Kybernetes

Reductionism ad absurdum: Attneave and Dennett cannot reduce Homunculus (and hence the mind)

Lance Nizami,

Reductionism ad absurdum

Attneave and Dennett cannot reduce Homunculus (and hence the mind)

Lance Nizami
Independent Research Scholar, Palo Alto, California, USA

Abstract

Purpose – Neuroscientists act as proxies for implied anthropomorphic signal-processing beings within the brain, Homunculi. The latter examine the arriving neuronal spike-trains to infer internal and external states. But a Homunculus needs a brain of its own, to coordinate its capabilities – a brain that necessarily contains a Homunculus and so on indefinitely. Such infinity is impossible – and in well-cited papers, Attneave and later Dennett claim to eliminate it. How do their approaches differ and do they (in fact) obviate the Homunculi?

Design/methodology/approach – The Attneave and Dennett approaches are carefully scrutinized. To Attneave, Homunculi are effectively "decision-making" neurons that control behaviors. Attneave presumes that Homunculi, when successively nested, become successively "stupider", limiting their numbers by diminishing their responsibilities. Dennett likewise postulates neuronal Homunculi that become "stupider" – but brain-wards, where greater sophistication might have been expected.

Findings – Attneave's argument is Reductionist and it simply assumes-away the Homuncular infinity. Dennett's scheme, which evidently derives from Attneave's, ultimately involves the same mistakes. Attneave and Dennett fail, because they attempt to reduce intentionality to non-intentionality.

Research limitations/implications – Homunculus has been successively recognized over the centuries by philosophers, psychologists and (some) neuroscientists as a crucial conundrum of cognitive science. It still is.

Practical implications – Cognitive-science researchers need to recognize that Reductionist explanations of cognition may actually devolve to Homunculi, rather than eliminating them.

Originality/value – Two notable Reductionist arguments against the infinity of Homunculi are proven wrong. In their place, a non-Reductionist treatment of the mind, "Emergence", is discussed as a means of rendering Homunculi irrelevant.

Keywords Brain, Homunculus, Neuron, Reductionism, Attneave, Dennett

Paper type Research paper

1. Introduction

In an earlier work in *Kybernetes*, Nizami (2015) examines how neuroscientists interpret Shannon's "general communication system", the basis for Shannon Information Theory. Nizami demonstrates that Shannon overlooks a crucial element that is implied to be in his system, namely, the *observer*, a concept of great importance to Second-Order Cybernetics. Shannon's system is symmetric; that is, its "reception" side is a mirror-image of its "transmission" side. But there is no "reception" side in a research animal's brain. Whatever neuroscientists imagine as Shannon's "signal", such as a train of neuronal voltage-spikes (reviewed in Nizami, 2017), is not converted back to a "message" (the communications-engineering operation called "decoding"). In fact, it is neuroscientists who act as Shannon's

My thanks to Claire S. Barnes PhD for her insights. Some arguments from the present paper appeared in rougher form in Nizami (2014b) and Nizami (2016). I sincerely thank the two anonymous Reviewers for their thought-provoking advice.

Kybernetes
© Emerald Publishing Limited
0368-492X
DOI 10.1108/K-10-2016-0266

missing observer(s), performing "decoding" within the laboratory. The neuroscientists thereby act as proxies for implied anthropomorphic signal-processing beings within the animal's brain – the *Homunculi*. Nizami (2015) re-describes the concept of the Homunculus and briefly summarizes two schemes that claim to rationalize-away the Homunculi, namely those of Attneave (1961) and of Dennett (1981). Nizami (2015) notes that the latter paper apparently derives from the former and that the former paper is wrong. But Nizami (2015) omitted the proof, for lack of space. That proof is now supplied, along with an exploration of the underlying Reductionism of Attneave (1961) and of Dennett (1981), which first necessitated the proof. Finally, an alternative way of thinking about Homunculi is advanced, which allows the elimination of Homunculi altogether. These issues continue to be timely; hence, all are discussed using the "present tense".

First: who is Homunculus? Neuroscientists and philosophers portray Homunculus as a miniature copy of its host, living within each host's head without the host's awareness. Homunculus has a long history. René Descartes (quoted in Kenny, 1971, p. 66) warned against believing in the "visual Homunculus", who "sees" images for us within the brain, a notion now known as the "Cartesian theater". What is allegedly seen by Homunculus are the neuronal voltage spikes that arrive at the host's brain, some from sensory receptors which interact with the external world (eyes, ears, skin, nose, mouth), some from internal sensor cells that interact with the internal world (the world of blood carbon dioxide, blood oxygen, muscle tension, etc.). Homunculus uses the incoming spikes to infer the state of the external and internal worlds.

Note that the Homunculus itself (henceforth called the Prime Homunculus) seemingly needs legs (or their equivalents), to roam the brain; arms, to isolate particular neurons and eyes, to observe spike firing. To coordinate these capabilities – and to remember the observations, to better "discriminate" between stimuli in the future – Homunculus needs a brain of its own. But such a brain needs its own internal Homunculus, likewise to perform "computations" (as the literature often states) on behalf of the previous Homunculus. An infinity of "nested" Homunculi is necessitated. Figure 1 illustrates the first few nestings.

An infinite regression of Homunculi can be unwittingly called-upon elsewhere in the nervous system, not just at the brain. For example, an infinite Homuncular regression lies hidden within sensory receptors (such as inner hair cells of the ear or retinal cells of the eye), in an allegedly Information-Theory-based multi-paper account of perception authored by K.H. Norwich. This Homunculus error is revealed in Nizami (2009a) and further criticism appears in *Kybernetes* (Nizami, 2011, 2013) and in other cybernetics journals (Nizami, 2009b, 2010, 2012a).

An infinity of Homunculi is absurd; experiencing any [internal or external] stimulus might require infinities of time and/or energy. Therefore, the number of successive nestings must be finite – or, better yet, Homunculus must be eliminated entirely. Two famous Homunculus-removal schemes exist, to which citations appear daily: Attneave (1961) and Dennett (1981), in books that are still in print. As Newell (1980, p. 715) notes (original internal quotation marks):

> A major item on the agenda of cognitive psychology is to banish the homunculus (i.e., the assumption of an intelligent agent (little man) residing elsewhere in the system, usually off stage, who does all the marvelous things that need to be done actually to generate the total behavior of the subject) [. . .] who is renamed the "executive" in many models.

Baddeley (1998) in fact claims credit for the so-called "central executive" (citing Baddeley and Hitch, 1974). Regardless, Newell's (1980) remarks remain true.

Schinkel (2006), Drayson (2012), Rose (2012) and Boone and Piccinini (2016) all attempt to banish Homunculi and all refer specifically and uncritically to the Homunculus-limiting schemes of both Attneave (1961) and Dennett (1981). Other authors agree uncritically with

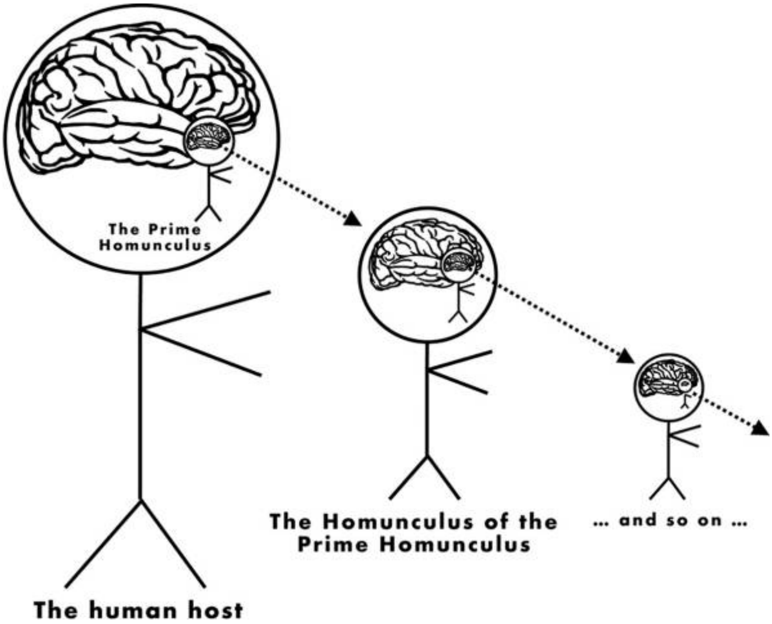


Figure 1. The infinity of Homunculi (see text)

Attneave (1961) but fail to cite Dennett (1981) (Sanders, 1998; Crick and Koch, 2000; Posner and DiGirolamo, 2000; Crick and Koch, 2003; Mumford, 2003; Baddeley, 2012; Woodman and Schroeder, 2012). Yet other authors fail to mention Attneave (1961), but agree (perhaps reluctantly) with Dennett (1981) (Monsell and Driver, 2000; Verbruggen *et al.*, 2014). Other writers cite neither Attneave nor Dennett, but nonetheless recognize that "the notion of a central executive [that controls the selection of actions and cognitive functions more generally] is tantamount to positing a homunculus (small man)" (Hazy *et al.*, 2007, p. 1601), hence there being a need "to deconstruct this implicit homunculus" (Hazy *et al.*, p. 1609). Numerous other citations could be mentioned.

The schemes of Attneave (1961) and of Dennett (1981) will be presently synopsized and will be found faulty.

2. According to Attneave (1961): a finite number of Homunculi

2.1 Attneave's verbal description

Attneave (1961) allegedly obviates the infinite regression of Homunculi and influential researchers have believed him (Crick and Koch, 2000, 2003). Indeed, the book *Sensory Communication* (reprinted in 2012), in which Attneave's work appears, has long been a standard reference-work in Cognitive Science. Altogether, Attneave (1961) must not be ignored.

