teorema
Vol. XXXVII/3, 2018, pp. 00-00
ISNN 0210-1602
[BIBLID 0210-1602 (2018) 37:3; pp. 00-00

The Discontinuity of Levels in Cognitive Science

Bryce Gessell and Felipe De Brigard

RESUMEN

Comenzamos caracterizando la concepción "funcionalista homuncular" de la mente de Dennett, tal como es descrita en su obra temprana. A continuación, la comparamos con la bosquejada en *From Bacteria to Bach and Back*. Argumentamos que los cambios recientes en la posición de Dennett han producido cierta tensión en su visión de las descomposiciones funcionales. Supuestamente, las descomposiciones funcionales basadas en la postura intencional terminan en un nivel inferior, no inteligente, que puede explicarse mecánicamente. Sin embargo, puesto que Dennett cree ahora que las neuronas podrían tener que ser descritas en términos intencionales, no queda claro si nuestras explicaciones de las funciones cognitivas van a casar con nuestras explicaciones de la conducta de neuronas y redes. Exploramos las consecuencias de esta tensión para la teoría de Dennett y para la neurociencia cognitiva en general.

PALABRAS CLAVE: *neurociencia cognitiva, funcionalismo, ontología cognitiva, postura intencional, psicología.*

ABSTRACT

We begin by characterizing Dennett's "homuncular functionalist" view of the mind, as described in his early work. We then contrast that view with the one outlined in *From Bacteria to Bach and Back*. We argue that recent changes in Dennett's view have produced tension in the way he conceives of functional decompositions. Functional decompositions based on the intentional stance are supposed to reach a bottom, "dumb" level which can be explained mechanically; however, since Dennett now believes that neurons may need to be described intentionally, it is not clear whether our explanations of cognitive functions can ever align with our explanations of neuronal and network behaviors. We explore the consequences of this tension for Dennett's view, and for cognitive neuroscience in general.

KEYWORDS: *Cognitive Neuroscience; Functionalism; Cognitive Ontology; Intentional Stance; Psychology.*

I. INTRODUCTION

Dennett's career has now spanned multiple generations — amazingly, next year will mark the *fiftieth* anniversary of his first book, *Content and Consciousness* [Dennett (1969)]. In that work and many following, Dennett

took up an enormous variety of topics, ranging from traditional questions in philosophy of mind to problems in the borderlands of other disciplines, including psychology, neuroscience, cognitive science, evolutionary biology, and more. As two of his former students, we can attest that his teaching and mentoring cover the same expansive ground.

From Bacteria to Bach and Back (hereafter, FBBB) brings together many threads of Dennett's thought, weaving a story about the origin of minds and human consciousness by way of memes permeating individuals and cultures. If that story sounds familiar, it is because Dennett has been telling a version of it since the beginning: "I have devoted half a century, my entire academic life, to the project in a dozen books and hundreds of articles tackling various pieces of the puzzle, without managing to move all that many readers from wary agnosticism to calm conviction. Undaunted, I am trying once again and going for the whole story this time" [Dennett (2017), p. 16]. But in telling "the whole story," FBBB transcends Dennett's past work and transforms his views by making them fully conversant with recent scholarship across the sciences.

In some places, however, certain claims of "the whole story" offered in FBBB differ from claims and views Dennett once held in the past. In most instances these modifications are innocuous, but at least one such alteration may be more consequential than Dennett himself seems to recognize. We are thinking about his revision of "homuncular functionalism" in chapter 8. In this brief comment, we suggest that Dennett's change of mind about homuncular functionalism brings out an internal tension between the intentional and the design stances that may even question the legitimacy of cognitive neuroscience — at least as it is currently carried out — as a reliable strategy to uncover the neural machinery of the mind.

