ARTICLE

WILEY

Evolving Concepts of Functional Localization

Joseph B. McCaffrey

Department of Philosophy, Neuroscience
Program, University of Nebraska Omaha,
Omaha, Nebraska, USA

Correspondence
Joseph B. McCaffrey.
Email: jmccaffrey@unomaha.edu

Abstract

Functional localization is a central aim of cognitive neuro-science. But the nature and extent of functional localization in the human brain have been subjects of fierce theoretical debate since the 19th Century. In this essay, I first examine how concepts of functional localization have changed over time. I then analyze contemporary challenges to functional localization drawing from research on neural reuse, neural degeneracy, and the context-dependence of neural functions. I explore the consequences of these challenges for topics in philosophy of science and philosophy of mind including localizationist versus anti-localizationist approaches to cognitive neuroscience, multiple realizability, reverse inference in functional neuroimaging, and the modularity of mind.

An adequate understanding of conceptual change requires the recognition that scientific concepts are tools for research, as much as glassware, microscopes, and automated sequencers.

–Paul Griffiths and Karola Stotz (2008, 11)

1 | THE PROBLEM OF FUNCTIONAL LOCALIZATION

Functional localization is one of neuroscience's central goals—"advancing our understanding of neuroscience centrally depends on characterizing how structure and function are related in the brain" (Anderson et al., 2013, p. 50). Much of human cognitive neuroscience is geared toward characterizing the functions of brain areas and/or mapping cognitive functions onto the brain. Hence influential articles propose that the right temporoparietal junction (RPTJ) is specialized for theory of mind, while working memory localizes to the dorsolateral prefrontal cortex (DLPFC),

etc. (Kanwisher, 2010). These putative mappings are used to investigate the neural basis of cognitive functions and may provide a step toward explaining those functions mechanistically and/or via cognitive models (Bechtel & Richardson, 2010).[1]

Due to the centrality of functional localization to cognitive neuroscience, there have been fierce debates about the prospects of localization from the 19th Century to the present (Mundale, 2002). Core questions in these debates include: Do cognitive functions (e.g., working memory) have definite seats in the brain? Is the cerebral cortex, a continuous ~2.5 mm thick structure with architectural similarities (e.g., its six layered structure) and differences, divided into functionally specialized regions (Kanwisher, 2010)? Are mental kinds such as "fear" or "pain" multiply realized in the human brain (Figdor, 2010)? Do brain areas have different functions at different times (Burnston, 2016b)? Do individuals vary in terms of structure-function mappings (De Brigard, 2017; Viola, 2021; Ward, 2022)? Is functional localization possible in human brains and other complex biological systems (Silberstein, 2021)? If so, how do neural structure and mental function relate (Anderson, 2014)?

Cognitive neuroscience traditionally presupposes that cognitive functions have stable neural realizers, that brain areas are specialized for particular cognitive functions, and that there is little variability in the brain's functional topography between healthy adults. This "Neo-Phrenological Framework of Cognitive Neuroscience," (Viola, 2017, p. 165) proposes one-to-one mappings between neural structures and mental functions. In a system so constituted, one could predict what deficits will arise from structural lesions or infer functions from activation patterns measured by neuroimaging techniques such as functional magnetic resonance imaging (fMRI).

Recent work in philosophy and cognitive neuroscience challenges this conventional framework (Anderson, 2014). Brain areas are frequently claimed to have many cognitive functions rather than one (Anderson, 2010; Burnston, 2016b; Klein, 2012; McCaffrey, 2015; Price & Friston, 2005). Furthermore, different areas are sometimes thought to implement the same function in different circumstances (Figdor, 2010; Price & Friston, 2002). There is also significant variability reported both within (Seghier & Price, 2018) and between (De Brigard, 2017; Viola, 2021) individuals in terms of structure-function mappings.

In response to these challenges, some theorists advocate preserving localization by focusing on large-scale brain networks rather than individual regions (Pessoa, 2014; Sporns, 2014) or by revising our taxonomy of mental functions to achieve better mappings (Poldrack, 2010; Price & Friston, 2005; Rathkopf, 2013).[2] Others call for developing new concepts of localization, or even shifting away from localizationism as the dominant paradigm in cognitive neuroscience (Anderson, 2014; Hutto et al., 2017; Klein, 2012; Pessoa, 2014, 2022; Silberstein, 2021; Viola, 2017; Westlin et al., 2023; Zerilli, 2021). After reviewing the complexities of contemporary brain mapping, neuroscientist and philosopher Michael Anderson writes: "I myself have become pessimistic that it will *generally* be possible to identify the specific neural mechanisms…to which we can assign cognitive operations" (2014, 81). Thus, our concepts of functional localization are in flux, or perhaps peril.

In this essay, I examine philosophical debates about functional localization in cognitive neuroscience. First, I analyze the concept of functional localization by briefly reviewing the history of debates between localizationists and anti-localizationists from the 19th Century to the present (Section 2). Then I examine research on three topics—neural reuse, neural degeneracy, and contextualism—that challenges a standard localizationist picture (Section 3, Section 4, Section 5) and discuss the philosophical implications of this research. I conclude by reflecting on how our concepts of functional localization have changed and whether localizationism will remain the guiding paradigm for cognitive neuroscience (Section 6).

2 | LAYING THE FOUNDATIONS: LOCALIZATIONISM AND ANTI-LOCALIZATIONISM

The history of neuroscience is rife with debates about the nature of functional localization (Mundale, 2002). An early proponent of cerebral localization was the neuroanatomist and phrenologist Franz Joseph Gall (1758-1828). Gall's "organology" held that the cortex is a collection of "cerebral organs," each of which carries out a specific mental

faculty (e.g., memory of words) (1835).[3] His opponent Marie Jean Pierre Flourens (1794-1867), an equipotentialist, instead claimed that the cortex is an undifferentiated mass of tissue that participates equally in its various faculties (1824).[4] Gall's view is an early form of **Localizationism** in which mental faculties have definite "seats" and the cortex is conceived as a mosaic of functionally distinct regions. Flourens' view is a form of **Anti-Localizationism**, which denies cortical specialization and the localization of mental function.

During the 19th Century, findings such as Paul Broca's (1824-1880) identification of an inferior frontal lobe region (now called Broca's area) where damage selectively impairs speech production (1861) swung the pendulum toward localizationism.[5] However, the localization issue was far from settled. First, there were debates about what localizationism entailed, with many adopting a middle ground between the Gall's organology and Flourens' equipotentialism (Bergeron, 2007; Gamboa, 2020; Young, 1968). Carl Wernicke (1848-1905) believed that basic sensory and motor functions localized to discrete structures while "higher" psychological functions did not localize (Bergeron, 2007). Friedrich Goltz (1834-1902) denied localizationism without embracing Flourens' equipotentialism (Gamboa, 2020). Second, despite the widespread acceptance of localizationism, radical anti-localizationist views kept emerging.

In the 20th Century, neuroscientist Karl Lashley (1890-1958) articulated an anti-localizationist theory (1929) akin to that of Flourens. Lashley found that in lesion studies with rats navigating mazes, the amount of cortex removed rather than the site of the lesion determined how much performance was impaired. This led him, like Flourens, to endorse the principles of equipotentiality (parts of the cortex are functionally interchangeable) and mass action (the cortex acts as a whole unit in learning and memory) (1929). But by the mid-20th Century, neuropsychology studies showed many selective structure-function mappings. For example, Scoville and Milner (1957) reported that bilateral resection of the hippocampus in the famous patient H.M. caused selective deficits in "recent memory"—that is, the formation of new episodic memories. Such findings solidified localizationism as the dominant research paradigm in the mind-brain sciences (Figdor, 2010).

In the 1980s, contemporary cognitive neuroscience was launched by studies using positron emission tomography (PET) to map distinct language functions (e.g., speech production, phonology, semantics) onto cortical regions (Petersen et al., 1988). This led to a period of exuberant localization in which neuroimagers assigned functions such as face recognition, visual word form recognition, semantic memory, etc. to specific cortical structures using techniques such as PET and fMRI (see Kanwisher, 2010 for a review). Some have suggested that, while the precise areas and functions are different, the overall picture is essentially Gall's (Anderson, 2014; Poldrack, 2010).

