COMPUTATIONAL NEUROSCIENCE AND LOCALIZED NEURAL FUNCTION

DANIEL C. BURNSTON
TULANE UNIVERSITY, PHILOSOPHY DEPARTMENT

Penultimate Draft. Please cite final version: 10.1007/s11229-016-1099-8

“My point is that with each new case it is an empirical question whether these models, or models from some other theory, or no models from any theory at all will fit.”

-----Nancy Cartwright

“You should understand that if you’ve looked through one window,
You’ve looked through the general idea of a window,
Although this claim is entirely conditional and doesn’t apply to stained glass or lancets,
So you have to repeat the procedure repeatedly.”

-----Jaswinder Bolina

1. Introduction

In this paper I criticize a meta-theoretic view of functional localization in neuroscience, which I call “absolutism.” Absolutism is the idea that the right kind of function ascription to make to parts or areas of the brain is to posit a single, univocal function for each area.

> *Absolutism:* For any area of interest A, there is some univocal description, D, such that D describes the function of A in any particular context.

Absolutism is inspired by some powerful, but (I will argue) ultimately mistaken intuitions about what constitutes good explanation in the case of functional localization. The main intuition is that explanations should be highly general and predictive, and that such goals are undermined if the explanandum is genuinely context-sensitive. Hence, as per the definition above, we should want a single function attribution that describes what a given A does *whenever it functions*. In particular, I assess a form of absolutism that has gained considerable momentum in recent years, which I call “computational absolutism” (CA). Traditional ways of establishing neural function argued that the function of a brain area is to process a particular type of information, and/or contribute to a particular type of task. Recent physiological results, however, suggest that individual brain areas contribute to a number of distinct tasks, and process distinct types of information, depending on the perceptual or behavioral context. Computational absolutists consider this a bad result for functional localization, since it undermines generality and predictive power, and they hope to avoid this outcome by positing a different type of function ascription. They claim that, even if the information processed by brain areas differs with context, the brain area *performs the same type of operation*—namely, a particular type of computation—in each case. If CA is successful, then contextual variation is *not* part of the function of a brain area, but explained away as the output of the same, type-identical function being employed in different settings.

It is certainly not a priori that localization *must* be construed in absolutist terms. Recently, some theorists have begun to consider the possibility that contextual variation and localization are not incompatible, adopting "contextualist" views of function (Klein, 2012; McIntosh, 2004). Elsewhere (Burnston, 2015, 2016), I argue that contextualist accounts, appropriately construed, are not subject to the negative epistemic consequences that absolutists assume. Here I claim only that CA does not succeed at its goal of avoiding contextual variation in functional analysis. CA-theorists seek explanations of what an area does *intrinsically*—viz., regardless of contextually-varying external influence—that are themselves highly general and predictive. I argue that these goals are incompatible; any account general enough to abstract away from context is not likely to either successfully decompose the brain or to succeed at being highly predictive. I will make the point via a case study of computational modeling projects describing the function of perceptual area MT.

My strategy will be as follows. I will first argue against CA on its own terms. The current state of the field consists in a diverse set of models, involving different computational descriptions that account for distinct phenomena. So, there is no currently available univocal description of MT's computational function. Moreover, the field is not reasonably thought of as *trending* in the direction of a single, highly general description that covers all cases. The reasons I will give are familiar from the scientific modeling literature, particularly the perspectives advanced by Cartwright (1983, 1999) and Morrison (2000). Highly general models, of the sort that might reasonably be expected to apply in a wide variety of cases, do not in and of themselves explain or predict anything—instead they must be conjoined with specific causal claims about the situations being modeled. Since these causal factors are likely to vary with context, context cannot be eliminated from function ascription. I thus end by outlining a perspective on modeling projects in neuroscience inspired by the work of Cartwright and Morrison. While this will at best be a sketch, I claim that the view has the resources to adequately describe the field of computational neuroscience and the relative roles that contextual function attributions and general mathematical models play in giving functional explanations.

In section (2), I further describe absolutism and its motivations. In section (3), I discuss recent physiological results from area MT which suggest that its function varies with perceptual and task context. In section (4) I argue that the current state of the field in computational modeling of MT's function does not support a CA account, and in (5) that the field is not *trending* in the direction of a CA account. In section (6) I give a version of a Cartwright/Morrison account for computational modeling of neural function.