To that end, we start — in section II — with a review of Dennett's homuncular functionalism as it was understood prior to FBBB. Next, in section III, we carefully reconstruct his re-examination of homuncular functionalism from FBBB, highlighting the differences between the original and the new formulations. Next, in section IV, we argue that the new formulation strongly suggests that the hierarchical picture resulting from a top-down functional decomposition — i.e., from people to neurons — may not match the hierarchical picture which emerges from a bottom-up decomposition, that is, from neurons to people. The consequences of this mismatch — we argue in section V — are profound, as they not only question whether the gap between the intentional and the design stances can be bridged, but more generally whether the scientific challenge of mapping mental predicates to neural categories can ever be met.

II. DENNETT'S FUNCTIONALISM

Dennett's views do not easily fit within the strict confines of philosophical labelling. Is he a realist or an anti-realist about propositional attitudes? Is he an instrumentalist or an eliminativist about the mind? Is he a realist or an anti-realist about consciousness? Since anthologies must be published, though, and chapters must fit into sections, it is not surprising that Dennett's proclivities toward computational explanations of mental processes made his early writings appear to belong alongside those of early functionalists, such as Putnam (1960); (1967), Armstrong (1968), and Lewis (1972). However, from the very beginning, there were fundamental differences between Dennett's version of functionalism, and that of what we may call *machine-table functionalists*. Putnam (1960); (1967), for example, following Turing's (1950) insight that intelligence could be understood as complex algorithmic processes, put forth the suggestion that mental states may refer to states of a Turing machine table. Related views were subsequently advocated by Armstrong (1968) and Lewis (1970), (1972). Armstrong's causal theory of the mind, for instance, argued that a mental state ought to be defined as a "state that is apt to be the cause of certain effects or apt to be the effect of certain causes" [Armstrong (1968)], independently of their material realization.

In this sense, Armstrong's causal theory is similar to Putnam's machine table functionalism, as they both define mental states in causal/computational terms independently of the material in which they are realized. Importantly, the idea that the same machine table could be instantiated by different realizers — be they neurons, chips, or vacuum tubes — motivated functionalists to postulate two kinds of mental identifications. First, there were *psychophysical* identifications: one-to-one relationships between mental states, understood as computationally (i.e., causally) defined states of a Turing machine table, and the particular realizers that bring them about. Here Lewis' insights proved critical. Roughly, his suggestion is that to identify mental states with their particular realizers you need first to "collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses, [and then] add also the platitudes to the effect that one mental state falls under another — 'toothache is a kind of pain', and the like." [Lewis (1972), p. 256]. These platitudes of folk psychology constitute the theoretical terms whose definitions are given in terms of particular causal roles. Finally, to identify the physical realizer of a particular theoretical term from the set of folk psychological platitudes, and thus establish a psychophysical identification between the

two, one needed to get the Ramsey-sentence including the relevant folk psychological term, and then identify the entity picked out by its bounded existential variable. The entity picked out by the bounded existential quantifier would be the ontological correlate of the mental state referred to by the relevant psychological term. Thus, if a term, *T*, belongs to the list of psychological platitudes, then the physical realizer of *T* would be the entity picked out by the bounded variable, *x*, from the Ramsey-sentence, $\exists!xT(x)$.

The second identification machine-table functionalism argues for are psychological identities, that is, one-to-one relations between mental states. Specifically, machine table functionalists were committed to the claim that two organisms, regardless of their physical make up, could nevertheless be in the same mental state as long as they were in the same state of an identical machine table. Just as two computers made of different materials could nevertheless be in the same computational state, two entirely different organisms could be said to be in the same state of pain or be entertaining the same belief as long as the state of their machine tables were computationally identical. Moreover, the theory seems to be committed to the claim that two entities, regardless of their physical makeup, could in principle instantiate the exact same mental life as long as both of them instantiate the exact same machine-table. Furthermore, two people would be in the same mental state if and only if they are in the same machine state. This implies that minds like ours have to have one and only one functional description. Entities with Turing machine tables that do not correspond to ours would not count as having minds like our own.