Yet the localization debate continues to rage. Some researchers, inspired by connectionist and network approaches to neuroscience, advocate mapping functions onto large-scale brain networks and criticize the field's focus on isolated regions (McIntosh, 2000; Pessoa, 2014; Sporns, 2014). By "region" I mean cortical structures of roughly the size and composition identified in Korbinian Brodmann's (1868-1968) cytoarchitectural maps, though the precise taxonomy is under significant revision (Foit et al., 2022; Loukas et al., 2011). Others argue that structure-function mappings are so complex that it is time to reexamine, or perhaps discard, the localizationist framework (Anderson, 2014; Burnston, 2020; Frisch, 2014; Klein, 2012; Pessoa, 2014, 2022; Silberstein, 2021; Silberstein & Chemero, 2013; Uttal, 2001; Viola, 2017; Westlin et al., 2023). For example, Anderson (2014) argues that since brain regions have "dispositional functions" and the brain flexibly recruits neural resources to perform tasks, neuroscience should move past localizationism as a guiding framework (see Section 6). Pessoa (2022) similarly claims that localization fails because the brain networks supporting cognitive functions are interactionally complex and are not functionally decomposable (see Boone et al., 2022 for a critical reply). Must cognitive neuroscience abandon localization to move forward? This perennial issue occupies the forefront of theoretical neuroscience once more.

One reason that localization debates are so intractable is that researchers neglect to analyze the concept of functional localization in the first place (Mundale, 2002). Functional localization is, like the concept of the cortical column (Haueis, 2016, 2021), the receptive field (Chirimuuta & Gold, 2009; Venturelli, 2021), and hierarchy (Burnston & Haueis, 2021), one of the central concepts organizing neuroscience research (Gold & Roskies, 2008; Mundale, 2002).

Yet it is often unclear what conceptions of localization those attacking or defending it hold (Mundale, 2002). Getting clear about the concept of functional localization is vital for making progress in these longstanding debates.[6]

Mundale analyzes functional localization as the concept that the brain has "functionally distinct, physically discrete" areas (2002, 314). In contrast, she distinguishes two forms of anti-localization held by theorists such as Lashley and Flourens. *Equipotentialism* alleges that the cortex is "functionally equivalent throughout" (Mundale, 2002, p. 315)— that is, all cortical tissue is capable of performing any of the cortex's functions. *Holism* proposes that "to understand how the brain subserves a given function, it is necessary to understand the whole brain (or at least large portions of it) acting as an integrated functional unit" (Ibid)—that is, all or most of the brain participates in any function. [7] But while these understandings of localizationism and anti-localizationism are faithful to the historical debates outlined above, they are inadequate for understanding the landscape of contemporary localization debates.

For one, virtually no anti-localizationists are committed to the radical theses of holism and equipotentialism. Even radical contemporary anti-localizationists (e.g., Anderson, 2014; Hutto et al., 2017; Pessoa, 2022) explicitly disavow equipotentialism and holism while insisting that localization has failed (though I later question whether their views entail holism). Merely denying equipotentialism and holism leaves open a vast space of possibilities that plausibly challenge localizationism as we know it. For example, if a typical brain region performs up to ten cognitive functions, but each region's functional profile is distinct (e.g., Anderson, 2010; Anderson et al., 2013), does this mean localization has failed? The analysis of localization offered above—that the brain has functionally distinct areas—is too vague to adjudicate such questions. So what does affirming or denying localizationism amount to today?

I argue the concept of functional localization operating in contemporary cognitive neuroscience has shifted from the simple conceptualizations above to a more nuanced set of commitments about structure-function mapping. The idea that contemporary localizationism entails a "one-to-one mapping" between neural structures and cognitive functions (e.g., Viola, 2017), is roughly right, but obscures some important subtleties. In my view, contemporary localizationism consists of three tenets: (1) **Structural Specialization** (brain structures are specialized for particular cognitive functions), (2) **Localization of Function** (cognitive functions stably localize to a specific neural structure), and (3) **Intrinsicality** (the functions of neural structures can be characterized or understood in relative isolation). I derived this conceptual analysis by looking at alleged challenges to localizationist models and observing that the specific target is not always the same (Sections 3-5).[8] These tenets serve as background assumptions in brain mapping research. In paradigm cases of localization, all of them are presumed to hold.

Consider Nancy Kanwisher's work on face recognition and the fusiform gyrus. Kanwisher et al. (1997) report that the so-called "fusiform face area" (FFA) is selectively activated by faces in fMRI experiments and that damage to the FFA results in prosopagnosia (face blindness) while sparing visual object recognition more generally. Thus, the FFA is considered to be specialized for faces (*Structural Specialization*). Further research implies that face recognition maps onto a broader network of areas (*Localization of Function*) and theorizes about the specific computation the FFA performs in encoding face identity (*Intrinsicality*) (Kanwisher, 2010; Kanwisher & Yovel, 2006).

Neuroscientists routinely dispute specific localization claims such as "the FFA is specialized for face processing." Gauthier et al. (2000) claim that the FFA is not specialized for faces but instead serves visual expertise more generally (e.g., it is activated when bird experts view birds and car experts view cars). But many philosophers and neuroscientists now question whether the tenets of localizationism *generally* hold (Anderson, 2010; Bergeron, 2007; Burnston, 2016a, 2016b; De Wit & Matheson, 2022; Figdor, 2010; Klein, 2012; Pessoa, 2014; Silberstein & Chemero, 2013). In this context, anti-localizationism is the claim that the brain's functional architecture *systematically* violates one or more of the core tenets of localizationism. These claims, if true, undermine localizationism as a guiding paradigm for cognitive neuroscience. Jointly, they may even signify a return to more radical forms of anti-localizationism (Anderson, 2014; Hutto et al., 2017; Pessoa, 2022).

It is to this research we now turn. In Section 3, I examine evidence of **Neural Reuse**—the idea that cortical regions are typically recruited for multiple cognitive functions and hence rarely specialized (Anderson, 2010). In Section 4, I examine evidence of **Neural Degeneracy**—the idea that different brain regions can perform the same cognitive function under different circumstances (Price & Friston, 2002). In Section 5, I examine evidence of **Contextualism**—the

claim that the functions of neural structures critically depend on the broader "neural context" in which the structure is embedded (Burnston, 2016b; Klein, 2012).

3 | NEURAL REUSE AND THE MASSIVE RE-DEPLOYMENT HYPOTHESIS

3.1 | Findings

Structural Specialization holds that neural structures—e.g., cortical regions or networks—are specialized for particular cognitive functions (e.g., speech production). Anderson's (2007, 2010) influential theory of neural reuse, called the "massive redeployment hypothesis," posits instead that brain regions are typically multi-functional.[9] Neural reuse occurs when "neural circuits are… exapted (exploited, recycled, redeployed) during evolution or normal development, and put to different uses, often without losing their original functions" (Anderson, 2010, p. 246). Some examples of neural reuse include: (1) FFA's recruitment for cars, birds, and other visual stimuli (e.g., chess configurations) in addition to faces (Gauthier et al., 2000), (2) the involvement of Broca's area's in non-speech functions such as motor imitation and musical syntax (Viola & Zanin, 2017), and (3) the anterior cingulate cortex's association with error monitoring, decision making, cognitive control, and motivation (Holroyd & Yeung, 2012).

The term "neural reuse" sometimes refers to the phenomenon of multi-functionality in which a cortical region participates in multiple cognitive functions (e.g., working memory, response inhibition, etc.) or domains (e.g., vision, memory, mathematics, etc.). This is called a "one-to-many" mapping from regions to functions (Viola, 2017). Elsewhere, "neural reuse" is shorthand for theories explaining the phenomenon of multi-functionality, including Dehaene's neuronal recycling theory (2005) and Anderson's (2007, 2010) massive re-deployment hypothesis (see Barack, 2019 for a philosophical discussion). Dehaene's neuronal recycling theory emphasizes the recruitment of brain areas for novel functions during development (e.g., visual areas being repurposed for reading). Anderson's massive redeployment hypothesis emphasizes changes over evolutionary time scales—that is, cortical regions evolved for one function and were subsequently exapted for other functions during our phylogeny.[10] The massive re-deployment hypothesis makes three predictions: (1) brain regions will be recruited for numerous cognitive functions across different cognitive domains instead of being functionally specialized, (2) all things being equal, older regions will be recruited for a broader range of functions than newer ones, and (3) older cognitive functions will be more localizable than newer ones, since there is more existing circuitry for newer functions to be cobbled from.