2. Traditional and Computational Absolutism

A function, for current purposes, is what a part of a system contributes to the overall behavior of the system of which it is a part (Cummins, 1975). Absolutism attempts to define each area's function according to *one* type of contribution, which is made in all contexts. Neuroscience has traditionally pursued absolutism of a particular form—namely, *representational* or *task* uniformity. On such accounts, which I call "traditional absolutism" (TA), neural areas are individuated in virtue of *representing* a specific kind of information, and/or *performing*, univocally, a particular kind of task: different brain areas process motion (area MT) and color information (V4), or are in charge of spatial memory (the hippocampus) or language comprehension (Wernicke's area). There is a close relationship between representing a specific kind of information and performing a particular task. As we will see, area MT is considered to contribute to tasks involving motion perception in virtue of representing motion. The language comprehension system, similarly, must work over representations of lexical constituents in order to contribute to lexical tasks. So, TA attempts to attribute univocal function ascriptions, and to do so by isolating a single representational or task description that can be assigned to the area in question.

Evidence against TA can come in the form of (i) showing that a given area processes multiple kinds of information, and/or contributes to multiple different tasks, or (ii) showing that the area is not in fact necessary for the task it is posited to perform. I will focus on type (i). Proponents of CA generally agree that this kind of evidence speaks against TA, but resist the conclusion that absolutism writ large is false. For instance, Anderson (2010) cites evidence from fMRI meta-analyses that the average cortical area is involved in nine distinct tasks. He then resorts to a CA-type description: he proposes a distinction between *working* and *use,* on which there is some computational process—an area's working—that it performs across any instance of its functioning or use, which may vary with context. So, while a given area may be involved in multiple tasks, it *does the same thing* in each. A working is "whatever single, relatively simple thing a local neural circuit does for or offers to all of the functional complexes of which the circuit is a part" (Anderson, 2010, p. 295).

Other CA theorists posit similar distinctions. Bergeron (2007) distinguishes *working* and *role,* where a working is a "sensory or motor operation" that is "basic" (p. 185). Rathkopf (2013) gives a more formal gloss. He argues that what is specific to a particular area is a *type of signal transform,* where this is to be read as a mathematical operation unique to the area. So, no matter what signal comes into an area (i.e., what information it is processing at a given time), the area applies the same, unique operation to that signal, and its output is the result of that operation. These accounts are still absolutist—they posit a univocal description that is supposed to describe what an area does in all contexts. It is just that the type of description has changed, from one couched in terms of information processing or task contribution to one describing a type of computation.

There are two specific motivations for positing CA, both inspired by absolutism writ large. The first is the desire to explain what a particular anatomical structure does. Rathkopf calls these "structurally oriented explanations"—explanations on which "the structural properties of the device … make sense" (p. 6). Price and Friston (2005) and Anderson (2010) make related claims, arguing that each part of the brain has a distinct anatomical structure, comprised of its pattern of internal and external connections, and thus should admit of a distinct function ascription. The goal, on CA, is to figure out what computation the fixed structure performs.

The second motivation is based on the ideal of absolutist explanations as highly projectable, predictive functional accounts of brain areas. This is supposedly impossible if contextualist variation, for instance in task contribution, is admitted into functional accounts. As Rathkopf complains, given contextual variability in the tasks to which an area contributes: "there could be no grounds for claiming of any list of task-bound functions that it is complete. … Any functional contribution to behaviors yet to be taken into account would demonstrate the incompleteness of the list" (2013, p. 10). The goal of functional localization, in standard absolutist terms, is to avoid these possible extensions and variations—to give as close to a "complete" theory of the brain area being described as possible. Price and Friston sum up the position well, sayingthat "it is most useful to label a region with a function that explains all patterns of activation" (2005, p. 268). If a successful CA account is posited, we will know, given that description, how the area will function in any particular context we investigate.[1]

I will argue that CA does not adequately describe the epistemology of computational neuroscience in one specific case: function attribution to perceptual area MT. Since perceptual areas are primary textbook success cases for functional localization, failure to account for localization in MT amounts to a major problem for CA. I will first, in section 3, discuss results in investigation of MT's function that undermine TA—that is, evidence that its function varies contextually. CA theorists, if they are right about localization, should be able to point to computational analysis of MT as providing a univocal function description despite this apparent variation. In the remainder of the paper I argue that this is not the case.