Years later, Block (1978) rightly pointed out that this view, which was supposed to reduce the liberalism of behaviorism, ends up being too liberal as well, as it wrongly attributes mentality to entities seemingly lacking it. But the alternative — to restrict machine tables to only those entities with psychologies isomorphic to ours — ends up being as chauvinistic as identity theory, for we end up denying mentality to entities which seemingly have it. However, Block's objections to machine functionalism never really affected Dennett's own computationalist view, for he does *not* think that psychological identities — i.e. the one-to-one mapping of mental states between organisms — are to be found at the level of the machine table, but at the level of the intentional stance. Thus, two entities could be in entirely different machine states, and thus require entirely different functional descriptions from the design stance, and yet share the same intentional descriptions from the intentional stance [Dennett (2005)]. For Dennett, therefore, psychological identifica-

tions do not depend on us running the exact same mental algorithm; they come from uniform intentional descriptions casted from the intentional stance [De Brigard (2007), p. 132].

But then, if two organisms can be in the exact same intentional state while differing in their underlying functional states, how can we ever map mental terms onto physical ones? Dennett's approach to answer this question finds inspiration in the way in which artificial intelligence operates:

> The AI programmer begins with an intentionally characterized problem, and thus frankly views the computer anthropomorphically: if he *solves* the problem he will say he has designed a computer that can understand questions in English. His first and highest level of design breaks the computer down into subsystems, each of which is given intentionally characterized tasks; he composes a flow chart of evaluators, rememberers, discriminators, overseers and the like. These are *homunculi* with a vengeance; the highest level design breaks the computer down into a committee or army of intelligent homunculi with purposes, information and strategies. Each homunculus in turn is analyzed into *smaller* homunculi, but, more important, into *less clever* homunculi. When the level is reached where the homunculi are no more than adders and subtractors, by the time they need only the intelligence to pick the larger of two numbers when directed to, then have been reduced to functionaries "who can be replaced by a machine." The aid to comprehension of anthropomorphizing the elements just about lapses at this point, and a mechanistic view of the proceedings becomes workable and comprehensible. [Dennett (1975/1978), pp. 80-81).

Lycan calls this strategy *homuncular functionalism* [e.g., Lycan (1981); (1987)], as it suggests that an intentional system, like us, can be hierarchically subdivided into nested sub-systems, with each lower tier requiring less intentional vocabulary for its description than the previous one, until the hierarchy bottoms out at a level of description that requires no intentional vocabulary at all: the level of neurons.

Importantly, homuncular functionalism is not only a claim about the structure of our mind: it is also a view that informs our scientific practices in cognitive psychology and neuroscience. As Dennett reminds us:

> The task of psychology is to explain human perception, learning, cognition, and so forth in terms that will ultimately unite psychological theory to physiology in one way or another, and there are two broad strategies one could adopt: a *bottom-up* strategy that starts with some basic and well-defined unit or theoretical atom for psychology, and builds these atoms into molecules and larger aggregates that can account for the complex phenomena we all observe, or a *top-down* strategy that begins with a more

> abstract decomposition of the highest levels of psychological organization, and hopes to analyze these into more and more detailed smaller systems or processes until finally one arrives at elements familiar to the biologists [Dennett (1978), p. 110].

Alongside Simon (1969) and Marr (1976), and anticipating Shallice (1988) and Cummins (2000), Dennett advocates for a top-down, functional decomposition approach that parallels the structure of homuncular functionalism. And he argues that "the bottom up strategy in psychology is unlikely to prove very fruitful," because the two "best developed attempts […] are now widely regarded as stymied" [Ibid.]. The first such attempt was stimulus-response behaviorism which, Dennett suggested, won't succeed because stimulus-response associations turn out to be the wrong kind of basic psychological atoms. The second attempt — more relevant for present purposes — is what he called back then "neuron signal physiological psychology," and he thought that "even if synapses and impulse trains are perfectly good atoms, there are just too many of them, and their interactions are too complex to study once one abandons the afferent and efferent peripheries and tries to make sense of the crucial center" [Ibid.].