Attneave (1961, p. 778) begins as follows: "If a homunculus exists it must certainly be composed of neurons". Hence, "We fall into a regress *only* if we try to make the homunculus do everything" (Attneave, p. 778; original italics). Further (Attneave, p. 778; original italics):

> The moment we specify certain processes that occur *outside* the homunculus, we are merely classifying or partitioning psychoneural [*sic*] functions; the classification may be crude, but it is not in itself regressive.

He continues (Attneave, p. 778):

> Indeed, one might even hypothesize a series of concentric or nested homunculi without falling into a regress, provided each contained an outer layer of functions not contained in the next smaller one.

Attneave (1961) then borrows from Bullock (1961). That is, Attneave (1961, p. 778, original internal quotation marks) proposes that:

> For any given behavior there must be at least one neuron that "decides", on the basis of activity in receptors and other neurons, whether to initiate that behavior or not.

Consequently "There may be as many 'homunculi' of this sort as there are coherent behavior patterns" (Attneave, p. 778; original internal quotation marks).

Attneave (1961) strongly implies that his "psychoneural functions" (see above) are either behaviors or, operationally, groups of interconnected neurons which altogether underlie those behaviors. The notion that behaviors arise from distinct, separate groups of neurons is *Reductionist*. And it is intensely popular. Hein *et al.* (2016) allegedly have found the circuit for motives, Matthews *et al.* (2016) have found the circuit for loneliness, Chou *et al.* (2016; reviewed by Desban and Wyart, 2016) have found the circuit that "determines who wins or loses a fight" (Desban and Wyart, p. 42), Zalocusky *et al.* (2016) identify a group of neurons for "risk selection" and Kondoh *et al.* (2016) identify a group of neurons for "an instinctive fear response in mice". All of these neural groups are Attneave's "psychoneural functions" – effectively, they are Homunculi or the things that Homunculi control.

This notion of neural groups as "psychoneural functions" arises because adding neuronal "circuits" to other neuronal "circuits" merely creates larger and more-complicated circuits, whereas adding conscious parts to other conscious parts presumably creates another conscious part – perhaps even a mind. Therefore, there is a drive to assign consciousness, in the form of a conscious Homunculus, to the smallest possible units of the nervous system. Those are usually taken to be individual neurons. In the literature, neurons have been anthropomorphized; they have been granted human qualities such as consciousness. This approach eliminates any need to explain how consciousness and a mind might "emerge" from a mass of neurons. The practice of assigning human qualities to neurons is so commonplace that an entire companion paper is devoted to examining it (Nizami, 2017). Note well, however, that Homunculus is traditionally invoked to *create* the mind from the brain (more on this below). "Mind" is, however, difficult to conceptually separate from "consciousness"; the issue will not be debated here. Presently, let us assume that a necessarily conscious Homunculus would have a mind, which would nonetheless be traditionally called upon to create a mind (mine or yours) from a neuron or a group of neurons, perhaps a brain. Such paradoxes as minds being required to create minds reveal the impossibility of Homunculus and are revisited in the Discussion.

2.2 A physical understanding of Attneave's verbal description

Attneave (1961) is remarkably vague, perhaps intentionally so. Let us assume that each "function" involves a system of interconnected neurons. Let us also assume that there are "Homuncular layers" of functions. A subset of the functions on each layer are controlled by an external Homunculus, one per layer. Consider the "Top Layer", having enough functions (here, "n") for a "mind". At Top Layer, there is Top-Layer Homunculus, controlling a number $m < n$ of the Top Layer functions. Figure 2 illustrates the scheme.

But Top-Layer Homunculus contains (according to Attneave) a sub-Homunculus, here called Second-Layer Homunculus. It controls l functions, where $l < m$. Within Second-Layer

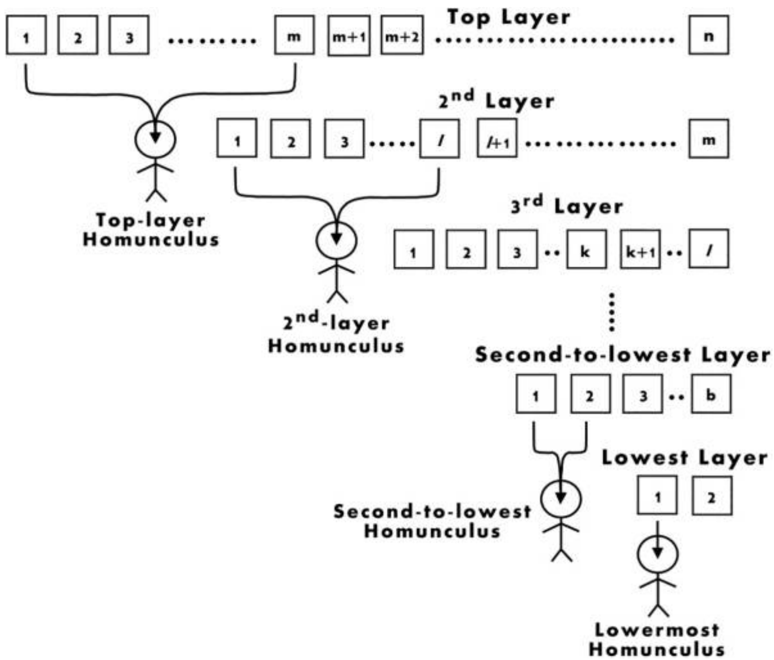


Figure 2.
An interpretation of Attneave's (1961) proposed hierarchy of Homunculi that deal with "psychoneural functions" (the boxes)

Homunculus is Third-Layer Homunculus, controlling k functions, k < *l*. Within Third-Layer Homunculus is Fourth-Layer Homunculus, controlling j functions, j < k [. . .] *and so on*. The number of functions controlled by each succeedingly lower-layer Homunculus decreases, until eventually we reach the Second-to-Lowest Layer (Figure 2), which has b functions (b ≪ j), of which only *two* are controlled by Second-to-Lowest Homunculus. Below the Second-to-Lowest Layer is the Lowest Layer, consisting of only two functions – one that is controlled by Lowermost Homunculus and one that is not.

3. Analysis: Attneave is wrong

3.1 How do nested Homunculi work as controllers?

If this scheme is indeed what Attneave (1961) intends, then Attneave is dead wrong. Here is why. Consider what happens if Second-to-Lowest Homunculus *hypothetically* controls only two functions, regardless of the number of *uncontrolled* functions on that layer. Then, after Attneave (1961), one of those two functions would *actually* be controlled by Lowermost Homunculus, leaving Second-to-Lowest Homunculus to *actually* control only one function. If, likewise, Third-to-Lowest Homunculus *hypothetically* controlled only three functions on its layer, then likewise it would *actually* control only one function. This effect propagates upward, such that Top-Layer Homunculus itself *actually* controls only one function (in Figure 2, it is function "m"), *if* the number of functions hypothetically controlled by each Homunculus at each layer continues to increase by just one. But all of this contradicts Attneave's notion that if any Homunculus controls only a single function, then it does not contain nested Homunculi controlling any other functions. The contradiction remains, even if we allow the number of functions that a given sub-Homunculus controls to be incremented by more than one over the number controlled by the next-lower Homunculus, or if the number of functions controlled by Lowermost Homunculus is likewise more than just one.

This argument may be difficult for the reader to follow. Certainly, it is difficult to depict graphically, using only two dimensions. But its essence is simple: when controllers are nested, each controller must have the same responsibilities as the others. That is, the number of "functions" controlled by each controller remains the same across controllers. This is what Figure 1 illustrates when each controller (i.e. Homunculus) actually controls *all* "psychoneural functions" associated with the brain.

In sum, "nested control" obviates the very notion of "control" as something done by a single "controller" (in contrast to an infinity of controllers), that is, a single distinct neuronal Homunculus "that 'decides', on the basis of activity in receptors and other neurons, whether to initiate that behavior or not" (Attneave, 1961, p. 778).

3.2 How do nested controllers serve not-nested neurons?

There is another issue. Recall Attneave's statement (1961, p. 778) that "If a homunculus exists it must certainly be composed of neurons". If so, then nesting of Homunculi within Homunculi implies nesting of neurons within neurons. "Nested" presently means "inside of". Now, neurons all have connections to other neurons, or to a sensor cell and another neuron, such that no neuron (or group of neurons) are actually "nested *around*" by other neurons or "nested *amidst*" other neurons. That is, no neuron is physiologically isolated from other cells.

How then would a Homunculus composed of neurons be realized physiologically? Maturana and Varela (1987/1998, p. 162) note that a part of the mammalian visual system called the lateral geniculate nucleus (LGN) receives voltage spikes from at least six different parts of the nervous system within the head (Zeki, 1993). Therefore, as Maturana and Varela (1987/1998) note, the LGN cannot simply be a "relay station" between the retina and other mammalian vision-associated nuclei. And indeed, Tibbetts (1995) notes that the reciprocal properties of the LGN are recognized by Harth (1993), who attempts to make them crucial to a theory of visual perception. Unfortunately, according to Tibbetts (1995), Homunculi arise in Harth's scheme. Another example of a neural Homunculus comes from Crick and Koch (2005; see the feature by Reardon, 2017). Crick and Koch (2005, p. 1276) postulate that a substructure of the brain called the *claustrum* is "A conductor coordinating a group of players in in the orchestra, the various cortical regions". Of course, this "conductor" is just the "central executive" by any other name: Homunculus.

Complicated reciprocal connections indeed occur in mammalian *auditory* neurophysiology (Ehret and Romand, 1997). Among other effects, this potentially confounds any interpretations of psychophysical data (Nizami, 2012b). Consider such connectivity, then, in the context of an infinite nesting of Homunculi (Figure 1). Back-and-forth exchanges of neuronal voltage-spikes between physical groupings of neurons, such as the LGN and its partners, occur without the neurons being "nested" in the way that a Homunculus is nested *within* a Homunculus. Nonetheless, spikes are often imagined as "information flow" (Whitfield, 1984, among countless others, some of them cited in Nizami, 2014a, 2015). As such, according to Attneave (1961), Homunculi would "direct" or "interpret" that flow. Indeed, each "information flow" might conceivably have its own dedicated Homunculus. Homunculi are imagined as being nested (e.g. Figure 1); but neurons cannot be *inside* neurons. How, then, could Homunculi that serve neurons nonetheless be *inside* Homunculi?