Anderson (2010) reports support for all three predictions in meta-analyses of neuroimaging studies. Regarding (1), Anderson reports that in a meta-analysis of nearly 1,500 fMRI studies, brain regions are recruited for tasks across an average of 9 cognitive domains (language, memory, vision, etc.).[11] Regarding (2), meta-analyses show that posterior brain regions (a rough proxy for phylogenetically older areas) are re-deployed across more domains. Regarding (3), these meta-analyses also suggest that phylogenetically newer functions such as language are more distributed (and hence less localized) than phylogenetically older ones such as attention and visual perception. These findings, which have been vindicated by subsequent research (see Zerilli, 2021, Ch. 3), suggest that neural reuse is typical for brain structures whereas specialization appears rare or non-existent.

The main empirical worry for neural reuse is that perhaps regions only appear multi-functional because the fairly large regions used in fMRI meta-analyses are actually constellations of smaller functional sub-regions. This "sub-divide and conquer" strategy of dividing multi-functional regions into smaller functionally-specialized regions (McCaffrey, 2015) may explain away some cases of reuse. For instance, Fedorenko and Blank (2020) claim Broca's area contains neural sub-populations, some of which are language-specific and others that are more domain general. But neural reuse has been observed both at smaller scales within the human brain (e.g., meta-analyses with smaller regions of interest) and in neural circuits at different scales in invertebrate and vertebrate models (Briggman & Kristan, 2008, Anderson, 2010, Anderson, 2014, Ch. 1, Zerilli, 2021, Ch. 3). Thus, reuse may be a general feature of neural organization at multiple scales.

3.2 | Implications

Neural reuse raises important questions about the nature of neural functions (Bergeron, 2007; Garson, 2011; McCaffrey, 2015; Rathkopf, 2013) and the practices surrounding functional attribution in cognitive neuroscience (Poldrack, 2006, 2010; Price & Friston, 2005). What does it mean to claim that a brain area has a particular function? How should we characterize the functions of brain areas if reuse is the norm? Neuroscientists often assign imprecise functions to brain regions using tasks that recruit them—e.g., the "fusiform face area" or the "visual word form area" (Price & Friston, 2005). Recently, philosophers have analyzed functional attribution in neuroscience more precisely.

Craver (2001) adopts a *causal role theory* (Cummins, 1975) of neural function, arguing that a region's function is the mechanistic role (e.g., converting phonology to semantics) that it plays in carrying out a broader cognitive capacity (e.g., reading). Garson (2011) adopts an *etiological theory* (Neander, 1991) of neural function in which a region's function is what it was selected for (either by natural selection or by Darwinian developmental processes such as neural pruning). According to Garson, "detecting orthographic regularities" is the visual word form area's function because it was selected for that during development.

Anderson's (2010) massive re-deployment hypothesis theorizes that brain regions perform a single "working" that is put to numerous cognitive "uses." Drawing an analogy with a Swiss Army Knife, brain regions possess characteristic workings (one has a knife, another a corkscrew), that each have different uses (e.g., a knife could cut a peach or a cherry). Thus, brain areas have many functions at one level of description (the cognitive process or task) and one function at another level of description (the working) (see also Bergeron, 2007; Price & Friston, 2005).

McCaffrey (2015) claims Craver's account explains the sense in which brain regions are multi-functional. A multi-functional region performs same role in different cognitive capacities. Neural reuse strikingly implies that diverse cognitive capacities—e.g., musical syntax and motor mimicry—share cognitive machinery and hence models of those capacities might be mutually informative (Anderson, 2010). But McCaffrey cautions against assuming a "conserved role"—some biological components perform the same role in different capacities, but others have "variable roles" where this kind of functional analysis fails. Remaining questions about functional attribution include: (1) Can etiological functions account for multi-functionality? (2) If so, should one privilege the function for which a region was originally selected? (3) Will any single account of "function" capture the diverse aims and uses in neuroscience?

Rathkopf (2013) claims that, in light of neural reuse, neuroscientists should abandon "task-bound functional analysis"—that is, they should stop assigning functions such as the "visual word form area" that denote a particular task or behavior. Instead, they should characterize brain regions in terms of their "intrinsic functions,"—e.g., the specific computations those regions perform (see also Shine et al., 2016). For Rathkopf, task bound functions will fail to explain why a region has the structural composition it does or to predict what tasks will recruit the region.

The latter relates to the controversy surrounding "reverse inference" in cognitive neuroimaging (Hutzler, 2014; Machery, 2014; Poldrack, 2006). Reverse inference occurs when one infers a mental process on the basis of brain activity measured via neuroimaging. Suppose researchers establish a link between fear—e.g., showing participants scary images—and amygdala activation measured by fMRI. Reverse inference would occur when, having established this link, researchers infer that amygdala activation in another study signifies that the participant was experiencing fear. But if the amygdala is recruited for a host of non-fear processes, as it has been shown to (Lindquist et al., 2012), then inferring fear on the basis of amygdala activation is unwarranted (Poldrack, 2006). The general problem for reverse inference is that brain regions seem to participate in multiple cognitive functions—e.g., region R has functions F_1, F_2, F_3, etc. Machery (2014) argues that performing reverse inference responsibly involves comparing competing hypotheses and knowing the likelihood that activity in R reflects F_1 versus, say, F_2 or F_3. But if brain regions have intrinsic functions that signify an underlying computation, as Rathkopf (2013) and others (e.g., Price & Friston, 2005; Shine et al., 2016) have argued, then perhaps reverse inference from activation in a region to its intrinsic function I is warranted across a broad range of tasks.[12]

Neural reuse also has important consequences for debates about embodied cognition (Goldman, 2012; Kiverstein, 2020) and the modularity of mind (Zerilli, 2021). Modularists argue that the mind's functional architecture is modular—i.e., it has components (modules) that are domain specific, automatic, informationally encapsulated,

innate, etc. (Coltheart, 1999; Fodor, 1983).[13] Modularists claim that cognitive capacities such as language are carried out by domain specific modules. Researchers often distinguish functional modularity (the mind has functionally distinct modules) from anatomical modularity (mental modules reside in distinct brain regions) (Bergeron, 2007).

Jungé and Dennett (2010) think neural reuse violates anatomical modularity while sparing functional modularity. But Zerilli (2021, Ch. 4) claims that reuse neural cuts against functional modularity too. For Zerilli, modularity entails that cognitive capacities such as language and folk physics are, in principle, functionally dissociable—i.e., they can be separately modified. But if every brain region is re-deployed across multiple capacities, then the circuitry underlying these capacities overlaps and hence they cannot be separately modified (e.g., lesioning a region common to language and folk physics will lead to deficits in both). One can respond that the mind is modular at the level of regional workings, but then modularity would not apply to cognitive capacities such as language.

4 | NEURAL DEGENERACY AND MULTIPLE REALIZABILITY

4.1 | Findings

Localization of Function holds that cognitive functions stably map to onto some anatomical locale. But recent research suggests that different brain areas sometimes perform the same cognitive function under different circumstances (Price & Friston, 2002). This kind of "many-to-one" mapping from brain regions to cognitive functions is called *neural degeneracy*, a situation where different neural structures perform the same function (Boone, 2018; Figdor, 2010; Price & Friston, 2002). Returning to the Swiss Army Knife analogy—perhaps one region is a scissors and another a knife, but both are capable of piercing a film in the same way.[14] If true, then cognitive functions would not necessarily possess a single neural substrate.

Edelman and Gally define *degeneracy*, a widespread feature of biological systems (e.g., gene networks and metabolic networks), as "the ability of elements that are structurally different to perform the same function or yield the same output" (2001, 13763). Researchers contrast degeneracy with *redundancy*, which occurs when a system contains multiple "back-up" copies of the same physical structure. Redundancy is the possession of multiple tokens of the same structural type (e.g., two copies of a gene); degeneracy occurs when different types of structures perform the same function (e.g., having one of two different genes suffices for a trait to emerge). Degeneracy and redundancy are both ways that biological systems maintain robustness in the face of damage and systemic change (Edelman & Gally, 2001; Price & Friston, 2002).

Price and Friston (2002) theorize that neural degeneracy—here the ability of different cortical areas to support the same cognitive function—might explain the ability of patients to overcome cognitive deficits arising from lesions (e.g., following stroke). But they also argue that such lesions might unmask degeneracy that already exists within healthy individuals—e.g., $Region_A$ or $Region_B$ can both, independently perform function *F*. For example, single word reading is accomplished by two different neural routes, one using sub-lexical spelling-to-sound rules and one using whole-word recognition (Seghier & Price, 2018). Individuals can still read most words if only one of these pathways is damaged.