In sum: the functionalist picture with which we end up, is one that differs from machine-table functionalism in fundamental respects. First, it rejects psychological identifications because the uniformity of our intentional descriptions does not depend on sharing the exact same machine table. Two intentional systems can differ in their physical and computational descriptions and nevertheless instantiate the same mental states. The second difference — which follows from the first one — is that psychophysical identifications are not going to proceed the way Lewis suggested, for even if we were able to compile all folk psychological platitudes, it is unlikely that there will be equivalent Ramsey-sentences across individuals such that one and only one physical realizer can be identified by a bounded variable. As a result, Dennett advocates for an empirical, rather than an a priori, strategy to identify the physical realizers of intentional states: a top-down, functional decomposition that parallels the structure of his homuncular functionalism.

III. Homuncular Functionalism Revised

Since the 1970's, neuroscience has developed at a staggering pace, and we now know much more about the neurobiology, neurochemistry, and functional neuroanatomy of human and non-human brains than ever

before. Unsurprisingly, Dennett has been well aware of these developments, which perhaps explains why, 40 years after the publication of *Brainstorms* and the introduction of his homuncular functionalism, he decided that it was time to revisit it in FBBB:

> I still think this is on the right track, but I have come to regret—and reject — some of the connotations of the two terms I used: "committee" and "machine". The cooperative bureaucracy suggested by the former, with its clear reporting relationships (an image enhanced by the no-nonsense flow charts of classical cognitive science models) captured the dream of top-down GOFAI [Good Old Fashioned Artificial Intelligence], but it suggested a profoundly unbiological sort of efficiency. Turing's strange inversion of reasoning is still intact: eventually in our decompositional cascade we arrive at elements whose tasks are so rigid and routinized that they "can be replaced by a machine", just like Turing's diligent human computer. The simplest moving parts *within* neurons, the motor proteins and microtubules and the like, really are motiveless automata, like the marching broomsticks in *The Sorcerer's Apprentice*, but neurons themselves, in their billions, play more enterprising and idiosyncratic roles than the obedient clerks I was imagining them to be, and that fact has major implications for the computational architecture of brains" [Dennett (2017), p. 162].

Indeed it does! Neurons are much more complex, and their behaviors are much more agential — even if unconscious —than we thought forty years ago. We now know that the simple neuronal models we inherited from McCulloch and Pitts (1943) are simply wrong, for neuronal dynamics are complex, and simple parametric models seem incapable of capturing the complexity of neuronal behavior.

Let us look at a few cases of this complexity. On the view of neurons as simple switches, these cells can say "yes" or "no" to incoming stimuli by firing or not firing, respectively. A slightly more elaborate view takes neurons as logic gates, which participate by their firing in logical operations like "AND," "OR," and "IF-THEN." The resulting view takes the nervous system as a kind of motherboard in a computer, whose operations are hardwired into the cells and their connections. The biological contingencies of the cellular environment are taken as inessential and can be ignored.

However, scientific evidence accrued in the last few decades strongly suggests that the behavior of neurons is much more nuanced than this, and the nuances may play important roles in cognition. Neurons, for example, are not only connected to other neurons synaptically, but also through structural links called gap junctions [Shimizu and Stopfer (2013)]. Signals can pass from cell to cell via these links, but importantly, the sig-

nals are not as quantifiable or as regular as action potentials. In addition, neurons can dump neurotransmitter molecules into cellular spaces in order to influence the behavior of entire circuits or networks. This neuronal behavior, called "volume transmission," pushes against the view of neurons as simple switches or gates with clear targets, which can be modeled at a very abstract level. It has also recently been discovered that neurons can share signals via the gene *Arc* and its proteins [Pastuzyn et al (2018)]. This means of communication may allow neurons which are not active to undergo changes as though they were active, by producing proteins in response to receiving mRNA from other neurons. Finally, neuronal mechanisms would likely not behave as deterministically as simple models suggested [Gessell (2017)], and neurons may multiplex, as they seem to carry signals from more than one stimulus at a time [Caruso et al. (2018)].