3.3 Attneave attempts to substitute one form of organization for another
What Attneave (1961) evidently tries to do and which Dennett (1981) also explicitly attempts (see Section 6, below), but under a business analogy, is the separation of "commanders" from their "platoons". Consider an Army Division, divided into battalions, themselves divided into regiments, themselves divided into companies, themselves divided into platoons. Each sub-unit has its own commander; each ultimate platoon-commander takes orders from his respective company commander and so on up to the level of the lone General who commands the Division. All of the Division's physical sub-units are hence groups of soldiers. When arrayed on the parade ground, these groups are *physically adjacent*, but they are not *physically nested*. Nonetheless, the organization of the Division can be diagrammed on paper as an *organizational nesting*, using boxes, with platoons as boxes within their companies and so on up to the largest organizational box within which all of the soldiers reside, namely the Division. Likewise, the various commanders are not physically nested, but they nonetheless form an *organizational hierarchy*, which could be represented on paper by nested boxes, but might best be represented by a chain-of-command *tree structure* (see, e.g. Feinberg, 2011, for the example of biological cells).

Consider now an infinite nesting of Homunculi. It is indeed a physical nesting. However, it is also an *organizational nesting*, not merely an *organizational hierarchy*. That is, each Homunculus within the infinite regression depends (for instructions) upon all of the Homunculi nested within itself. In these respects, the infinite nesting of Homunculi differs from the organization of military commanders. Nonetheless, if soldiers (possibly including their commanders too) can be thought of as neurons, then what Attneave (1961) and Dennett (1981) apparently attempt to do is to replace the classic nesting of Homunculi by the military-command hierarchy. But such a transformation appears impossible *prima facie*. (For example, Feinberg (2011) posits "nested" groups of neurons, but on closer inspection they appear to be merely *adjacent* groups.)

Altogether, Attneave's scheme for the elimination of the infinite nesting of Homunculi does not work. And yet others (see above) believe in his scheme. They, like Attneave, may be making crucial errors, perhaps in the form of inappropriate assumptions. It is important, therefore, to scrutinize Attneave's assumptions.

4. The source of Attneave's error: revisiting the traditional infinity of Homunculi

A crucial assumption underlies the imagined need for the Homunculus. The assumption in question has been stated already, but bears repeating: namely, that the nervous system of any species is inherently "stupid", indeed, that it is *useless* because it contains only a mass of neuronal voltage spikes which form "internal representations" of external phenomena. Those representations are believed to be "stupid" because they lack "sense" or "meaning". Lörincz *et al.* (2002, p. 189) point out the erroneous belief that "an internal representation is still meaningless unless someone can "read" or interpret it" (original internal quotation marks) (see also McMullen, 2001). In other words, the erroneous belief is that "the internal representation should *make sense*" (Lörincz, 2009, p. 126; italics supplied). Such things as "sense" or "meaning" allegedly only therefore arise when the spikes are subjected to "examining", "remembering", "knowing" and "deciding" by a Prime Homunculus, to create an actual *mind*. Of course, such activities are what already *defines* an actual mind – so that a mind is needed in order to have a mind and so on *ad infinitum*. This is what is shown by Figure 1. More is said on this below.

Note well an important point: that *each mind-within-a-mind must be equally capable*. Even nested Homunculi within the brain of an insect must achieve the same degree of sophistication as the insect's Prime Homunculus, even if that Prime Homunculus is less capable than a human being's. But Attneave (1961) allegedly escapes the infinity: "We fall into a regress *only* if we try to make the homunculus do everything". Attneave says:

> Indeed, one might even hypothesize a series of concentric or nested homunculi without falling into a regress, provided each contained an outer layer of functions not contained in the next smaller one.

Effectively, our so-called Prime Homunculus is replaced by a Homunculus of lesser responsibility and hence less-than-fully-human capability – namely, Top-Layer Homunculus (Figure 2). Similarly, Top-Layer Homunculus does not fully generate the human mind; hence, hypothetically, it likewise needs nothing within it to generate a *complete* mind – something of yet lesser capability is enough, namely, Second-Layer Homunculus. The latter, in turn, needs within it something of yet lesser capability still, namely, Third-Layer Homunculus [. . .] and so on, until a "stupid" Homunculus is reached that cannot be stupider, namely, Lowermost Homunculus.

Altogether, Attneave (1961) simply ignores the very definition of Homunculus; after all, if it does not "do everything", then what operates the remaining "psychoneural functions"?

Attneave's (1961) notion of Homunculi becoming progressively "stupider" was to be replicated by Dennett (1981), as follows.

5. Dennett (1981) and how *he* claims to dismiss the infinite regression of Homunculi

5.1 "Brainstorms" by Daniel Dennett: an introduction

In 1981, The MIT Press published a book copyrighted to Bradford Books, Publishers, in 1978. That book was *Brainstorms: Philosophical Essays on Mind and Psychology*, by Daniel C. Dennett (D. Phil., Oxon.). Dennett's book is an academic best-seller. It is still in print. It has been cited more than 4,000 times (GoogleScholar), far in excess of the minimum of 100 citations presently considered to be "significant" for a scientific work. It receives new citations daily.

5.2 How Dennett (1981) dismisses the infinite regression of Homunculi

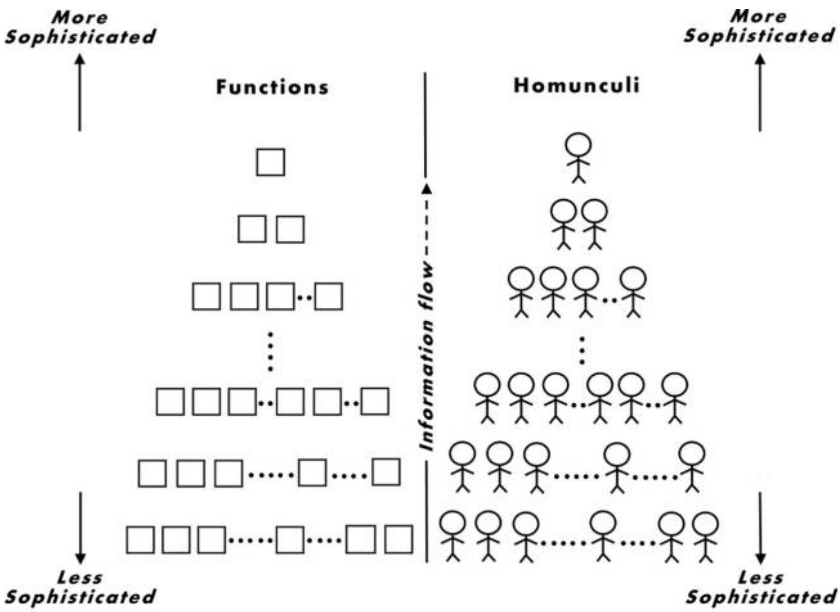
5.2.1 Replacing Homunculi by machines. Dennett's book largely concerns Homunculus. Nonetheless, Dennett does not get to the crux of his arguments until his eightieth page. There, he discusses artificial intelligence (AI), as a context within which to introduce an assumption that underlies many of his numerous remaining pages. In Dennett's words, "AI program designers work backwards on the same task behaviorists work forwards on" (Dennett, 1981, p. 80). That is, to Dennett, an AI researcher starts with a problem such as making computers "understand" and then "breaks it [the problem] down into sub-problems" (Dennett, p. 80). Those sub-problems involve operations such as "recognize", "distinguish" or "ignore". From there, the researcher "breaks *these* problems down still further until finally he reaches problem or task descriptions that are obviously mechanistic" (Dennett, p. 80; italics added). That is, "The AI programmer [. . .] frankly views the computer anthropomorphically" (Dennett, 1981, p. 80). Further, "the highest level design breaks the computer down into a committee or army of intelligent *homunculi* with purposes, information and strategies" (Dennett, p. 80; italics added).

Dennett (p. 80) states that "Each homunculus in turn is analysed into *smaller* homunculi, but, more important, into *less clever* homunculi" (original italics). Dennett (p. 80, original internal quotation marks) elucidates:

> 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, they have been reduced to functionaries "who can be replaced by a machine".

Figure 3 shows Dennett's hypothetical relation of functions to Homunculi.

5.2.2 Making the machines: the stupidifying of Homunculi. By now, the reader may have noted some similarity of Dennett (1981) to Attneave (1961). But to continue: Dennett (1981, p. 81) concludes that "if the [computer] program works then we can be certain that all homunculi have been discharged from the theory"; that is, "obviated from the program". Dennett (p. 101) further notes, "Any psychology with undischarged homunculi is doomed to circularity or infinite regress". Dennett's word "psychology" seems, in context, to mean a suite of human behaviors, perhaps all of those that characterize a human being (rather than, e.g. any body-of-thought purporting to provide psychological explanation). Dennett then proceeds to provide the missing details of his scheme, that is, *how* to "discharge" all Homunculi from a "psychology", viz.: "If one can get a team or committee of *relatively* ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress" (p. 123; original italics). But this is also Attneave (1961). In Attneave, as the Homunculi are successively nested they assume fewer responsibilities – that is, they get "stupider". Oddly, Dennett (1961) does not credit Attneave (1961). Regardless, Dennett now



Notes: The function(s) on each horizontal level control those below them. As the functions become more sophisticated, so do the Homunculi that control *them*. Sensory "information" flows from bottom to top, i.e., from less-sophisticated functions to more-sophisticated functions (that is, brain-wards). Note that each function might itself contain a similar hierarchy of functions

Figure 3.
An interpretation of Dennett (1981): Homunculi as *organization*. For each "function" (left-hand side), there is a Homunculus (right-hand side)

returns to his mind-analogue from artificial intelligence. He visualizes the mind-analogue anthropomorphically, as a corporate organization (Dennett, pp. 123-124; original italics):

> A flow chart is typically the organizational chart of a committee of homunculi (investigators, librarians, accountants, executives); each box specifies a homunculus by prescribing a function *without saying how it is to be accomplished* (one says, in effect: put a little man in there to do the job).