There are further reports of degeneracy related to reading (Seghier et al., 2012), affective experience (Doyle et al., 2022), and many other functions. Many neuroscientists now believe that degeneracy is a widespread phenomenon in the human brain (Pessoa, 2014). Degeneracy may occur both within individuals and between individuals (McCaffrey & Wright, 2022; Viola, 2021; Ward, 2022). De Brigard (2017) claims that some cognitive functions are carried out by different brain areas in younger versus elderly participants. Tang et al. (2006) report that native Chinese and English speakers have different patterns of brain activation when doing arithmetic. Seghier et al. (2012) report that healthy individuals differ in what circuits they employ to recognize words under rapid presentation. These findings suggest that the localization of function is strictly speaking false since degenerate functions map disjunctively onto anatomical loci.

4.2 | Implications

Neural degeneracy raises important questions about the relationship between mental kinds and their neural substrates (Figdor, 2010), the search for neural correlates or biomarkers of cognitive functions (Coninx, 2021), and individual differences in cognitive neuroscience (Viola, 2021; Ward, 2022). Is a mental function such as pain equivalent to the operation of a localizable brain mechanism? Is a pain realized by the same neural mechanisms in different individuals?

Multiple realizability is the idea that mental kinds such as "pain" can have different *neural realizers*—e.g., that human and octopus pain are realized by different neural structures. This concept has often been considered a barrier to mind-brain identity theories and to the reduction of psychology to neuroscience (Fodor, 1974; Putnam, 1967). Bechtel and Mundale (1999) argue that while multiple realizability is a conceptual possibility, it is not a live empirical option in cognitive neuroscience. They claim that putative cases of multiple realization within the human brain—e.g., multiple brain structures support "memory"—typically vanish when researchers adopt a more fine-grained taxonomy of mental functions—e.g., distinguishing episodic memory from procedural memory reveals mappings to specific structures such as the hippocampus and basal ganglia. Pain seems to be realized by multiple neural systems, but perhaps distinguishing different kinds of pain—e.g., nociception from affective pain—will resolve this apparent multiple realizability (Coninx, 2021; Michel, 2019).

But Figdor (2010, see also Polger & Shapiro, 2016) argues that evidence of neural degeneracy implies that even fine-grained mental kinds such as "word form recognition" are multiply realized within the human brain. Boone (2018) argues that multiple realization occurs within neural systems (e.g., the crab stomatogastric ganglion), but analyzes it in causal explanatory terms rather than as a relationship between higher-level and lower-level kinds. According to Boone's (2018) causal explanatory mechanism realization account (CEMR), multiple realization occurs when there is sameness or stability in the functional role (in Cummins' 1975 sense) performed by two neural structures despite causal differences in the mechanisms underlying that function.

Whatever one's analysis of multiple realization, neural degeneracy is increasingly seen as an important confound in literature on the search for the biomarkers of pain (Coninx, 2021) and the neural correlates of consciousness (Klein et al., 2020; Michel & Lau, 2020). Assessing whether neural degeneracy implies multiple realization requires grappling with philosophical issues such as the right way to individuate cognitive functions and the relationship between mental kinds and neural mechanisms (Boone, 2018; Craver, 2009; Hochstein, 2016; McCaffrey & Wright, 2022; Michaelian, 2011). It also requires confronting unresolved methodological and conceptual issues surrounding evidence of degeneracy.

First, it is uncertain how common degeneracy actually is. Khan et al. (2022), who developed an fMRI analysis pipeline to detect degeneracy, note that there are no standard analyses for measuring degeneracy in neuroimaging datasets. The best evidence comes from studies combining neuropsychology and neuroimaging findings (Seghier et al., 2012; Seghier & Price, 2018). But these lines of research are laborious and hence established examples remain relatively few. Second, theorists have proposed many different types of degeneracy that have different theoretical implications.

Inter-individual or *participant degeneracy* occurs when different subjects seem to perform the same cognitive function using different brain regions (Khan et al., 2022). Viola (2021) claims that cultural and individual differences in brain mapping are so common that neuroscientists should abandon the assumption of a "Platonic Brain"—i.e., of a structure-function mapping shared by healthy individuals. Ward (2022) analyzes the nature of cognitive variation, typing instances of cognitive variation by: (1) origin—whether the variation is fundamental versus evoked by environmental context, and (2) structure—whether the variation is continuous or discrete/categorical. Such concepts are necessary for understanding what individual differences in structure-function mappings signify. If participant degeneracy is the norm, then perhaps there are stable functional mappings within individuals or groups of individuals—e.g., perhaps there are stable mappings when we group participants by age (De Brigard, 2017), or by some dimension of cognitive variation (Ward, 2022).

By contrast, *intra-individual* or *condition degeneracy* occurs when an individual performs a cognitive function using different neural structures at different times or under different circumstances (Khan et al., 2022). This could theoreti-

cally occur during different trials within a study (Khan et al., 2022), or over one's lifespan (De Brigard, 2017). If condition degeneracy is widespread, then mappings would not be stable within individuals or populations (Anderson, 2014; Hutto, Seegers, and Segundo-Ortin, 2017).

Seghier and Price (2018) further distinguish intrinsic—i.e., subjects possess different underlying cognitive machinery—versus strategic—i.e., subjects possess the same cognitive resources but differ in how they consciously deploy them—forms of degeneracy. Thus, it is unclear whether examples of degeneracy signify that two different structures can perform the same computation, or whether they perform different computations that both suffice for carrying out a task. It would be most damaging to localizationism and suggestive of multiple realizability if condition degeneracy was common and this was not explained by the use of different cognitive strategies.

5 | CONTEXTUALISM

5.1 | Findings

Intrinsicality holds that the functions of brain structures can be understood relatively independently of the function of other structures. For example, if a network realizes the subjective experience of fear, then "fear" is properly attributed to that network and not the whole brain or to the network plus other interacting structures. Burnston construes intrinsicality as the assumption that "the function of a part *P* is to be specified with limited reference to the system in which *P* functions" (2020, 5). "Limited reference" captures the fact that functional attributions typically presume stable background conditions against which neural functions are performed (e.g., inferotemporal cortex processes shape information conditioned upon receiving the right set of inputs from early visual areas) (Klein, 2018).

Contextualism (Burnston, 2016a, 2016b, 2020; De Wit & Matheson, 2022; Hutto et al., 2017; Klein, 2012) proposes that intrinsicality is false: neural structures perform radically different functions in different contexts. I say "radically different functions" because functional architectures such the massive re-deployment hypothesis (Anderson, 2010) presuppose a form of *Weak Context-Sensitivity* where brain regions have an intrinsic working that is deployed for many higher-level cognitive uses. Contextualists claim that brain regions exhibit a form of *Strong Context-Sensitivity* where they lack an intrinsic working or computation (Burnston, 2016a; De Wit & Matheson, 2022). Contextualists believe functional mappings only hold in particular contexts—i.e., one cannot specify a structure's function in a universal, decontextualized way. Recalling the Swiss Army Knife analogy, a brain region might function as a scissors or a corkscrew depending on the context—e.g., what another region instructs it to do.[15]

McIntosh (2000) proposes that brain regions have different functions depending on their "neural context"—that is, the network of structures the region is currently engaged with. Klein (2012) argues that functional attribution in cognitive neuroscience must therefore be local and contextual rather than holding across the board. Klein claims that some parts of mechanical or informational processing systems lack functions that hold across every context. He illustrates this by considering the pistons of diesel trucks with engine brakes. In normal driving conditions, the piston speeds the truck up by compressing a fuel air mixture; when the brake is engaged, exhaust valves release this mixture and the piston compresses air in the cylinder, which slows the truck down. He argues that in such cases, a decontextualized functional description—i.e., the piston speeds or slows the truck—is meaningless. Assigning functions to individual brain regions is similarly uninformative because they have contextually varying functions.

Burnston (2016b, 2016a, 2020) uses detailed empirical examples to argue that cognitive neuroscience must embrace contextualism. In a representative example, Burnston (2016a) analyses the functions of the middle temporal area of extrastriate cortex (MT). MT is involved in multiple functions including motion processing (sensitivity to speed and direction of motion) and depth perception. Assigning one function to MT would entail dividing MT into sub-populations dedicated to motion versus depth perception or claiming that MT performs a single computation that is used for both functions (McCaffrey, 2015; Shine et al., 2016). Burnston argues that many of the same neurons encode information about motion and depth. Furthermore, he contends that the computational models of the MT's

contribution to motion and depth perception do not point toward a single underlying computation. He concludes that MT neurons enact one coding scheme in connection with motion processing and another with depth perception.