Needless to say, this very brief survey only scratches the surface of what recent neurobiological and neurochemical findings overwhelmingly suggest: that neurons are not simple on-and-off switches. From this new understanding of neurons, two very important consequences follow that affect Dennett's original formulation of homuncular functionalism. First, we must reconsider the actual scope of multiple realizability. When machine table functionalism was all the rave, philosophers claimed that it was conceivable to have minds implemented in all sorts of different substrata, including financial markets in Bolivia or the inhabitants of the nation of China [Block (1978)]. But based on recent developments in neurobiology and neurochemistry, Dennett has been insisting — at least since 2003 — on what he calls *minimalist functionalism*, the view that although multiple realizability as initially conceived by functionalists may be true in principle, in reality it is likely that no other materials can actually do what neurons do [see also Clark (1986); Godfrey-Smith (2016)]. That does not mean, as he reminds us in the quote above, that Turing's "strange inversion of reasoning" was wrong: it is still intact. Eventually our functional decomposition will find "elements whose tasks are so rigid and routinized that they 'can be replaced by a machine'" [Dennett (2017), p. 162]. But these elements are to be found *within* neurons, at the level of proteins, microtubules, and the like, whose behavior will be — one would hope — ultimately captured by algorithmic descriptions suitably formulated from the design stance. If so, then a second consequence for homuncular functionalism follows, namely that contrary to its initial formulation, a neuron's complex behavior may be better captured from the *intentional* rather than the *design* stance. To stress this point Dennett congenially alludes, for instance, to Fitch's notion of "nano-intentionality" and to Seung's ideas of

"selfish neurons" and "hedonistic synapses", which seem to capture the now-evident fact that neurons, as it were, have a mind of their own.

IV. Mismatching Hierarchies

Superficially, these two consequences wouldn't seem too earth-shattering for the original notion of homuncular functionalism. After all, isn't Dennett just moving the goal-posts a bit by telling us that it isn't neurons, but the stuff within neurons, that constitutes the bottom level of functional decomposition? In a sense, yes; he is certainly stressing the point that the tiered structure assumed by GOFAI architectures is overly simplistic and unnatural:

> The top-down intelligent design of the classic computer created a hyper-competent but deeply unbiological marvel. This is not because the computers are made of the wrong kind of stuff but because the stuff they are made of is organized into the wrong kind of hierarchies: the kind of planned bureaucracies that may be "well-oiled machines" but depend on regimentation of the moving parts, suppressing both exploration and improvisation at every level [Dennett (2017), p. 163].

Notice that the picture of functional decomposition offered here suggests that the hierarchy of levels may look very different when viewed from the top compared to when viewed from the bottom. In its original formulation, homuncular functionalism hypothesized a decomposition of complex cognitive processes into simpler and simpler sub-processes, each one — quite literally — *dumber* than the previous one, where "dumber" is understood as being less in need of intentional descriptions than the system above. By the time we reached the level of microscopic neurons, we expected to be at a point where no intentional descriptions were required to capture their behavior.

However, the story told in FBBB is different. Now the neuronal level is not as dumb as we thought — in fact, we may need the intentional stance to fully capture the complexity of single-neuron behaviors. If so, there is a strong possibility that the structure of the hierarchy assumed by GOFAI simply does not correspond to the ascending hierarchy of a bottom-up approach, i.e., from neurons to people. This is because our descriptions of neuronal behavior will themselves be intentional. Remember that the reason for rejecting the bottom-up approach of "neuron signal physiological psychology," back in *Brainstorms*, was that there were just too many neurons with too many complex interactions.

Nevertheless, the message we got was that in principle, if we were able to study all the neurons and their interactions, we would be able to walk our way upwards, as it were, and re-compose the functional hierarchy going from the dumb neuronal atoms characterized at the design stance up to cognitive agents described from the intentional stance.

It may be possible to interpret Dennett's revision in FBBB as simply suggesting that the right atoms might be microtubules or mitochondria, and that the hierarchical decomposition is still intact; we just got the bottom level wrong. But our concern is that in considering neurons as agents, even if Skinnerian [Dennett (2017), p. 165], we may find that the best ways to characterize their behavior don't fit with any of the subsystems into which higher levels were supposed to be decomposed from the top-down. In other words, Dennett suggested that our top-down functional decomposition was going to include recorders, discriminators, etc. — subsystems whose behaviors were clearly characterized as gears within a larger, nested mechanism. But now it turns out that if we start from the neuronal level and move up, we may find that the best characterizations of neuronal behavior do not correspond to any of the nested mechanisms that were supposed to be part of the agent as a whole.

