtue of our minds; for functionalism is a product of our minds, and everything else is just physics. What is called for is a "here and now" *local* theory of intentionality, which I shall now attempt to provide.

### III

As stated above, we observe things in the world that exhibit similar characteristics. We group these together into categories, defined by mutual agreement through the common currency of language. We learn these definitions as children through a process of feature identification. These features tend to be easily identifiable characteristics. Take, for example, a family that has a young child and a dog. The child learns the word "dog" from his parents, and may identify this with "covered with hair" and "walks on all fours" (although not perhaps in those words, but more

in the way they *look*). If the mother brings home a cat one day, the child may point to it and say "dog". His mother would correct him, and so he may then distinguish "cat" from "dog" on the basis of "size" and "meows/barks". Suppose then that the family visits the zoo, and goes to the bear exhibit. The child will probably point to the bears and say "dog". Once again, his parents will correct him, and he will learn to distinguish "bear" from "dog" in the same way that he learned to distinguish "cat" from "dog". Through this process, the child will eventually learn the meaning of "dog" as it is defined through the mutual agreement of the society in which he lives. Such definition through mutual agreement within a society is evidenced by the fact that different societies may have different taxonomies for things such as colors, types of snow, etc... But there is more to this story than identifiable characteristics; these objects also have various semantic characteristics.

It will be worthwhile to consider how members of these categories appear to us in our phenomenological field. These objects (e.g. cats, dogs, balls, etc...) seem to have meanings contained within them. When I look at a ball, I don't *just* see something of a certain size that is round and can bounce; I see something that will entertain, that we can play a game with. In this way the object, which I perceive to belong to a particular category, also has various *semantic attachments*. Our minds associate certain categories with other categories, such that they appear to us as unified. Note that these semantic attachments are themselves categories, having their own defining conditions. Also, the conditions of categories can themselves be categories (i.e. "walks" and "legs" in "walks on four legs"). The question here is whether there is some terminus to this process of analysis, and if there is, will it provide a suitable foundation for intentionality?

I think that the only suitable candidate here would be something like analytic a priori concepts (perhaps a better description would be *atomistic noumena*, but characterizing them is an

area for future research). Let's briefly revisit the "ball" example. This category has the characteristics of "round", "bounces", and "size". The first is geometrical, and hence an instantiation of an a priori concept. The second can be analyzed as a change in direction with minimal loss of momentum upon striking another object, even when moving at low velocity. Direction, striking, velocity, etc., are a priori concepts. The third, size, is mere quantification, and hence a priori. These a priori concepts are not, in and of themselves, *sensory* concepts. Sensory modalities are just ways in which they manifest themselves. What they are is primitive logical elements, and hence they stand alone, without need for definition. These seem to be excellent candidates for primitive elements in a foundationalist theory of intentionality.

We can now apply this idea of non-linear supervenience to Searle's Chinese Room argument to achieve a novel sort of conclusion. The most obvious consideration is that most Chinese symbols (or English ones, for that matter) are not formal symbols, since they represent categories that cannot be formally defined. Furthermore, since they exhibit non-linear supervenience, their interactions cannot be detailed by a list of formal principles. The conclusion is that the Chinese Room, as Searle describes it, would not and could not pass a rigorous Turing test. Since these symbols refer to non-linear categories, they cannot be *directly* instantiated by a Turing machine. The Turing machine must implement the processes by which the a priori constituents interact. Such an implementation would pass a Turing test, provided it had learned (or been programmed with) the categories referred to by the language. As expressed earlier, being a priori, these concepts contain intentionality within themselves; *they need no further basis for their existence nor explanation of their meaning*. One point worth noting is that, on introspection, we don't *seem* to process language in this way. Rather, we seem to operate directly

on the symbols themselves, rather than breaking them down into components. I shall return to this point rather quickly.

#### IV

Intentionality is not the only thing that needs explaining here. Along with "intentional inexistence", Franz Brentano gave two other characteristics of mental phenomena that I consider just as important. The first is that they are "...the exclusive object of inner perception; they alone, therefore, are perceived with immediate evidence". This is important because, due to non-linear supervenience, the world we perceive is carved up into categories that we create. In a sense, the world *as we perceive it* is a creation of our own minds. The second characteristic is that "...mental phenomena which we perceive, in spite of all their multiplicity, always appear to us as a unity" (1995, 483). This brings us back to the point above: that we seem to operate directly on the symbols of language themselves, rather than their components. It is one thing to understand the machinations of the physical brain, and quite another to characterize the operations of our mental existence. This problem plagues all potential theories of the mind, in that there is no law of physics that allows for the apparent unity of consciousness. But then again, I am not looking to physics for an explanation. Nor am I seeking a functionalist account. The answer is to be found with the explanation of intentionality, in what I refer to here as the *a priori domain*, or (loosely) what Kant called the *noumenal realm*.