Burnston (2020) claims that multiplexing—a phenomenon in which multiple information processing streams share a common neural substrate—might support this form of strong context-sensitivity. Multiplexing occurs when one signal has distinct embedded components within it that different receivers can extract (Akam & Kullmann, 2014). In some cases, temporal dynamics of the local field potential (LFP)—extracellular electrical potential reflecting aggregate neuronal activity—can extract distinct signals from a single neural population (Watrous et al., 2015). Neuroscientists have often characterized strong context-sensitivity in invertebrate models when changes in the inputs to a circuit and/or its neuromodulatory environment (i.e., what neuromodulators are present in the extracellular milieu) shift the circuit's coding properties (Bargmann, 2012; Marder, 2012). Some neuroscientists think these models explain the multi-functionality of cortical regions better than models that presume segregated sub-populations or invariant computations (De Wit & Matheson, 2022; Gu, van Rijn, and Meck, 2015).

5.2 | Implications

Contextualism has profound implications for neuroscience and its philosophy (Burnston, 2020; Klein, 2012; Silberstein, 2021). Cognitive neuroscience typically seeks univocal functional attributions that capture a structure's function in a universal, decontextualized fashion. Both task-bound functional attributions—e.g., "the visual word form area's function is to detect orthographic regularities"—and intrinsic functional descriptions referencing a structure's specific computation (Rathkopf, 2013; Shine et al., 2016)—presuppose that there is some way of describing a structure's function that is stable across contexts. Burnston (2020, see also Anderson, 2014, Viola, 2017, De Wit & Matheson, 2022) claims that cognitive neuroscience should abandon this approach.

Some neuroscientists view systematic or one-to-one mappings between cognitive functions and neural structures as one of cognitive neuroscience's primary goals (Poldrack, 2010; Price & Friston, 2005). In this framework, neuroscientists are instructed to revise their taxonomy of cognitive functions and their division of neural structures until one-to-one mappings emerge. This implies that one-to-many (reuse) and many-to-one (degeneracy) mappings between structures and functions only occur when neuroscientists get things wrong. But contextualists counter that even when neuroscientists get things right, they will fail to achieve one-to-one mappings (Anderson, 2014; Burnston, 2020; Pessoa, 2014).

Contextualism provides a different perspective on the problem of reverse inference in neuroimaging (Section 3.2). If brain regions have contextually varying functions, then one should take context into account when performing reverse inference (Hutzler, 2014; Klein, 2012; Nathan & Del Pinal, 2017). Klein (2012) argues that reverse inference on a region R should account for the neural context C—i.e., the network of areas activated along with R.[16] Forms of reverse inference that apply machine learning techniques—e.g., multi-voxel pattern analysis (MVPA)—to distributed patterns of brain activation seem to outperform traditional region of interest approaches (Haxby et al., 2001; Nathan & Del Pinal, 2017; Poldrack, 2011). This could signify that the functions of networks are more stable than the functions of individual regions or that neural structures exhibit strong context-sensitivity. One worry here is that some contexts relevant to neural functioning (e.g., neuromodulatory context), may require specialized methodologies to detect or be inaccessible given current neuroimaging techniques.

Contextualism raises a host of unresolved issues. First, do *brain regions* typically exhibit strong contextualism? McCaffrey (2015) argues that the brain is "functionally heterogeneous"—i.e., some neural structures have intrinsic functions where others exhibit more radical forms of context-sensitivity. Second, *what types of neural structures* exhibit strong contextualism? Klein's (2012) analysis of reverse inference presupposes that the functions of networks are more stable than those of individual regions. But some authors claim that networks also have contextually varying functions and thus strong context-sensitivity occurs at multiple scales of neural organization (Pessoa, 2014). Lastly, does multiplexing capture the sense in which human brain regions are contextual and, if so, is that the primary means through which strong contextualism occurs (Burnston, 2020)?

6 | EVOLVING CONCEPTS AND THE FUTURE OF LOCALIZATIONISM

6.1 | How Radical a Shift is Needed? TALONs, Protean Brains, and the Return of Radical Anti-Localizationism

Contemporary localizationism assumes that neural structures are specialized for particular cognitive functions (*Structural Specalization*), cognitive functions have stable, localizable realizers (*Localization of Function*), and that the functions of neural structures can be understood in relative isolation (*Intrinsicality*). Recent work suggests that the cerebral cortex departs from some or all these tenets. What does this mean for localizationism?

Some researchers claim it is time for cognitive neuroscience to abandon localizationism (Anderson, 2014; Pessoa, 2022; Silberstein, 2021; Silberstein & Chemero, 2013; Westlin et al., 2023). This would not mean jettisoning attempts to link neural structure to mental functioning, but rather acknowledging that structure-function mappings will be highly unstable and contextual and that our analytic tools and theoretical assumptions should adapt accordingly (Anderson, 2014, Ch. 4). Silberstein and Chemero (2013, Silberstein, 2021) argue that localization fails in complex systems exhibiting robustness, feedback, and contextually-shifting dynamic interactions between components. They believe that contextualism signifies a return to holism since one cannot understand a structure's function without characterizing the complex dynamics of the entire system in which it is embedded. Burnston (2020) counters that functions genuinely localize to specific structures (e.g., regions or networks) *within a specific context*.

Who is right depends on how radically the brain violates localizationism. The brain's functional architecture departs from strict localizationism along at least three independent dimensions (*Structural Specialization*, *Localization of Function*, and *Intrinsicality*). Departures along these dimensions vary in both *degree* and *scope*. By "degree," I mean how severe the violation is. A region with 20 distinct functions violates specialization more than one with three. By "scope," I mean how common the violation is. If 300 brain regions exhibit degeneracy, this is a broader scope than if 30 regions do. The proposals reviewed so far—e.g., Anderson's massive re-deployment hypothesis (2010)—abandon some aspects of localizationism (e.g., *Structural Specialization*) while retaining others (e.g., *Localization of Function*). But a functional architecture that violates all three tenets to a large degree and over a large scope would signal a return to radical anti-localizationism—i.e., to architectures closer to equipotentialism and/or holism.

Some researchers have recently embraced radical anti-localizationism (Anderson, 2014; Hutto et al., 2017; Pessoa, 2022). Anderson (2014, Ch. 4) argues that brain regions perform a broad range of overlapping functions and primarily differ in their disposition to engage in some functions rather than others. Returning to this Swiss Army Knife analogy: multiple regions possess a scissors, knife, and corkscrew; they differ only in their tendency to deploy each instrument. In this architecture, cognitive functions (e.g., working memory) are realized by transiently assembled local neural sub-systems (TALONs)—i.e., structural coalitions whose composition differs in each iteration of the function.[17]

Anderson's new model reflects a strong pivot toward anti-localizationism compared to his earlier work (e.g., 2010). The dispositional account of regional functions is a step toward equipotentialism; the rejection of localization entailed by TALONs is a step toward holism. Hutto et al. (2017) similarly explore the possibility of "Protean Brains." Their Protean Brain Hypothesis (PBH), "conjectures that brain structures are functionally malleable: they make use of neural structures in inventive and on the fly improvisations to suit circumstances and context" (2017, 210). While certainly not mainstream, these proposals are gaining traction and suggest that cognitive neuroscience might revive the anti-localizationist views of Lashley and Flourens with contemporary twists.

6.2 | Progress and Conceptual Change

In this essay, I examined how concepts of functional localization have changed since the 19th Century to the present and the empirical and philosophical controversies these shifts create. Philosophers of science have recently argued that scientific concepts are tools for formulating lines of investigation and interpreting empirical results. Griffiths and

Stotz argue that, for scientists, concepts are "tools which classify experience in ways that meet their specific needs and which are reshaped in the light of new empirical findings (2008, 10). The concept of "functional localization" historically emerged with the notion of a discrete "cerebral organ" onto which cognitive functions neatly mapped. This conception of localization informed the search for lesions affecting specific cognitive functions (Young, 1968) and the development of neuroimaging to localize cognitive functions (Figdor, 2010). Over time, neuroscientists uncovered findings that challenged this ideal. This necessitated the development of new concepts (and importing concepts from other disciplines) such as "neural degeneracy," "neural context," and "neural reuse," to refine the concept of functional localization and to test its limits.