This possible mismatch becomes more evident when we ask, with Dennett, "what could a neuron 'want'?" [Dennett (2017), p. 162]? More generally, what kind of intentional states could help us explain their complex behavior? Dennett's answer is that, just like their prokaryote cousins, neurons are designed to deal with affordances, or the things they care about. And what do neurons care about? Physiologically, neurons care about exactly those things which we would expect any one-celled organism to care about: maintenance and repair of cellular structures, the production and use of energy, and so on. But neurons also care about information. Specifically, Dennett's suggestion is that neurons care about wiring together in order to detect affordances. The "affordances" in this case offer information about the stimuli a neuron can expect to receive, and so affordances allow neurons to make hierarchical predictions. The predictions anticipate input to neurons, and the term "hierarchical" refers to the fact that predictions may be made at different levels of abstraction. Feedback loops from areas up the hierarchy influence a predictive code and modify it in response to the accuracy of its predictions.

Now consider the predictive coding approach in light of what we've learned about Dennett's homuncular functionalism. Neuronal behavior is very complex, and we have come to believe that it requires a higher level of description than the simple design stance. That is, it re-

quires the intentional stance. But if this is the case — and if neurons are computing predictions about the presence or absence of things that matter *to them* — then the possibility emerges that the right functional categories to describe the behavior of neurons, along with other bottom-up nervous structures like brain areas, networks, and systems, just may not correspond to the functional levels of description assumed by top-down decompositions [De Brigard (2012)]. These are the functional decompositions of standard cognitive science and GOFAI, which we attempt to ground in the activities of neural populations or some other functional units. But what individual neurons care about, or what they make predictions about, may not be relevant to what seems to us to matter in decompositions of cognitive functions. What the brain is doing and how it is organized from the bottom-up may look very different from what it is doing and how it is organized from the top-down.

Let's look at an example of how this might be the case. Suppose we are interested in characterizing the contribution a certain neuronal network makes to cognition, and the hierarchical predictive coding approach suggests that the network's function is to calculate expectations for the presence of a certain property *P*. On this bottom-up perspective, we have identified a functional unit — the network — and also the unit's function — predict and check for the presence of *P*. Were we to continue to identify functional units in this manner, we might identify some other network of which this smaller one is a part, with the larger network having some functional description that we also provide from the bottom-up.

The question is, why should it be the case that our top-down functional decomposition of a larger cognitive system — let's say memory — also involves property *P*? In other words, what if *P* is something that we have good reason to think our network really does care about, even though *P* doesn't appear anywhere in a detailed decomposition of our ability to remember? What if the things that matter to neural networks, or even neurons, are not the things that matter to us? Dennett's new thinking about homuncular functionalism has opened this possibility, for our intentional descriptions of neuronal behavior should be like intentional descriptions of prokaryote behavior: they only contain things that matter to the system being described. So while what matters to us will no doubt figure in functional decompositions of our cognitive systems, it may be impossible to connect the corresponding subsystems to neuronal functional units, since those units may have a functional description which is proper to them, independent of our top-down analysis.

V. Cognitive Ontologies and the Mind-Brain Mapping

The previous section ended with the problem of aligning descriptions with mechanisms from two different perspectives: one proceeding from the bottom-up, and the other proceeding from the top-down. We think that this problem may have another far-reaching consequence for Dennett's view. Like many others interested in cognitive neuroscience, Dennett reasonably assumes that a main goal of this science is to produce a mapping between personal-level cognitive processes and subpersonal level neural mechanisms. This is true when we treat cognitive systems only as information-processing systems, but it's also true when we consider the phenomenology involved. Dennett, for example, describes the importance of the mapping this way:

> We won't have a complete science of consciousness until we can align our manifest-image identifications of mental states by their contents with scientific-image identifications of the subpersonal information structures and events that are causally responsible for generating the details of the user-illusion we take ourselves to operate in [Dennett (2017), p. 367].