Hence, according to Dennett (p. 124):

> If we then look closer at the individual boxes we see that the function of each is accomplished by subdividing it via another flow chart into still smaller, more stupid homunculi.

Dennett continues (p. 124, original internal quotation marks):

> Eventually this nesting of boxes within boxes lands you with homunculi so stupid (all they have to do is remember whether to say yes or no when asked) that they can be, as one says, "replaced by a machine".

Of course, "to say yes or no" is to make a decision, something done by the whole human being. Regardless, Dennett (1981, p. 124) summarizes his model as follows: "One *discharges* fancy homunculi from one's scheme by organizing armies of such idiots to do the work" (original italics). But Dennett (1981), like Attneave (1961), provides no illustration of his Homuncular hierarchy. Rather, he provides a number of complicated but seemingly-irrelevant flowcharts, perhaps to emphasize his aesthetic of mind-as-computer.

5.3 Dennett's interpretation of Homunculus' place in physiology

Dennett (1981) also translates his scheme into physiology, specifically, the visual system. Dennett declares (1981, p. 124) that "The more raw and uninterpreted the [sensory] representation – e.g., the mosaic of retinal stimulation at an instant – the more sophisticated the interpreter or user of the representation". The "interpreter" presumably is Homunculus, lying close to the retinal receptors, perhaps as a neuron. Dennett continues (1981, p. 124), "The more interpreted a representation – the more *procedural* information is *embodied* in it, for instance – the less fancy the interpreter need be" (original italics). In other words, as the "representation" gets further from the retina and closer to the brain, the *less* sophisticated is the respective observing-controlling Homunculus. Dennett (p. 124, original italics) proceeds,

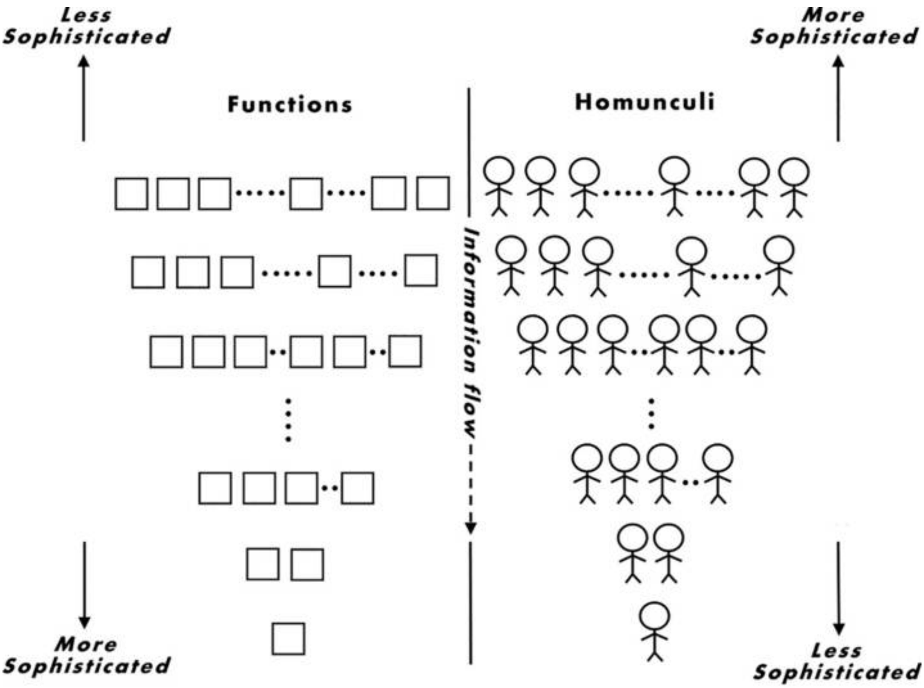
> It is this fact [*sic*] that permits one to get away with *lesser* homunculi at high levels, by getting their earlier or lower brethren to do some of the work.

Thereby, "All homunculi are ultimately discharged" (Dennett, p. 124). Dennett presumably means that Homunculi are reached who are so stupid that they are merely "machines" that make Yes/No decisions. An army of idiots does the work – but those idiots hypothetically work more centrally to the brain than the retina does.

Dennett (1981) provides no illustration of his model. Figure 4 shows the present author's interpretation. Of course, if Homunculi are neurons, then the order of neuronal sophistication in Dennett's scenario is exactly the *opposite* of what many physiologists and psychologists have postulated for decades; otherwise, the brain has nothing to do.

6. Analysis: Dennett is wrong

Dennett's (1981) concept of how to eliminate Homunculi strongly resembles that of Attneave (1961). And, like Attneave, Dennett proves to be wrong, as follows. Dennett (1981) imagines a Homuncular committee, each member controlling a function that can be subdivided into



Notes: Unlike Figure 3, the function(s) on each horizontal level control those *above* them. Figure 3 is effectively turned upside-down, because Dennett has changed the organization such that, as the functions become less sophisticated, the Homunculi that control them become *more* sophisticated (see text). Hence, sensory "information"-flow, being from less-sophisticated functions to more-sophisticated functions (i.e. toward the brain), is now graphically downwards. (Note that each function might itself contain a similar hierarchy of functions.)

Figure 4. An interpretation of Dennett (1981): Homunculi as physiology. As in Figure 3, for each "function" (left-hand side) there is a Homunculus (right-hand side)

other functions. Each Homunculus contains less intelligent Homunculi, within them yet-less-intelligent Homunculi, all controlling functions. Eventually a level is reached at which each of that level's Homunculi "can be replaced by a machine".

But just like Attneave (1961), Dennett (1981) ignores the fact that any Homunculus, by its very definition, is infinitely nested with Homunculi of *identical* capability, whatever that capability may be. And just like Attneave (1961), Dennett (1981) uses "stupidifying" to *redefine* Homunculus, not to remove it. Finally Dennett, just like Attneave (1961), assumes that Homunculi are neurons; but for neurons, "nesting" is a physical impossibility. *Organization-wise*, Dennett (1981) imagines Homunculi as employees within offices within departments within businesses. But *neuron-wise*, such structure is not possible. Indeed, the "stupidifying" of Homunculus-neurons with *rising* cognitive level opposes Dennett's own "committee" scheme, in which a "higher executive" (e.g. something at the brain) would be expected to be more sophisticated than an "employee" (e.g. a retinal cell). Compare Figures 3 and 4. The notion that cognitive "processing" becomes *more* sophisticated brain-wards was common belief by 1981. So why did Dennett (1981) postulate that the *brain* had the stupider Homunculi? The apparent answer is a paper called "Some informational aspects of visual perception" (Attneave, 1954;

3,200+ citations). Attneave (1954, p. 183) approaches perception "as an information-handling process" which "seems to involve a set of processes whereby information is *predigested* before it ever reaches awareness" (p. 187, original italics). That is (Attneave, 1954, p. 189):

> It appears likely that a major function of the perceptual machinery is to strip away some of the redundancy of stimulation, to describe or encode incoming information in a form more economical than that in which it impinges on the [visual photo-] receptors.

Cue to Dennett (1981), whose sophisticated interpreter "predigests" the "incoming information" to produce something that is sufficiently "economical" for *stupider Homunculi*, namely, those whom Dennett implies to be *closer* to the brain. In short, the work of Dennett transpires again to be the work of Attneave.

7. Discussion: the mind as computer

7.1 Neuronal "computation"

Dennett's (1981) approach to eliminating Homunculi is to treat brains as computers. That analogy is perpetuated by others in the fields of Philosophy and Psychology. For example, Margolis (1980, p. 246) asks "What, for instance, are the informational sub-routines neurophysiologically processed in the brain that facilitate what we call perception [. . .]?". Tibbetts (1995, p. 405) likewise asks, "How does simple data acquisition become transformed into perception?" and speaks of "neural computers" (Tibbetts, p. 407). Neuroscientists, too, imagine that sensory neurons that provide input to the brain and brain neurons themselves are computers. Indeed, sensory neurons may "adopt" some sort of "coding strategy" (Rieke *et al.*, 1997, p. 175). As such, it is "the specific computations being performed by the [individual] neuron under study" which must be discovered by the neurophysiologist, according to Chacron *et al.* (2003, p. 803; see also Passaglia and Troy, 2004, p. 1229). Neurons have many lab-imagined "coding strategies" to choose from (MacKay and McCulloch, 1952; Stein, 1967; Walløe, 1970; Optican and Richmond, 1987; Victor, 2006).

And naturally, the eye's retina has some such "computational strategy" (Atick and Redlich, 1992, p. 200). How accurate is it? Bialek *et al.* (1991, p. 1856) allegedly know, at least for flies: "The fly visual system approaches optimal real-time computation". Remarkable! An animal's vision may be as good as man-made algorithms! This notion has been seconded by (among others) Baden *et al.* (2016, p. 345), as follows: "Information channels from the mouse eye to the mouse brain" may have "an encoding strategy resembling that used in state-of-the-art artificial vision systems".

These are just a few examples of the computer analogy; the literature is vast. The computer analogy is likewise applied to neurons deep within the brain; consider, for example, that "Neurons at higher processing stations may become largely independent to allow for easier readout" (Chechik *et al.*, 2006, p. 359). Similarly, Ciocchi *et al.* (2015, p. 560) declare that some particular part of the brain "computes diverse information" and then sends "computations transmitted [*sic*]" to various brain "targets". Readers are referred to the companion paper (Nizami, 2017) for further examples of the use of the computer analogy in the literature.