Such conceptual innovations reflect a form of progress in cognitive neuroscience on par with the development of novel imaging and intervention tools (Griffiths & Stotz, 2008; Haueis, 2022). Haueis (2022) stresses that experimental findings can lead to the formation of novel concepts, which in turn shape the future course of empirical inquiry. Whether the brain's functional architecture exhibits high degrees of reuse, degeneracy, and context-sensitivity, and whether this signifies a return to radical forms of anti-localizationism, are questions we can now ask with increasing conceptual and experimental precision (Anderson, 2014; Anderson et al., 2013; Burnston, 2020). The brain's "true" functional architecture appears to reside in a vast conceptual space between the simplistic forms of localizationism and anti-localizationism neuroscientists first articulated. Exploring the precise contours of this space will require interdisciplinary collaborations of the kind reviewed in this essay and further conceptual innovations.

ORCID

Joseph B. McCaffrey https://orcid.org/0000-0001-6210-0152

ENDNOTES

[1] While much of this piece applies to functional localization in general, I focus on the localization of cognitive functions, as opposed to non-cognitive sensory, motor, or maintenance functions (Haueis, 2018, see also Klein, 2018), within the neocortex.

[2] Some propose that cognitive neuroscience lacks one-to-one mappings between mental functions and neural structures because we have the wrong "cognitive ontology" or set of mental functions (Anderson, 2015; McCaffrey & Wright, 2022; Poldrack, 2010; Price & Friston, 2005; Viola, 2017).

[3] Gall identified his 27 faculties through studying child prodigies, individual variation in artistic and intellectual abilities, and clinical cases. He inferred the existence of cerebral organs from studying skull convolutions in humans and animals. While Gall is sometimes viewed as a pseudoscientist for his phrenological system, his organology was influential in establishing the localization of cortical function (Finger, 2009; Young, 1968).

[4] Flourens conducted lesion studies on animals that allegedly demonstrated that as the cortex was resected, animals did not lose specific functions (e.g., sensory, motor, motivational), but rather were functionally impaired across the board as more cortex was lost.

[5] Broca famously described a patient dubbed "Tan" with damage to a circumscribed part of the left inferior frontal lobe who could only speak the word "tan," but could understand speech. This instance of Broca's aphasia, which contrasts with Wernicke's aphasia in which temporal lobe damage impairs language comprehension but not speech production, is considered the first empirical validation of localizationism (Young, 1968). Broca's area resides in the inferior frontal lobe of the dominant hemisphere, typically on the left.

[6] Some authors—e.g., Young (1968), Mundale (2002), Gold and Roskies (2008)—consider functional localization a scientific concept while others construe localization as a thesis, research program (e.g., Anderson, 2014; Figdor, 2010), explanatory strategy (Silberstein & Chemero, 2013), or heuristic (Bechtel & Richardson, 2010). These formulations are largely interchangeable for our present purposes.

[7] While one can endorse both equipotentialism and holism, the concepts are independent. A cortex in which any isolated part could perform any of its functions equally would be equipotential but not holistic. A cortex whose parts functionally differ, but in which every part participates in every mental function would be holistic but not equipotential.

[8] On my view, this is not merely polysemy where "localization" sometimes refers to "structural specialization" and sometimes refers "intrinsicality" as when "innate" variously means "genetically encoded," "present at birth," "not learned," etc. (Mameli & Bateson, 2006). Instead, I believe these tenets comprise the dimensions (Godfrey-Smith, 2009) or parameters (Akagi, 2022) that structure the contemporary concept of functional localization.

[9] Though this section focuses mainly on cortical regions, neural reuse may apply to brain networks as well (Pessoa, 2014; Zerilli, 2021).

[10] In evolutionary biology, "exaptation" occurs when a structure that evolved for one purpose, or no purpose, is coopted for a novel role over phylogenetic timescales (Gould & Vrba, 1982).

[11] While some regions were more specialized than others, nearly all regions were recruited for multiple domains.

[12] A reviewer noted that a region's intrinsic function may not signify a mental or cognitive process at all, which would complicate attempts to perform reverse inference.

[13] Modularists dispute the precise criteria. For example, Coltheart (1999) argues that domain specificity is the hallmark of modularity while innateness is not essential.

[14] Or perhaps the better analogy is if is two regions are each a differently sized and shaped pair of scissors.

[15] One might question whether systems can change their functional properties without a corresponding change in structural properties, since a system could not be a knife or a corkscrew without changing its structural arrangement. While this issue is metaphysically fraught, it is more apparent how this would work with information processing in neural systems (e.g., the same neuron fires in different burst patterns in different contexts) than with mechanical artifacts.

[16] Hutzler (2014) advocates taking the behavioral context or task setting into account.

[17] See Anderson (2016) for a review of these proposals and Anderson (2014) for a more detailed explication. For some philosophical criticisms of these proposals, see Kaplan and Craver (2016), McCaffrey and Machery (2016), and Shine et al. (2016).

REFERENCES

Akagi, M. (2022). Cognition as the sensitive management of an agent's behavior. *Philosophical Psychology*, *35*(5), 718–741. https://doi.org/10.1080/09515089.2021.2014802

Akam, T., & Kullmann, D. M. (2014). Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nature Reviews Neuroscience*, *15*(2), 111–122. https://doi.org/10.1038/nrn3668

Anderson, M. L. (2007). The Massive Redeployment Hypothesis and the Functional Topography of the Brain. *Philosophical Psychology*, *20*(2), 143–174. https://doi.org/10.1080/09515080701197163

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences*, *33*(4), 245–266. https://doi.org/10.1017/S0140525X10000853

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. The MIT Press. https://doi.org/10.7551/mitpress/10111.001.0001

Anderson, M. L. (2015). Mining the Brain for a New Taxonomy of the Mind: A new Taxonomy of the Mind. *Philosophy Compass*, *10*(1), 68–77. https://doi.org/10.1111/phc3.12155

Anderson, M. L. (2016). Précis of After Phrenology: Neural Reuse and the Interactive Brain. *Behavioral and Brain Sciences*, *39*, e120. https://doi.org/10.1017/S0140525X15000631

Anderson, M. L., Kinnison, J., & Pessoa, L. (2013). Describing functional diversity of brain regions and brain networks. *NeuroImage*, *73*, 50–58. https://doi.org/10.1016/j.neuroimage.2013.01.071

Barack, D. L. (2019). Cognitive recycling. *The British Journal for the Philosophy of Science*, *70*(1), 239–268. https://doi.org/10.1093/bjps/axx024

Bargmann, C. I. (2012). Beyond the connectome: How neuromodulators shape neural circuits. *BioEssays*, *34*(6), 458–465. https://doi.org/10.1002/bies.201100185

Bechtel, W., & Mundale, J. (1999). Multiple Realizability Revisited: Linking Cognitive and Neural States. *Philosophy of Science*, *66*(2), 175–207. https://doi.org/10.1086/392683

Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT press.

Bergeron, V. (2007). Anatomical and Functional Modularity in Cognitive Science: Shifting the Focus. *Philosophical Psychology*, *20*(2), 175–195. https://doi.org/10.1080/09515080701197155

Boone, T., Van Rooy, N., & De Brigard, F. (2022). Not Every Thing Must Go. *Journal of Cognitive Neuroscience*, *35*(3), 1–4. https://doi.org/10.1162/jocn_a_01931

Boone, W. (2018). Multiple realization and robustness. *Biological Robustness: Emerging Perspectives from within the Life Sciences*, 75–94. https://doi.org/10.1007/978-3-030-01198-7

Briggman, K. L., & Kristan, W. B. (2008). Multifunctional Pattern-Generating Circuits. *Annual Review of Neuroscience*, *31*(1), 271–294. https://doi.org/10.1146/annurev.neuro.31.060407.125552

Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). *Bulletin de la Société Anatomique*, *6*, 330–357.