On my view, there is only one way to explain perceptions. All of one's perceptions, be they visual, auditory, tactile, kinematic, or even the awareness of one's own thoughts, constitute one's phenomenological field. Included in this field are various perceptions that have the semantic attachment of *being about oneself*. For example, I associate various sensations, and

even a portion of my visual field, as being *about* my body. I associate the *awareness* of my thoughts, insofar as I "hear" them in my head, as being *about* what I am thinking. There is also a great deal of my phenomenological field that is interpreted as being *about* "physical" things distinct from myself. This phenomenological field *is* consciousness, and "I" am nothing but a process. Indeed, the entirety of the a priori domain is a process of which you and I are small parts. But we have points of view, and types of points of view are fundamental to the universe. This is something that needs to be explained.

## V

I look toward a Spinoza-like view of *dual-aspect realizability*, which will greatly enhance our understanding of what we once thought of as merely "physical". On Spinoza's view, the one fundamental substance has infinite aspects, two of which are physical and mental. One may then wonder what happened to the infinite number of other aspects of which we are unaware; we would be right to deny them. Instead, where Spinoza takes the physical and the mental as being two ways of looking at the same substance, we can construe "looking" as looking from a *point of view*. This takes us back to the above quote from Nagel, who has rightly observed that types of points of view are fundamental characteristics of our universe. There are only two *types* of points of view: first person and third person. The third person point of view is looking at something from the *outside*, and this is how we perceive the physical world (including other minds). The first person point of view is looking at something from the *inside*, and this is how we perceive the mental. Here we have a complete set of types of points of view; there is nothing else but inside and outside, and there are only two types of aspects instead of an infinite number. This is why I refer to "dual-aspect *realizability*" rather than "dual-aspect *theory*". The fundamental

constituents of our universe are not dual aspect, but rather *appear to be* depending on one's point of view. Of course, looking at something from the inside entails *being* that thing, for no matter where you look in the physical universe, you are only looking at matter from the outside. If you attempt to break something apart to see it from the inside, you find that it is made of smaller parts that you are still viewing from the outside.

### **Conclusion**

This, in a nutshell, is my theory of mind. While it may be quite strange, I believe that it is a simple account that allows us to hold onto our intuitions. How it all works is an area for further inquiry; here I have given only a framework. One may object that such an answer is a cop-out, a plea to some mysterious "realm". My reply is simply that, whenever physicists think that they have hit upon the fundamental constituents of reality, they find something else, something deeper. And this something deeper is always much stranger than what they had before (e.g. quantum mechanics). It seems that the world is quite contrary to first appearances. To assume that we have the requisite concepts in hand to fully explain the entirety of existence is simply naïve; there is much work left to be done, and that should be good news to philosophers.

\*\*\*

## References

- Brentano, Franz, Linda L. McAlister (Translator), Anto Rancurello (Translator), D. B. Terrell (Translator); *Psychology from an Empirical Standpoint*; Routledge, 1995.
- Dennett, Daniel C.; True Believers: The Intentional Strategy and Why It Works, *The Nature of Mind*; Oxford University Press, New York, 1991.
- Fodor, Jerry; *Psychosemantics*; MIT Press, Massachusetts, 1987.
- Nagel, Thomas; *The View From Nowhere*; Oxford University Press, New York, 1986.
- Smart, J. J. C.; Sensations and Brain Processes, *The Philosophical Review*; Vol. 68, No. 2 (Apr., 1959), 141-156.

**NLS Supplement: my explicit argument against physicalism  
and in favor of dual-aspect theory**

1. The physical domain is causally closed (observation)
2. We observe functional taxonomies in the world that are not physical taxonomies (even though they both describe the same "stuff")
3. Physics cannot explain the functional taxonomies (otherwise they would be physical taxonomies)
4. Functional taxonomies cannot explain functional taxonomies (due to vicious circularity)
5. Therefore, physicalism is false (since it cannot explain functional taxonomies)
6. Functional taxonomies must supervene on something else (they are not fundamental)
7. The explanation for functional taxonomies is in the minds of those who observe them (there is nowhere else to look for their explanation)
8. Functional taxonomies cannot explain the mind (by 4 & 7)
9. The mind cannot be fundamental (it is the most complex thing in the known universe)
10. The physical and the mental must both supervene on something else (by 5, 6, 8, & 9)
11. There is only one fundamental substance (by simplicity)
12. Dual-aspect theory is true (10 & 11)