7.2 Homunculus, the computer programmer

Any present-day computer depends upon a Homunculus – namely, the computer programmer. The programmer not only designs the computer's "thinking" (i.e. the software/ program), but provides input ("stimuli" or "data"). The programmer also reads the output, to reach conclusions (Presumably, so would one or more hypothetical Homunculi within the programmer's own brain, according to Dennett (1981)). Von der Malsburg (2014, personal communication) notes that "All interesting things like planning, problem solving, debugging and interpretation happen in the mind of the programmer, not in the computer. Here the programmer is the homunculus". If the brain is a computer, then, the programmer must reside within the brain – in the form of Homunculus. Searle (1994, p. 212) notes of the literature that "The idea always is to treat the brain as if there were some agent inside it [i.e. the Homunculus] using it to *compute* with" (italics added). That is, "Without the homunculus, there is no computation, just an electronic circuit. So how do we get computation into the brain without a homunculus?" (Searle, p. 221). Searle (pp. 212-225) further notes that a model of a neurophysiological process is not itself a neurophysiological process and *need not reveal anything about such a process*.

7.3 Homunculus, retriever of memories

As noted in the Introduction, the Prime Homunculus needs a memory to perform "remembering", "knowing", "deciding", "discriminating" and so forth. In the literature, there is an awareness of the need to describe human memory without inadvertently invoking Homunculi (Dalla Barba, 2001; Lörincz *et al.*, 2002; Hazy *et al.*, 2007; among others). But there is another, closely-related danger, namely, that of confounding human memory with computer memory. Both of these concerns are inherently noted by von Foerster (2003/2010). He cautions against viewing human memory as a "document storage and retrieval system" (von Foerster, p. 103) that involves memory retrieval by a mobile "little demon" (von Foerster, p. 103), a demon that needs its own memory of where memories were stored, that is, a Homunculus. Of course, the Homunculus becomes part of an infinite nesting of Homunculi, namely, the very one shown in Figure 1.

Von Foerster implies that the "demon" arises because scientists confuse the "*vehicles* for potential information with *information*" (von Foerster, p. 103; original italics). That is, scientists confuse storage media for what is being stored, namely, *memories*. Von Foerster (p. 103) further implies that the role of *cognition*, the job traditionally directed by Homunculus, should be to "lift" information from its carrier signals (von Foerster's original quotation marks). Von Foerster (p. 103) then states his particular aim, namely, to "ultimately dismiss the demon and put his brain right there where ours is". That is, von Foerster wishes to render Homunculus superfluous. His proposed method involves neural "memory units". But it is unclear whether von Foerster (2003/2010) achieves his goal, as there are too many missing details, a characteristic of his other chapters in that particular book as well. Regardless, von Foerster seems to pursue a computer analogy, like Dennett (1981). Regarding computers and memory, Naka and Sakai (1993, p. 80) might offer the last words here: "Matter is not information itself, just as chips are essential to a computer but are really sets of instructions that turn a computer into a computing machine".

Such analogies represent Reductionist approaches. Reductionism is of ongoing concern in cybernetics and for good reasons, as follows.

8. Discussion: Attneave, Dennett and Reductionism

8.1 Attneave (1961) and Dennett (1981) as Reductionists

Reductionism is the attempt to explain a complicated behavior or action, emanating from a complicated mechanism, in terms of simpler behaviors or actions, emanating from simpler mechanisms (Nizami, 2017). The Attneave (1961) and Dennett (1981) schemes are fundamentally Reductionist. Many others follow their lead. Crick and Koch (2000, p. 10) declare that the "problem of [explaining] consciousness [. . .] needs to be approached in a reductionist, scientific manner". Note well the equating of "scientific" to "reductionist", by two highly-placed contemporary scientists. Similarly, Monsell and Driver (2000, p. 27) favor "the progressive fractionation and localization of control subfunctions, through the combination of chronometric performance analyses, neuropsychology, functional imaging, electrophysiology and neuropharmacology". However, they warn that "It remains to be seen whether, in due course, the control homunculus will turn out to have been merely fractionated or completely dissolved" (Monsell and Driver, p. 27). They continue: "Although the picture remains murky, each new result adds a little light and we are beginning to discern the identities of some recruits to the army of control 'idiots'" (Monsell and Driver, p. 28; original internal quotation marks). Of course, those "idiots" are merely Dennett's (1981, p. 124) "homunculi so stupid [. . .] that they can be, as one says, 'replaced by a machine'". This is Reductionism, as echoed in Verbruggen *et al.* (2014, p. 514):

> We believe that in order to understand how control is achieved, boxes [*sic*] have to be broken down until we understand how complex behavior arises from a basic set of cognitive processes that can be implemented by our neural system.

Or, as Hommel *et al.* (2004, p. 103) declare:

> Cutting human performance into pieces that are simple enough for us to understand seems to be the only way to get rid of the homunculi still hiding behind our theories.

The present author would, of course, disagree.

8.2 Margolis (1980) and Sober (1982): how can we "reduce" the "intentional"?

The Reduction favored by Monsell and Driver (2000) and Verbruggen *et al.* (2014) and others may not even be conceptually possible. After completing the analysis shown above, the present author became aware of a critique of Dennett that was published in 1980, based upon the original 1978 printing of Dennett's book (see Section 5.1). In the critique, Margolis (1980) argues in terms of the "intentional", that is, the "purposive" (Margolis, p. 245), which involves "ascriptions of real psychological states, conscious and cognitively qualified" (Margolis, p. 255). Margolis also uses the words "molar", referring to the whole human and "molecular", referring to its components. In those terms, Margolis states a conclusion that was eventually reached by the present author (Section 3), namely, that the lowermost Homunculi, which Dennett (1981) describes as "stupid", cannot be so: "There remains a sense in which, in discharging "fancy homunculi", our "more stupid homunculi" never actually perform in a way that approaches purely non-intentional functioning" (Margolis, 1980, p. 251; original internal quotation marks). Margolis (p. 251) further declares that:

> The only purpose the description [Dennett's discharging of Homunculi] can serve is to suggest the fair sense in which intentional molar states and processes may be successfully analyzed into component intentional states and processes at the molecular level: we can make the components as numerous and as simple as we please, but (admitting the irreducibility of the intentional and the reality of intentional states) we are logically constrained to treat the relevant molecular processing as intentionally or informationally qualified.

In other words, "Neither he [Dennett] nor we know of any way to reduce the intentional to the non-intentional" (Margolis, 1980, p. 255).

Or, as McMullen (2001, p. 163) notes under the assumption that "intentionality" requires "knowing":

> One cannot derive cognitive activities from non-cognitive ones by breaking the former down to subsets of the latter and this is because the concept of "completely stupid knowers" is self-contradictory.

In short, Dennett and his predecessor, Attneave (1961), were bluffing. And even if Dennett (1981) thinks that he can reduce intelligent Homunculi to stupid ones, the crucial detail of *how* is absent. Or, as Margolis (1980, p. 254; original italics) puts it,

> Dennett's program makes no sense unless he can show how the conscious and cognitive capacities of molar agents can be analyzed, without remainder, into the systematically linked capacities of molecular homunculi.

Note the crucial element *without remainder*; or, as noted above (Section 4), if an Attneave (1961) Homunculus does not "do everything", then what operates the remaining "psychoneural functions"?

The same point about Dennett (1978/1981) is made by Sober (1982), in a roundabout fashion. Sober differentiates between "smart" Homunculi and "stupid" Homunculi as follows: "Smart homunculi [as in the Cartesian theater] may explain why I now see the page in front of me, but they do not explain what seeing *is*" (Sober, p. 422; italics added). Seeing is a sensory mechanism and the Homunculi that explain sensory mechanisms are *compelled* to be stupid, so Sober implies. Nonetheless, the intentional cannot be reduced to the non-intentional. Hence, each nested Homunculus must be as capable as the next, resulting in an infinity of Homunculi, the infinity of Figure 1. But nesting the Homunculi still does not explain what seeing *is*.

The dichotomy of "seeing" versus "perceiving" has long been noteworthy. Kenny (1971) had warned about the persistence of hidden Homunculi (more on this below). Rorty (1971) produced a directed retort, in which she criticizes Kenny's (1971) warnings in a fashion that is unintentionally ironic (Rorty, p. 78; original internal quotation marks):

> It might seem as though Kenny were saying: on the one hand there is a mechanistic account of the physical states "associated" with seeing. These states can be explained scientifically. On the other hand, there is an activity of persons, called "perceiving", which must be given quite a different sort of explanation.

Rorty declares her disapproval of such a dichotomy. But that dichotomy was our state of understanding in 1971 and it remains our state of understanding today.

9. Discussion: Kenny (1971), Rorty (1971) and Reductionism

9.1 Kenny (1971) warns of Homunculi

Kenny (1971) focuses on Homunculus, in the context of criticizing attempts by René Descartes (see citations in Kenny) and by Richard Gregory (1966) to express their own warnings about "the postulation of a little man within a man to explain human experience and behavior" (Kenny, 1971, p. 66). Descartes and Gregory are concerned with vision, where Homunculus hypothetically manifests as a pair of "second eyes", inside the brain of whatever creature experiences sight. But Kenny interprets Descartes and Gregory as both inadvertently invoking Homunculus while they are warning against invoking Homunculus! Elsewhere, Tibbetts (1995, p. 406) provides one argument against a visual Homunculus at the brain:

> Ironically, only at the retinal level is all the information that will be relayed to different feature-extraction areas [closer to the brain] still intact in one location. If a Homunculus were to be anywhere, then, it would be sitting just behind the retina.

As for the putative physiological "feature-extraction areas", those at the brain itself are described by Zeki (1993), amongst others.