The assumption of the end-goal for cognitive neuroscience is clear here — what we want is an "alignment" of mental states, which are individuated by their contents, with the neural structures and events that are causally involved in producing mentality. We can call the collection of such mental states and processes our "cognitive ontology" [Anderson (2014)]. The purpose of cognitive neuroscience, then, is to produce the mapping which moves from elements in our cognitive ontology to neural structures, which are themselves described by the design and physical stances.

The line of argument from our previous section has an important consequence here. Cognitive neuroscience, as initially conceived, has been modeled after the GOFAI and top-down functional-decomposition approaches to cognition. That is, cognitive neuroscientists start with models of higher-order cognitive systems, like memory, and assume that these break down into simpler, dumber sub-systems, like encoders and retrievers. As a result, many cognitive neuroscientists design their experiments in order to locate the mechanisms that act as encoders during encoding, or as retrievers during retrieval. Good old-fashioned cognitive neuroscience tries, as it were, to find the dumb homunculi in the brain.

But now Dennett is suggesting that the neural structures we seek may not be the dumb homunculi that we hoped they'd be. Rather, they may involve a much more complex set of entities organized and con-

nected in ways that do not correspond to the levels of the top-down decomposition. Consequently, the functional descriptions for the neural structures we identify will differ depending on whether we characterize them from the bottom-up or top-down perspectives. But if this is the case, then how are we to interpret neural "activations" measured in neuroscience? Do we interpret them as performing computations relevant to a function characterized from the top-down perspective, or to one characterized from the bottom-up perspective? What do we do when the neural networks whose activity we measure do not map onto "encoders" and "retrievers," but should rather be described in ways that make sense of the behavior of neurons and neuronal networks as agents?

A more pointed way of asking these questions is the following: what happens to cognitive neuroscience if the assumption of its goal — the existence of a clean (though perhaps difficult to find) mapping between our cognitive ontology and neural structures — is flawed? What if there is no clean mapping, or we find that we are able to generate many plausible but incompatible mappings? Should we reconsider cognitive neuroscience itself, if the assumption that drives it turns out to be false or misconceived? How do we do cognitive neuroscience at all, if there are principled reasons why we cannot bridge intentional descriptions made at the level of people and at the level of neurons and neural networks?

The shifts in Dennett's views on homuncular functionalism raise interesting questions about how he conceives of the mind-brain mapping. But they also reveal tensions endemic to cognitive neuroscience itself, for issues about intentional descriptions at different levels of organization lie at the heart of our attempts to move between psychology and neuroscience. In some ways, the pragmatic tendencies found elsewhere in Dennett's work may be of value here: perhaps all cognitive neuroscience can do is try to elaborate the best mapping or set of mappings, given the philosophical constraints we've discussed. Does this make Dennett a sort of instrumentalist or a sort of realist about the mapping itself?

VI. Conclusion

In its early form, Dennett's homuncular functionalist view argued for a top-down functional decomposition of cognition. At higher levels in the decomposition, the descriptions of these states will require intentional language. But as we decompose each level into its subcomponents, we begin to eliminate the intentional descriptions, until finally we arrive

at a "dumb" level whose behavior can be reproduced by a machine following simple instructions.

A functional decomposition like this one assumes that the bottom level really is mindless, and that we do not need any intentional language to describe its behavior. It is on this point, however, that FBBB suggests an important change — we can no longer consider neurons to be so dumb. That is, their behavior is complex enough that accounting for it requires the intentional stance as well. Thus our top-down functional decomposition appears to reach a level where intentional language re-enters the picture.

We have argued that this version of the homuncular functionalist view faces a problem: the functional descriptions used to capture the behavior of neurons may not fit well in the nested hierarchy of functional decompositions achieved from a traditional top-down analysis. If this is the case, then mapping mental states and processes to neural structures may be difficult or even impossible.