Burnston, D. C. (2016a). Computational neuroscience and localized neural function. *Synthese*, *193*(12), 3741–3762. https://doi.org/10.1007/s11229-016-1099-8

Burnston, D. C. (2016b). A contextualist approach to functional localization in the brain. *Biology and Philosophy*, *31*(4), 527–550. https://doi.org/10.1007/s10539-016-9526-2

Burnston, D. C. (2020). Getting over Atomism: Functional Decomposition in Complex Neural Systems. *The British Journal for the Philosophy of Science*, *72*(3), 743–772. https://doi.org/10.1093/bjps/axz039

Burnston, D. C., & Haueis, P. (2021). Evolving concepts of "hierarchy" in systems neuroscience. In F. Calzavarini & M. Viola (Eds.), *Neural Mechanisms: New Challenges in the Philosophy of Neuroscience* (pp. 113–141). Springer. https://doi.org/10.1007/978-3-030-54092-0

Chirimuuta, M., & Gold, I. (2009). The embedded neuron, the enactive field? In J. Bickle (Ed.), *The oxford handbook of philosophy and neuroscience (pp. 200-225)*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195304787.001.0001

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences*, *3*(3), 115–120. https://doi.org/10.1016/S1364-6613(99)01289-9

Coninx, S. (2021). *The notorious neurophilosophy of pain: A family resemblance approach to idiosyncrasy and generalizability*. Mind & Language. mila.12378 Retrieved from: https://doi.org/10.1111/mila.12378

Craver, C. F. (2001). Role Functions, Mechanisms, and Hierarchy. *Philosophy of Science*, *68*(1), 53–74. https://doi.org/10.1086/392866

Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575–594. https://doi.org/10.1080/09515080903238930

Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, *72*(20), 741. https://doi.org/10.2307/2024640

De Brigard, F. (2017). Cognitive systems and the changing brain. *Philosophical Explorations*, *20*(2), 224–241. https://doi.org/10.1080/13869795.2017.1312503

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: the neuronal recycling hypothesis. In S. Dehaene, J.-R. Duhamel, M. D. Hauser, & G. Rizzolatti (Eds.), *From Monkey Brain to Human Brain* (pp. 133–157). MIT Press.

De Wit, M. M., & Matheson, H. E. (2022). Context-sensitive computational mechanistic explanation in cognitive neuroscience. *Frontiers in Psychology*, *13*, 903960. https://doi.org/10.3389/fpsyg.2022.903960

Doyle, C. M., Lane, S. T., Brooks, J. A., Wilkins, R. W., Gates, K. M., & Lindquist, K. A. (2022). Unsupervised classification reveals consistency and degeneracy in neural network patterns of emotion. *Social Cognitive and Affective Neuroscience*, *17*(11), 995–1006. nsac028. https://doi.org/10.1093/scan/nsac028

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Fedorenko, E., & Blank, I. A. (2020). Broca's Area Is Not a Natural Kind. *Trends in Cognitive Sciences*, *24*(4), 270–284. https://doi.org/10.1016/j.tics.2020.01.001

Figdor, C. (2010). Neuroscience and the Multiple Realization of Cognitive Functions. *Philosophy of Science*, *77*(3), 419–456. https://doi.org/10.1086/652964

Finger, S. (2009). The birth of localization theory. In *Handbook of clinical neurology* (Vol. 95, pp. 117–128). Elsevier.

Flourens, M. J. P., & Nadel, L. (1824). Recherches experimentales sur les propietes et fonctions due systeme nerveus dans les animaus vertebres. *Paris.Polster, MR, Nadel, L, and Schacter, DL (1991) Cognitive neuroscience analysis of memory: A historical perspective. Journal of Cognitive Neuroscience*, *3*(2), 95–116. https://doi.org/10.1162/jocn.1991.3.2.95

Fodor, J. (1974). The disunity of science as a working hypothesis. *Synthese*, *28*(2), 97–115. https://doi.org/10.1007/bf00485230

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Foit, N. A., Yung, S., Lee, H. M., Bernasconi, A., Bernasconi, N., & Hong, S. J. (2022). A whole-brain 3D myeloarchitectonic atlas: Mapping the Vogt-Vogt legacy to the cortical surface. *NeuroImage*, *263*, 119617. https://doi.org/10.1016/j.neuroimage.2022.119617

Frisch, S. (2014). How cognitive neuroscience could be more biological and what it might learn from clinical neuropsychology. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00541

Gall, F. J. (1835). *On the functions of the brain and of each of its parts: The influence of the brain on the form of the head: The difficulties and means of determining the fundamental qualities and faculties, and of discovering the seat of their organs*. (W. Lewis, Trans.). Marsh, Capen, & Lyon.

Gamboa, J. P. (2020). Goltz against cerebral localization: Methodology and experimental practices. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *84*, 101304. https://doi.org/10.1016/j.shpsc.2020.101304

Garson, J. (2011). Selected effects and causal role functions in the brain: The case for an etiological approach to neuroscience. *Biology and Philosophy*, *26*(4), 547–565. https://doi.org/10.1007/s10539-011-9262-6

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 8–197. https://doi.org/10.1038/72140

Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199552047.001.0001

Gold, I., & Roskies, A. L. (2008). Philosophy of Neuroscience. In M. Ruse (Ed.), *The Oxford Handbook of Philosophy of Biology* (pp. 349–380). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195182057.003.0016

Goldman, A. I. (2012). A Moderate Approach to Embodied Cognitive Science. *Review of Philosophy and Psychology*, *3*(1), 71–88. https://doi.org/10.1007/s13164-012-0089-0

Gould, S. J., & Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology*, *8*(1), 4–15. https://doi.org/10.1017/S0094837300004310

Griffiths, P. E., & Stotz, K. (2008). Experimental Philosophy of Science. *Philosophy Compass*, *3*(3), 507–521. https://doi.org/10.1111/j.1747-9991.2008.00133.x

Gu, B.-M., van Rijn, H., & Meck, W. H. (2015). Oscillatory multiplexing of neural population codes for interval timing and working memory. *Neuroscience & Biobehavioral Reviews*, *48*, 160–185. https://doi.org/10.1016/j.neubiorev.2014.10.008

Haueis, P. (2016). The life of the cortical column: opening the domain of functional architecture of the cortex (1955–1981). *History & Philosophy of the Life Sciences*, *38*(3), 1–27. https://doi.org/10.1007/s40656-016-0103-4

Haueis, P. (2018). Beyond cognitive myopia: a patchwork approach to the concept of neural function. *Synthese*, *195*(12), 5373–5402. https://doi.org/10.1007/s11229-018-01991-z

Haueis, P. (2021). The death of the cortical column? Patchwork structure and conceptual retirement in neuroscientific practice. *Studies In History and Philosophy of Science Part A*, *85*, 101–113. https://doi.org/10.1016/j.shpsa.2020.09.010

Haues, P. (2022). Exploratory concept formation and tool development in neuroscience. *Philosophy of Science*, 1–44. https://doi.org/10.1017/psa.2022.79

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hochstein, E. (2016). Categorizing the Mental. *The Philosophical Quarterly*, *66*(265), 745–759. https://doi.org/10.1093/pq/pqw001

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. https://doi.org/10.1016/j.tics.2011.12.008

Hutto, D. D., Peeters, A., & Segundo-Ortin, M. (2017). Cognitive ontology in flux: The possibility of protean brains. *Philosophical Explorations*, *20*(2), 209–223. https://doi.org/10.1080/13869795.2017.1312502

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, *84*, 1061–1069. https://doi.org/10.1016/j.neuroimage.2012.12.075

Jungé, J. A., & Dennett, D. C. (2010). Multi-use and constraints from original use. *Behavioral and Brain Sciences*, *33*(4), 277–278. https://doi.org/10.1017/s0140525x1000124x

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170. https://doi.org/10.1073/pnas.1005062107

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kaplan, D. M., & Craver, C. F. (2016). A registration problem for functional fingerprinting. *Behavioral and Brain Sciences*, *39*, 15–16. https://doi.org/10.1017/s0140525x15001582

Khan, Z., Wang, Y., Sennesh, E., Dy, J., Ostadabbas, S., van de Meent, J.-W., Hutchinson, J. B., & Satpute, A. B. (2022). A Computational Neural Model for Mapping Degenerate Neural Architectures. *Neuroinformatics*, *20*(4), 965–979. https://doi.org/10.1007/s12021-022-09580-9

Kiverstein, J. (2020). Embodied cognition and the neural reuse hypothesis. In A. J. Lerner, S. Cullen, & S. Leslie (Eds.), *Current controversies in philosophy of cognitive science* (pp. 87–107). Routledge.