Kenny (1971) was immediately followed by a retort by Rorty (1971), a retort that expresses the spirit of Reductionism in neuroscience and perhaps its motivation too, as follows (Rorty, p. 79):

> Someone might hold the extreme view that once we have a really good neuro-physiological account of the mechanisms of perception, we could tell whether and possibly even what (although this seems more questionable) a subject has seen by having him wired up appropriately.

This relates to an important statement made by Gregory (1966) and mentioned by Kenny (1971, p. 68), viz., "How is information from the eyes coded into neural terms, into the language of the brain and reconstituted into experience of surrounding objects?". Kenny (1971, p. 68) correctly protests that the brain itself does not speak a "language". But he fails to dwell upon another important point, namely that Gregory (1966) is talking about *reconstitution* of voltage-spikes (from the retina) into percepts – reconstitution, otherwise known as "reconstruction" or "decoding". Nizami (2017) explains these terms and notes that they represent tasks often assigned to one or more anthropomorphized neurons, which effectively, become Homunculi.

9.2 A weak rebuttal from Rorty (1971): "Homunculus is a metaphor"

According to Kenny (1971, p. 65), invoking Homunculus is "a dangerous practice which may lead to conceptual and methodological confusion". Rorty (1971) retorts by dismissing Homunculus as a "metaphor", but fails to clearly define what she means by "metaphor". Baddeley (1998) likewise offers the "metaphor" excuse, in place of "metaphor" employing "*useful* concept [...] [that] does not have to be in any absolute sense "true" or "correct"" (Baddeley, p. 524; original italics and internal quotation marks). But Homunculus is used by hard-nosed math modelers, who are not accustomed to employing "metaphors" (Oram *et al.*, 1998; Westover and O'Sullivan, 2008; Seriès *et al.*, 2009). Indeed, in the highly-cited book *Spikes* (Rieke *et al.*, 1997), the words "homunculus" or "homunculi" appear at least twenty-two times just in the last six pages of their Chapter 1, titled "Central Claims of this Book". Such repeated attention might be lavished upon a "central claim", but seems inappropriate for a metaphor. In response to Rorty (1971), a reader might ask why *any* metaphor would appear in a scientific discourse.

10. Discussion: an alternative to Homunculus: *Emergence*

Regarding Reductionism, von Foerster (2003/2010) differentiates "hard sciences" from "soft sciences" according to whether Reductionism "inevitably leads to success" (hard sciences), or whether (von Foerster, p. 192; original italics and quotation marks):

> Scientists are dealing with essentially nonlinear systems whose salient features are represented by the *interactions* between whatever one may call their "parts" whose properties in isolation add little, if anything, to the understanding of the workings of these systems when each [such system] is taken as a whole.

The latter defines Emergent phenomena – and also the mind. Consider the following.

10.1 An early-contemporary "simple" definition of emergence: Goldstein (1999)

In an oft-cited paper, Goldstein (1999, p. 49) defines "emergence" as "the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems". Goldstein considers "complexity theory", which is a field of mathematics, as central to emergence. In practice, this may not always be the case. It is always the case, however, that "Emergent phenomena are conceptualized as occurring on the macro level, in contrast to the micro-level components and processes out of which they arise" (Goldstein, p. 49). Of course, this definition depends upon what are meant by macro, micro and so on; which, as it transpires, is not always clear. Regardless, Goldstein (p. 50) further characterizes emergent phenomena as "neither predictable nor deducible from lower or micro-level components". Emergent phenomena "appear as integrated wholes that tend to maintain some sense of identity over time" and they occur "at a global or macro level" (Goldstein, p. 50). Further, they "arise as a complex system evolves over time" (Goldstein, p. 50). Finally, they "are recognized by showing themselves", a concept that Goldstein does not pursue.

10.2 A detailed classification of emergence: Fromm (2005)

Goldstein (1999, p. 57) subsequently notes that emergent phenomena are not "reducible to the parts alone". Hence, emergence may offer a non-Reductionist, non-Homuncular account of consciousness. Note well, however, that Goldstein's well-cited paper oversimplifies emergence, even despite what was known at the time. Fromm (2005) provides a broader review. He identifies four forms of "emergence", depending upon how much feedback and what types, are associated with emergent phenomena. Now, feedback is a basic aspect of first-order cybernetics (Wiener, 1948) and Goldstein (1999, p. 55) identifies cybernetics as among the roots of emergence *theory*. But, according to Fromm (2005), feedback is also an active aspect of emergence.

Fromm's (2005) first of four forms of emergence is "simple intentional/nominal", which involves no top-down feedback from the macroscopic level to the microscopic level. Fromm gives examples, which include a mechanical timepiece, in which a measure of time emerges from the movement of gears; and a volume of gas, in which measures called "temperature" and "pressure" emerge from molecular motion. Measures, of course, require an Observer; the Observer's role is discussed further, below.

Note well that even the primitive first level of emergence is important, because we might try to understand the mind through the mind's product, technology. Consider the Antikythera Mechanism, described in the book "Pebbles to Computers" (Blohm *et al.*, 1986). The Antikythera Mechanism is an ancient, corroded-bronze geared device retrieved from a shipwreck, a device believed to have indicated the positions of heavenly bodies. Blohm *et al.* (1986) refer to the Antikythera Mechanism and similar devices, such as mechanical adding machines, as "analog computers". Blohm *et al.* describe various devices, from the simple to the complicated and they illustrate how the latter were designed from the former. But the complicated cannot simply be deconstructed back-down to the simple; too many refining processes (as in the literal refining of silicon for integrated circuits from "pebbles") occur in the build-up stages. The various gears of the Antikythera Mechanism were presumably designed to work only as an integrated whole and hence would have proven useless when separated into gear-groups or individual gear-wheels. Likewise, as noted in Section 3, the integrated whole called "consciousness" might not be divisable into portions that are directed by (or contain) "sub-Homunculi" working as "pieces of" consciousness.

Fromm's second form of emergence (Fromm, p. 9) is weak emergence, which involves:

> Unpredicted patterns, structures or properties – emergent phenomena – which are not directly specified in the interaction laws [which govern the component "entities"] and which in turn influence the low-level interactions of the entities via a feedback process.

This is perhaps the "emergence" meant by Goldstein (1999); it is subject to "top-down feedback" (Fromm, 2005, p. 9), also called "downward causation" (Fromm, p. 9; Goldstein, 1999, p. 62). Downward causation consists of either negative feedback (involved, for example, in the swarming of animals; Fromm, 2005, pp. 10-11), or consists of positive feedback (involved, for example, in stock-market "bubbles"; Fromm, pp. 12-13). Fromm then names a third kind of emergence, one which is more complicated insofar as it can involve short-range positive feedback and long-range negative feedback. For this kind of emergence, Fromm gives the examples of pattern formation in animal coats and the evolution of the Earth's ecosystem.

Finally, Fromm (2005, p.18, original italics) reaches his fourth category of emergence, namely, "strong emergence":

> *Strong Emergence* can be defined as the appearance of emergent structures on higher levels of organization or complexity which possess truly new properties that cannot be reduced, even in principle, to the cumulative effect of the properties and laws of the basic parts and elementary components.

Fromm names just two "strong emergents", namely, life and culture. Remarkably, he omits *consciousness*. In contrast, the philosopher Chalmers (2006) names consciousness as the only sure "strong emergent". Chalmers also distinguishes "strong downward causation", in which "the causal impact of a high-level phenomenon on low-level processes is not deducible even in principle", from "weak downward causation", in which such impact is merely "unexpected" (Chalmers, p. 249).

10.3 Why emergence is a second-order cybernetics concern

Chalmers' latter definition begs the question, of course, of what is meant by "unexpected". And Chalmers (2006) spends several pages inherently toying with that question. As such, he indirectly reveals that "unexpected" is relative to our own understanding of a system, that is, that "unexpected" is *subjective* (Goldstein's "are recognized by showing themselves"). But Chalmers' term "deducible", which he uses to distinguish "strong" downward causation from "weak", seems likewise to be subjective. All told, what phenomena exemplify which kind of "emergence" is relative to the expectations and capabilities of the Observer – which perhaps explains the differences over what phenomena are "strongly emergent". As such, emergence is, in itself, a Second-Order Cybernetics concern.

10.4 Evidence for an emergent mind?

We might ask what self-organization would produce integrated wholes that tend to maintain some sense of identity over time. Such "wholes" include the macro-level (and apparently non-deducible and non-reducible) phenomenon called Consciousness, perhaps exercising downward causation. Thompson and Varela (2001) suggest a candidate for self-organization: ensembles of neurons, neurons whose discharges synchronize locally and globally for a fraction of a second, providing transient linking of widely distributed parts of the brain. "Downward causation" can then, for example, be an influence of intentional mental exercises on neuronal activity (Thompson and Varela, 2001). However, Thompson and Varela (2001, p. 419) emphasize that "the available evidence so far regarding synchronization in the vertebrate brain is only correlative, not causal". Indeed, anything beyond correlation may be undiscoverable. We might take this as a definition of true emergence, so that true emergence represents not merely the uncomprehended but the incomprehensible.

One final note: Winograd and Flores (1987) discuss a stance taken by the philosopher Heidegger, viz.: "The interpreted and the interpreter do not exist independently: existence is interpretation and interpretation is existence" (Winograd and Flores, p. 31). In other words, everything is *interaction* – requiring an existing world to have thoughts and feelings *about*. According to Winograd and Flores (p. 32), Heidegger also claims that our own beliefs cannot be fully comprehended, essentially because we are ensconced in them. The present author would extend the word "beliefs" to that from which they emanate, namely, "brain processes".