Dennett's *From Bacteria to Bach and Back* takes up a number of views he has discussed in previous work. Here we identify one, homuncular functionalism, and show how Dennett's change of mind opens powerful explanatory possibilities for neural structures. But this power may come with a price, as it becomes unclear how to link the different approaches we can take to characterizing the mind, brain, and their functions.

Department of Philosophy
Duke University
201 West Duke Building
Campus Box 90743
Durhan, NC 27708 USA
E-mail: bgesslo7277@gmail.com

Department of Philosophy
Duke University
201 West Duke Building
Campus Box 90743
Durhan, NC 27708 USA
E-mail: felipe.debrigard@duke.edu

REFERENCES

ANDERSON, M. (2014), *After Phrenology: Neural Reuse and the Interactive Brain*; MIT Press.

ARMSTRONG, D. M. (1968), *A Materialist Theory of Mind*. Routledge, 1968.

BLOCK, N. (1978), "Troubles with Functionalism," *Minnesota Studies in the Philosophy of Science* 9, pp. 261-325.

CARUSO, V. *et al.* (2018), "Single Neurons May Encode Simultaneous Stimuli by Switching Between Activity Patterns," *Nature Communications* 9 (2715), pp. 1-16.

CLARK, A. (1989), *Microcognition*; The MIT Press.
CUMMINS, R. (2000), "'How Does it Work?' versus 'What Are the Laws?': Two Conceptions of Psychological Explanation"; in *Explanation and Cognition*, edited by Frank Keil and Robert Wilson The MIT Press.
DE BRIGARD, F. (2007), Review of "Sweet Dreams: Philosophical Obstacles to a Science of Consciousness", by Daniel C. Dennett. (MIT Press: Cambridge, MA. 2005). *Ideas y Valores: Revista Colombiana de Filosofía.* 56 (134), pp. 129-136.
— (2012), "Predictive Memory and the Surprising Gap". Commentary on Andy Clark's "Whatever Next? Predictive Brains, Situated Agents and the Future of Cognitive Science". *Frontiers in Psychology.* 3, p. 420.
DENNETT, D. (1969), *Content and Consciousness*; London, Routledge.
— (1978), *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.
— (2005), *Sweet Dreams*. Cambridge, MA. The MIT Press.
— (2017), *From Bacteria to Bach and Back: The Evolution of Minds*; WW Norton & Company.
GESSELL, B. (2017), "Indeterminism in the Brain"; *Biology and Philosophy* 32 (6), pp. 1205-1223.
GODFREY-SMITH, P. (2016),"Mind, Matter, and Metabolism," *Journal of Philosophy* 113, pp. 481-506.
LEWIS, D. (1970), "How to Define Theoretical Terms"; *Journal of Philosophy* 67 (13), pp. 427-444.
— (1972), "Psychophysical and Theoretical Identification"; *Australasian Journal of Philosophy* 50 (3), pp. 249-258.
LYCAN, W. (1981), "Form, Function, and Feel"; *Journal of Philosophy* 78 (1), pp. 24-50.
— (1987), *Consciousness*; The MIT Press.
MARR, D. (1976), "Early Processing of Visual Information"; *Philosophical Transactions of the Royal Society B* 275 (942), pp. 483-519.
MCCULLOCH, W. and W. PITTS (1943), "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics* 5, pp. 115-133.
PASTUZYN, E. *et al.* (2018), "The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer," *Cell* 172, pp. 275-288.
PUTNAM, H. (1960),"Minds and Machines;" in *Dimensions of Mind; A Symposium*, edited by S. Hook. New York University Press.
— (1967),"Psychological Predicates"; in *Art, Mind, and Religion*, edited by W. H. Capitan and D. D. Merrill. University of Pittsburgh Press, 1967.
SHALLICE, T. (1988), *From Neuropsychology to Mental Structure*; Cambridge University Press.
SHIMIZU, K. and M. STOPFER (2013), "Gap Junctions"; *Current Biology* 23 (23): pp. R1026-R1031.
SIMON, H. (1969), *The Sciences of the Artificial*; The MIT Press.
TURING, A. (1950), "Computing Machinery and Intelligence," *Mind* 49, pp. 433-460.