Klein, C. (2012). Cognitive Ontology and Region- versus Network-Oriented Analyses. *Philosophy of Science*, *79*(5), 952–960. https://doi.org/10.1086/667843

Klein, C. (2018). Mechanisms, resources, and background conditions. *Biology and Philosophy*, *33*(5–6), 36. https://doi.org/10.1007/s10539-018-9646-y

Klein, C., Hohwy, J., & Bayne, T. (2020). Explanation in the science of consciousness: From the neural correlates of consciousness (NCCs) to the difference makers of consciousness (DMCs). *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.60

Lashley, K. (1929). *Brain Mechanisms and Intelligence*. Chicago University Press.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, *35*(3), 121–143. https://doi.org/10.1017/S0140525X11000446

Loukas, M., Pennell, C., Groat, C., Tubbs, R. S., & Cohen-Gadol, A. A. (2011). Korbinian Brodmann (1868-1918) and His Contributions to Mapping the Cerebral Cortex. *Neurosurgery*, *68*(1), 6–11. https://doi.org/10.1227/NEU.0b013e3181fc5cac

Machery, E. (2014). Defense of Reverse Inference. *The British Journal for the Philosophy of Science*, *65*(2), 251–267. https://doi.org/10.1093/bjps/axs044

Mameli, M., & Bateson, P. (2006). Innateness and the Sciences. *Biology and Philosophy*, *21*(2), 155–188. https://doi.org/10.1007/s10539-005-5144-0

Marder, E. (2012). Neuromodulation of Neuronal Circuits: Back to the Future. *Neuron*, *76*(1), 1–11. https://doi.org/10.1016/j.neuron.2012.09.010

McCaffrey, J., & Wright, J. (2022). Neuroscience and Cognitive Ontology: a Case for Pluralism. In F. De Brigard & W. Sinnott-Armstrong (Eds.), *Neuroscience and Philosophy*. MIT Press. https://doi.org/10.7551/mitpress/12611.001.0001

McCaffrey, J. B. (2015). The Brain's Heterogeneous Functional Landscape. *Philosophy of Science*, *82*(5), 1010–1022. https://doi.org/10.1086/683436

McCaffrey, J. B., & Machery, E. (2016). The reification objection to bottom-up cognitive ontology revision. *Behavioral and Brain Sciences*, *39*, e125. https://doi.org/10.1017/S0140525X15001594

McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Networks*, *13*(8–9), 861–870. https://doi.org/10.1016/S0893-6080(00)00059-9

Michaelian, K. (2011). Is memory a natural kind? *Memory Studies*, *4*(2), 170–189. https://doi.org/10.1177/1750698010374287

Michel, M. (2019). Fish and microchips: On fish pain and multiple realization. *Philosophical Studies*, *176*(9), 2411–2428. https://doi.org/10.1007/s11098-018-1133-4

Michel, M., & Lau, H. (2020). On the dangers of conflating strong and weak versions of a theory of consciousness. *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.54

Mundale, J. (2002). Concepts of localization: Balkanization in the brain. *Brain and Mind*, *3*(3), 313–330. https://doi.org/10.1023/A:1022912227833

Nathan, M. J., & Del Pinal, G. (2017). The Future of Cognitive Neuroscience? Reverse Inference in Focus. *Philosophy Compass*, *12*(7), e12427. https://doi.org/10.1111/phc3.12427

Neander, K. (1991). Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, *58*(2), 168–184. https://doi.org/10.1086/289610

Pessoa, L. (2014). Understanding brain networks and brain organization. *Physics of Life Reviews*, *11*(3), 400–435. https://doi.org/10.1016/j.plrev.2014.03.005

Pessoa, L. (2022). *The entangled brain: How perception, cognition, and emotion are woven together*. MIT Press.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*(6157), 585–589. https://doi.org/10.1038/331585a0

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A. (2010). Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? *Perspectives on Psychological Science*, *5*(6), 753–761. https://doi.org/10.1177/1745691610388777

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Polger, T. W., & Shapiro, L. A. (2016). *The multiple realization book*. Oxford University Press.

Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, *6*(10), 416–421. https://doi.org/10.1016/S1364-6613(02)01976-9

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3–4), 262–275. https://doi.org/10.1080/02643290442000095

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, mind and religion* (pp. 37–48). University of Pittsburgh Press.

Rathkopf, C. A. (2013). Localization and Intrinsic Function. *Philosophy of Science*, *80*(1), 1–21. https://doi.org/10.1086/668878

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology Neurosurgery and Psychiatry*, *20*(1), 11–21. https://doi.org/10.1136/jnnp.20.1.11

Seghier, M. L., Neufeld, N. H., Zeidman, P., Leff, A. P., Mechelli, A., Nagendran, A., Riddoch, J. M., Humphreys, G. W., & Price, C. J. (2012). Reading without the left ventral occipito-temporal cortex. *Neuropsychologia*, *50*(14), 3621–3635. https://doi.org/10.1016/j.neuropsychologia.2012.09.030

Seghier, M. L., & Price, C. J. (2018). Interpreting and Utilising Intersubject Variability in Brain Function. *Trends in Cognitive Sciences*, *22*(6), 517–530. https://doi.org/10.1016/j.tics.2018.03.003

Shine, J. M., Eisenberg, I., & Poldrack, R. A. (2016). Computational specificity in the human brain. *Behavioral and Brain Sciences*, *39*, e131. https://doi.org/10.1017/s0140525x1500165x

Silberstein, M. (2021). Constraints on localization and decomposition as explanatory strategies in the biological sciences 2.0. In F. Calzavarini & M. Viola (Eds.), *Neural Mechanisms: New Challenges in the Philosophy of Neuroscience* (pp. 363–393). Springer. https://doi.org/10.1007/978-3-030-54092-0

Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, *80*(5), 958–970. https://doi.org/10.1086/674533

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nature Neuroscience*, *17*(5), 652–660. https://doi.org/10.1038/nn.3690

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., Reiman, E. M., & Liu, Y. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences*, *103*(28), 10775–10780. https://doi.org/10.1073/pnas.0604416103

Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. The MIT press.

Venturelli, A. N. (2021). Conceptual Change in Visual Neuroscience: The Receptive Field Concept. *International Studies in the Philosophy of Science*, *34*(1), 41–57. https://doi.org/10.1080/02698595.2021.2002675

Viola, M. (2017). Carving Mind at Brain's Joints. The Debate on Cognitive Ontology. *Phenomenology and Mind*, 162–172. https://doi.org/10.13128/PHE_MI-21115

Viola, M. (2021). Beyond the Platonic Brain: Facing the challenge of individual differences in function-structure mapping. *Synthese*, *199*(1–2), 2129–2155. https://doi.org/10.1007/s11229-020-02875-x

Viola, M., & Zanin, E. (2017). The standard ontological framework of cognitive neuroscience: Some lessons from Broca's area. *Philosophical Psychology*, *30*(7), 945–969. https://doi.org/10.1080/09515089.2017.1322193

Ward, Z. B. (2022). Cognitive Variation: The Philosophical Landscape. *Philosophy Compass*, *17*(10), e12882. https://doi.org/10.1111/phc3.12882

Watrous, A. J., Fell, J., Ekstrom, A. D., & Axmacher, N. (2015). More than spikes: Common oscillatory mechanisms for content specific neural representations during perception and memory. *Current Opinion in Neurobiology*, *31*, 33–39. https://doi.org/10.1016/j.conb.2014.07.024

Westlin, C., Theriault, J. E., Katsumi, Y., Nieto-Castanon, A., Kucyi, A., Ruf, S. F., Brown, S. M., Pavel, M., Erdogmus, D., Brooks, D. H., Quigley, K. S., Whitfield-Gabrieli, S., & Barrett, L. F. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences*, *27*(3), 246–257. https://doi.org/10.1016/j.tics.2022.12.015

Young, R. M. (1968). The functions of the brain: Gall to Ferrier (1808-1886). *Isis*, *59*(3), 250–268. https://doi.org/10.1086/350395

Zerilli, J. (2021). *The Adaptable Mind: What Neuroplasticity and Neural Reuse Tells Us About Language and Cognition*. Oxford University Press.

AUTHOR BIOGRAPHY

Joseph B. McCaffrey is an Assistant Professor of Philosophy, Faculty in the Neuroscience Program, and Faculty in Medical Humanities at the University of Nebraska, Omaha. He specializes in the philosophy of neuroscience and cognitive science. His work focuses on the nature of structure-function mapping in the brain, the epistemology of neuroimaging experiments, and cognitive ontology. Prior to joining UNO, Joe received a Ph.D. in History and Philosophy of Science from the University of Pittsburgh and completed a postdoctoral fellowship in Washington University in St. Louis' Philosophy-Neuroscience-Psychology program