11. Conclusions

The imaginary personage called Homunculus resides within the brain, examining arriving neuronal spike-trains to infer the state of the internal and external worlds. But Homunculus, by necessity, is infinitely nested. Attneave (1961) and later Dennett (1981) allegedly remove the infinity through Reduction. In particular, they attempt to limit the number of Homunculi, essentially by trivializing them, by rendering them into "stupider" parts. But there is a mild problem: all such Homunculi still "behave", that is, they perform things done by the whole human – even when those whole humans "are no more than adders and subtractors" or when they merely "pick the larger of two numbers" (Dennett, 1981, p. 80). Behavior is behavior and thereby Attneave (1961) and Dennett (1981) fail to eliminate Homunculi.

What was the underlying thinking? Nagel (1974) notes (albeit in a very convoluted fashion) that we expect all subjective experience (e.g. perceptions) to correspond to observable physical phenomena. Hence, there arose "the effort to substitute an objective concept of mind for the real thing, in order to have nothing left over which cannot be reduced" (Nagel, p. 445). In other words, to reduce the "intentional" to the "non-intentional". But, as Nagel (p. 445) cautions:

> If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done.

Nagel's words could have been written yesterday. To actually remove the infinity of Homunculi, the Homunculi need to be rendered irrelevant. Or as von Foerster (2003/2010, p. 103) puts it, we need to "dismiss the demon [Homunculus] and put his brain right there where ours is". Historically, the "little idiots" of Attneave and of Dennett were either neurons or "controllers" of neurons, which performed tasks – including even automatic unconscious activities, such as the regulation of breathing – tasks that required "discriminating" (as in Signal Detection Theory; Green and Swets, 1966/1988), "directing", "interpreting", "thinking", "examining", "comparing", "computing" and so on (Nizami, 2017). But these words describe the behaviors of the whole conscious human being, not those of the individual neurons (Nizami, 2017). If we regard "discriminating", "directing", "interpreting", "thinking", "examining", "comparing", "computing" and so on as capabilities that *emerge* from the activity of the enormous number of neurons in the brain, then Homunculi finally become irrelevant.

Meanwhile, Homunculi continue to lurk in scientists' explanations of cognition. To remove Homunculi once and for all, the culture of Cognitive Science itself needs to change. In particular, Reductionism's limited scope needs to be recognized. That limitation was well-understood by Dr David Suzuki over 30 years ago: "Our knowledge of how single neurons work in the brain is of little or no help in treating a psychotic person or coping with loneliness" (Suzuki, in Blohm *et al.*, 1986, p. 10). And yet, there has been a recent slew of miraculous Reductionist discoveries of circuits for behaviors (Chou *et al.*, 2016; Hein *et al.*, 2016; Kondoh *et al.*, 2016; Tovote *et al.*, 2016; Zalocusky *et al.*, 2016), including the identification by Matthews *et al.* (2016) of the circuit for loneliness! But is this alleged loneliness circuit of any clinical use? For that matter, should Matthews *et al.*, or anyone else making such claims, be trusted? After all, as Dr Suzuki notes (in Blohm *et al.*, 1986, p. 11), "By looking at nature in bits and pieces, our understanding of it can only be fragmentary, for nature is not the sum of its isolated parts". This is Nature as an Emergent phenomenon. And certainly, any brain "circuit" is only one small part of the brain. We can only imagine the time and money that has been wasted on the Reductionist search for neuronal Homunculi.

References

Atick, J.J. and Redlich, A.N. (1992), "What does the retina know about natural scenes?", *Neural Computation*, Vol. 4 No. 2, pp. 196-210.

Attneave, F. (1961), "In defense of homunculi", in Rosenblith, W.A. (Ed.), *Sensory Communication*, MIT Press, Cambridge, MA, pp. 777-782.

Attneave, F. (1954), "Some informational aspects of visual perception", *Psychological Review*, Vol. 61 No. 3, pp. 183-193.

Baddeley, A. (1998), "The central executive: a concept and some misconceptions", *Journal of the International Neuropsychological Society*, Vol. 4 No. 5, pp. 523-526.

Baddeley, A. (2012), "Working memory: theories, models and controversies", *Annual Review of Psychology*, Vol. 63, pp. 1-29.

Baddeley, A.D. and Hitch, G. (1974), "Working memory", in Bower, G.A. (Ed.), *The Psychology of Learning and Motivation*, Academic Press, New York, NY, Vol. 8, pp. 47-89.

Baden, T., Berens, P., Franke, K., Rosón, M.R., Bethge, M. and Euler, T. (2016), "The functional diversity of retinal ganglion cells in the mouse", *Nature*, Vol. 529 No. 7586, pp. 345-350.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R.R. and Warland, D. (1991), "Reading a neural code", *Science*, Vol. 252 No. 5014, pp. 1854-1857.

Blohm, H., Beer, S. and Suzuki, D. (1986), *Pebbles to Computers: The Thread*, Oxford University Press, Toronto.

Boone, W. and Piccinini, G. (2016), "The cognitive neuroscience revolution", *Synthese*, Vol. 193 No. 5, pp. 1509-1534.

Bullock, T.H. (1961), "The problem of recognition in an analyzer made of neurons", in Rosenblith, W.A. (Ed.), *Sensory Communication*, MIT Press, Cambridge, MA, pp. 717-724.

Chacron, M.J., Longtin, A. and Maler, L. (2003), "The effects of spontaneous activity, background noise and the stimulus ensemble on information transfer in neurons", *Network*, Vol. 14 No. 4, pp. 803-824.

Chalmers, D.J. (2006), "Strong and weak emergence", in Clayton, P. and Davies, P. (Eds), *The Re-Emergence of Emergence*, Oxford University Press, New York, NY, pp. 244-254.

Chechik, G., Anderson, M.J., Bar-Yosef, O., Young, E.D., Tishby, N. and Nelken, I. (2006), "Reduction of information redundancy in the ascending auditory pathway", *Neuron*, Vol. 51 No. 3, pp. 359-368.

Chou, M.-Y., Amo, R., Kinoshita, M., Cherng, B.-W., Shimazaki, H., Agetsuma, M., Shiraki, T., Aoki, T., Takahoko, M., Yamazaki, M., Higashijima, S.-I. and Okamoto, H. (2016), "Social conflict resolution regulated by two dorsal habenular subregions in zebrafish", *Science*, Vol. 352 No. 6281, pp. 87-90.

Ciocchi, S., Passecker, J., Malagon-Vina, H., Mikus, N. and Klausberger, T. (2015), "Selective information routing by ventral hippocampal CA1 projection neurons", *Science*, Vol. 348 No. 6234, pp. 560-563.

Crick, F. and Koch, C. (2000), "The unconscious homunculus", *Neuropsychoanalysis*, Vol. 2 No. 1, pp. 3-11.

Crick, F. and Koch, C. (2003), "A framework for consciousness", *Nature Neuroscience*, Vol. 6 No. 2, pp. 119-126.

Crick, F.C. and Koch, C. (2005), "What is the function of the claustrum?", *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 360 No. 1458, pp. 1271-1279.

Dalla Barba, G. (2001), "Beyond the memory-trace paradox and the fallacy of the homunculus", *Journal of Consciousness Studies*, Vol. 8, pp. 51-78.

Dennett, D.C. (1981), *Brainstorms: Philosophical Essays on Mind and Psychology*, MIT Press, Cambridge, MA.

Desban, L. and Wyart, C. (2016), "A brain conditioned for social defeat", *Science*, Vol. 352 No. 6281, pp. 42-43.

Drayson, Z. (2012), "The uses and abuses of the personal/subpersonal distinction", *Philosophical Perspectives*, Vol. 26 No. 1, pp. 1-18.

Ehret, G. and Romand, R. (1997), *The Central Auditory System*, Oxford University Press, New York, NY.

Feinberg, T.E. (2011), "The nested neural hierarchy and the self", *Consciousness and Cognition*, Vol. 20 No. 1, pp. 4-15.

Fromm, J. (2005), "Types and forms of emergence", *Complexity Digest*, Vol. 25, pp. 1-23.

Goldstein, J. (1999), "Emergence as a construct: history and issues", *Emergence*, Vol. 1 No. 1, pp. 49-72.

Green, D.M. and Swets, J.A. (1966/1988), *Signal Detection Theory and Psychophysics*, Peninsula Publishing, Los Altos, CA.

Gregory, R.L. (1966), *The Eye and the Brain*, McGraw-Hill, New York, NY.

Harth, E. (1993), *The Creative Loop: How the Brain Makes a Mind*, Addison-Wesley, New York, NY.

Hazy, T.E., Frank, M.J. and O'Reilly, R.C. (2007), "Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system", *Philosophical Transactions of the Royal Society B*, Vol. 362 No. 1485, pp. 1601-1613.

Hein, G., Morishima, Y., Leiberg, S., Sul, S. and Fehr, E. (2016), "The brain's functional network architecture reveals human motives", *Science*, Vol. 351 No. 6277, pp. 1074-1078.

Hommel, B., Daum, I. and Kluwe, R.H. (2004), "Exorcising the homunculus, phase two: editors' introduction", *Acta Psychologica*, Vol. 115 Nos 2/3, pp. 99-104.

Kenny, A.J.P. (1971), "The homunculus fallacy", in Grene, M. (Ed.), *Interpretations of Life and Mind: Essays around the Problem of Reduction*, Humanities Press, New York, NY, pp. 65-74.

Kondoh, K., Lu, Z., Ye, X., Olson, D.P., Lowell, B.B. and Buck, L.B. (2016), "A specific area of olfactory cortex involved in stress hormone responses to predator odours", *Nature*, Vol. 532 No. 7597, pp. 103-106.

Lörincz, A. (2009), "Hebbian constraint on the resolution of the homunculus fallacy leads to a network that searches for hidden cause-effect relationships", *Proceedings of the 2nd Conference on Artificial General Intelligence, AGI 2009, Arlington, VA*, pp. 126-131.

Lörincz, A., Póczos, B., Szirtes, G. and Takács, B. (2002), "Ockham's razor at work: modeling of the "homunculus"", *Brain and Mind*, Vol. 3 No. 2, pp. 187-220.

McMullen, T. (2001), "Getting rid of the Homunculus: a direct realist approach", in Morss, J.R., Stephenson, N. and van Rappard, H. (Eds), *Theoretical Issues in Psychology*, Springer Science+Business Media, New York, NY, pp. 159-167.

MacKay, D.M. and McCulloch, W.S. (1952), "The limiting information capacity of a neuronal link", *The Bulletin of Mathematical Biophysics*, Vol. 14 No. 2, pp. 127-135.

Margolis, J. (1980), "The trouble with homunculus theories", *Philosophy of Science*, Vol. 47 No. 2, pp. 244-259.

Matthews, G.A., Nieh, E.H., Vander Weele, C.M., Halbert, S.A., Pradhan, R.V., Yosafat, A.S., Glober, G.F., Izadmehr, E.M., Thomas, R.E., Lacy, G.D., Wildes, C.P., Ungless, M.A. and Tye, K.M. (2016), "Dorsal raphe dopamine neurons represent the experience of social isolation", *Cell*, Vol. 164 No. 4, pp. 617-631.

Maturana, H.R. and Varela, F.J. (1987/1998), *The Tree of Knowledge*, Shambhala Publications, Boston, MA.

Monsell, S. and Driver, J. (2000), "Banishing the control homunculus", in Monsell, S. and Driver, J. (Eds), *Control of Cognitive Processes: Attention and Performance XVIII*, MIT Press, Cambridge, MA, pp. 3-32.

Mumford, S. (2003), *Dispositions*, Oxford University Press, New York, NY.

Nagel, T. (1974), "What is it like to be a bat?", *The Philosophical Review*, Vol. 83 No. 4, pp. 435-450.

Naka, K. and Sakai, H.M. (1993), "Can we ever understand how the retina communicates with the brain?", in Shapley, R. and Lam, D.M.-K. (Eds), *Contrast Sensitivity*, MIT Press, Cambridge, MA, pp. 75-81.

Newell, A. (1980), "Reasoning, problem-solving and decision processes", in Nickerson, R. (Ed.), *Attention and Performance VIII*, Erlbaum, Hillsdale, NJ, pp. 693-718, available at: http://repository.cmu.edu/compsci/2482/

Nizami, L. (2009a), "Sensory systems as cybernetic systems that require awareness of alternatives to interact with the world: analysis of the brain-receptor loop in Norwich's Entropy Theory of Perception", *Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics*, *San Antonio, TX*, pp. 3477-3482.

Nizami, L. (2009b), "A lesson in the limitations of applying cybernetics to sensory systems: hidden drawbacks in Norwich's model of transmitted Shannon information", *IIAS-Transactions on Systems Research and Cybernetics (International Journal of the International Institute for Advanced Studies in Systems Research and Cybernetics)*, Vol. 9, pp. 1-9.

Nizami, L. (2010), "Interpretation of absolute judgments using Information Theory: channel capacity or memory capacity?", *Cybernetics and Human Knowing*, Vol. 17, pp. 111-155.

Nizami, L. (2011), "Norwich's Entropy Theory: how not to go from abstract to actual", *Kybernetes*, Vol. 40 Nos 7/8, pp. 1102-1118.

Nizami, L. (2012a), "Confusing the "confusion matrix": the misapplication of Shannon Information Theory in sensory psychology", *Acta Systemica (International Journal of the International Institute for Advanced Studies in Systems Research and Cybernetics)*, Vol. 12, pp. 1-17.

Nizami, L. (2012b), "Loudness "fatigue" with two ears but not with one: simultaneous dichotic loudness balance (SDLB) explained", in Leth-Steensen, C. and Schoenherr, J.R. (Eds), *Fechner Day 2012: Proceedings of the 28th Annual Meeting of the International Society for Psychophysics*, *Ottawa*, pp. 6-11.

Nizami, L. (2013), "Paradigm versus praxis: why psychology "absolute identification' experiments do not reveal sensory processes", *Kybernetes*, Vol. 42, pp. 1447-1456.

Nizami, L. (2014a), "Information theory's failure in neuroscience: on the limitations of cybernetics", Paper No. 158, *Proceedings of the IEEE 2014 Conference on Norbert Wiener in the 21st Century, IEEE*, *Piscataway, NJ*.

Nizami, L. (2014b), "Why "information transmitted" tells us nothing about cognition", *Cognition: Abstracts of Papers Presented at the LXXIX Cold Spring Harbor Symposium on Quantitative Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 106.

Nizami, L. (2015), "Homunculus strides again: why "information transmitted" in neuroscience tells us nothing", *Kybernetes*, Vol. 44 Nos 8/9, pp. 1358-1370.

Nizami, L. (2016), "Attneave was wrong: the infinity of Homunculi", *Abstract 83, p87, TSC 2016: The Science of Consciousness*, Tucson, AZ, available at: www.consciousness.arizona.edu/documents/TSC2016_BOOK_of_Abstracts_final.pdf

Nizami, L. (2017), "I, NEURON: the neuron *as* the collective", in press, *Kybernetes*.

Optican, L.M. and Richmond, B.J. (1987), "Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis", *Journal of Neurophysiology*, Vol. 57, pp. 162-178.

Oram, M.W., Földiák, P., Perrett, D.I. and Sengpiel, F. (1998), "The "ideal homunculus": decoding neural population signals", *TINS*, Vol. 21 No. 6, pp. 259-265.

Passaglia, C.L. and Troy, J.B. (2004), "Information transmission rates of cat retinal ganglion cells", *Journal of Neurophysiology*, Vol. 91 No. 3, pp. 1217-1229.

Posner, M.I. and DiGirolamo, G.J. (2000), "Executive attention: conflict, target detection and cognitive control", in Parasuraman, R. (Ed.), *The Attentive Brain*, MIT Press, Cambridge, MA, pp. 401-423.

Reardon, S. (2017), "Giant neuron encircles entire brain of a mouse", *Nature*, Vol. 543 No. 7643, pp. 14-15.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. and Bialek, W. (1997), *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA.

Rorty, A. (1971), "Not every homunculus spoils the argument", in Grene, M. (Ed.), *Interpretations of Life and Mind: Essays Around the Problem of Reduction*, Humanities Press, New York, NY, pp. 75-80.

Rose, D. (2012), "Three dimensions of psychosemantics", *Gestalt Theory*, Vol. 34, pp. 259-286.

Sanders, A.F. (1998), *Elements of Human Performance: Reaction Processes and Attention in Human Skill*, Lawrence Erlbaum, Mahwah, NJ.

Schinkel, M.P. (2006), "Disequilibrium dynamics and aggregate excess demand: on a Homunculus fallacy in economic theory", *History of Political Economy*, Vol. 38 No. 1, pp. 189-212.

Searle, J.R. (1994), *The Rediscovery of the Mind*, MIT Press, Cambridge, MA.

Seriès, P., Stocker, A.A. and Simoncelli, E.P. (2009), "Is the homunculus "aware" of sensory adaptation?", *Neural Computation*, Vol. 21 No. 12, pp. 3271-3304.

Sober, E. (1982), "Why must homunculi be so stupid?", *Mind*, Vol. 91 No. 363, pp. 420-422.

Stein, R.B. (1967), "The information capacity of nerve cells using a frequency code", *Biophysical Journal*, Vol. 7 No. 6, pp. 797-826.

Thompson, E. and Varela, F.J. (2001), "Radical embodiment: neural dynamics and consciousness", *Trends in Cognitive Sciences*, Vol. 5 No. 10, pp. 418-425.

Tibbetts, P. (1995), "Neurobiology and the Homunculus thesis", *Man and World*, Vol. 28 No. 4, pp. 401-413.

Tovote, P., Soledad Esposito, M., Botta, P., Chaudun, F., Fadok, J.P., Markovic, M., Wolff, S.B.E., Ramakrishnan, C., Fenno, L., Deisseroth, K., Herry, C., Arber, S. and Lüthi, A. (2016), "Midbrain circuits for defensive behaviour", *Nature*, Vol. 534, pp. 206-212.

Verbruggen, F., McLaren, I.P.L. and Chambers, C.D. (2014), "Banishing the control homunculi in studies of action control and behavior change", *Perspectives on Psychological Science*, Vol. 9 No. 5, pp. 497-524.

Victor, J.D. (2006), "Approaches to information-theoretic analysis of neural activity", *Biological Theory*, Vol. 1 No. 3, pp. 302-316.

von Foerster, H. (2003/2010), *Understanding Understanding: Essays on Cybernetics and Cognition*, Springer-Verlag, New York, NY.

Walløe, L. (1970), "On the transmission of information through sensory neurons", *Biophysical Journal*, Vol. 10 No. 8, pp. 745-763.

Westover, M.B. and O'Sullivan, J.A. (2008), "Achievable rates for pattern recognition", *IEEE Transactions on Information Theory*, Vol. 54 No. 1, pp. 299-320.

Whitfield, I.C. (1984), *Neurocommunications: An Introduction*, John Wiley and Sons, New York, NY.

Wiener, N. (1948), *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, MA.

Winograd, T. and Flores, F. (1987), *Understanding Computers and Cognition*, Addison-Wesley, San Francisco, CA.

Woodman, G.F. and Schroeder, C.E. (2012), "Using nonhuman primates to study the micro- and macrodynamics of neural mechanisms of attention", in Posner, M.I. (Ed.), *Cognitive Neuroscience of Attention*, The Guilford Press, New York, NY, pp. 219-228.

Zalocusky, K.A., Ramakrishnan, C., Lerner, T.N., Davidson, T.J., Knutson, B. and Deisseroth, K. (2016), "Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making", *Nature*, Vol. 531 No. 7596, pp. 642-646.

Zeki, S. (1993), *A Vision of the Brain*, Blackwell Scientific Publications, Cambridge, MA.

Corresponding author
Lance Nizami can be contacted at: nizamii2@att.net