3. MT and TA

3.1. The Traditional TA Account of MT

One of the most successful TA-style accounts in neuroscience is the "modular functional hierarchy" view of the visual system (Burnston, 2015, 2016). The framework explains the functioning of the visual system via decomposition into specified areas, each of which functions to represent a specific kind of information. As Zeki (1978) describes it, each area "functionally specialized to analyze different features of the visual environment" (p. 423) such that "at every area a different type of information is analyzed" (p. 428). Early levels of the hierarchy represent simple features, such as luminance discontinuities (edges), displacement (a precursor to motion), and wavelength. Higher levels, such as those of "mid-level" vision, take these representations as input, and use them to extract more general features, such as motion, color, and form. Each of these features is represented in a specific area—motion in MT, color and form in distinct parts of V4.[2] Finally, different high-level visual areas have responses that represent particular shapes regardless of perspective, use motion to guide eye movements, and recognize objects.

---

[1] CA theorists differ in whether they describe their chosen operations quantitatively or qualitatively. I will focus on quantitative descriptions here, since in computational neuroscience most functional descriptions are given in quantitative terms (but see section 5). It should also be noted that these views have developed for each individual theorist over time—Friston (2010) gives a more quantitative reading than Price and Friston (2005) do, for instance. Anderson (2014), in more recent work, seems to soften his CA stance somewhat. I am attempting to focus on the idea of CA itself, so will not catalog these nuances here.

[2] While I have mostly discussed absolutism as a description of the function of areas, it can also apply to functional divisions within an area. In the visual system distinct parts of V1 are classically posited to process displacement, orientation, and wavelength information, and this continues in areas V2 and V3 (Livingstone & Hubel, 1988). It is compatible with TA to subdivide areas and give absolutist function ascriptions to each, but this consideration is not pertinent in the arguments I will make regarding MT. See note 4 below, and for more thorough discussion (Burnston, 2015).

MT's function is defined, according to the hierarchical view, by its specific role in motion perception. Individual MT cells have preferred directions of motion and preferred speeds, meaning that the cell will be most active when a stimulus moving at a particular direction and speed occurs within the cell's receptive field (the area of the visual field in which a stimulus can activate the cell). While there are cells sensitive to displacement in V1, at MT response properties to motion emerge that are distinct from elsewhere in the visual system. Particularly important is the ability to detect "pattern motion" in the stimulus. The cells in V1 that respond to displacement respond in a way that is ambiguous regarding the direction of motion, due to their small spatial receptive fields (the so-called "aperture problem"; see Bradley & Goyal, 2008, for a review). That is, the displacement they detect could be caused by an object moving in a number of different directions. Moreover, different parts of the same object, which moves in a single overall trajectory, may move in different directions. Pattern motion involves the ability to extract the overall direction of the object from the "component" motions of its parts. MT has different groups of cells that respond to component and pattern motion, which allow it to detect motion even in stimuli without clear object boundaries, including overlapping stimuli, transparent stimuli, and fields of moving dots. Importantly, MT responses are also correlated with perceptual *judgments* about pattern motion (Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996) —MT signals are thus taken to play a functional role in perceiving motion, particularly by projecting to the lateral interparietal area (LIP; Shadlen, Britten, Newsome, & Movshon, 1996), an area whose responses underlie perceptual decisions.

A final difference between MT and earlier parts of the visual hierarchy is that MT cells also exhibit motion "opponency," meaning that motions in directions other than their preferred direction (most strongly, the opposite direction) reduce, and sometimes cancel, their responses. MT's combined properties also differ from areas later on in the hierarchy, which have specific responses to non-linear types of motion (e.g., rotations and expansions in area MST; Van Essen & Gallant, 1994). Due to its specific place in the visual hierarchy, and its particular response properties, the standard view of MT is that it is "a motion area" (Zeki, 1978, p. 426)—that is, it has representing motion as its univocal function.

Altogether, MT is an incredible success story for functional explanation in neuroscience, and it is a textbook case of TA-style description. However, physiological analysis in the last 15 years has slowly uncovered much more diversity in MT responses than the traditional hierarchy view predicts; investigators have slowly come to accept that MT responses represent more than motion information, and are involved in more than motion-judgment tasks. I discuss one example of these results in the next section.

3.2. Evidence against a TA-Account of MT

Binocular disparity—basically, the angle formed by an object and the points on the two retinae where light reflecting from the object hits the eyes—is an important cue for *depth* perception. Objects at two different depths will exhibit two different disparities relative to a "plane of fixation," the depth at which the two eyes are focused. A wide range of results have suggested that, not only are MT responses influenced by disparity information, but that that information is relevant for depth perception. DeAngelis and Newsome (1999) found that a majority of individual MT cells have responses to particular preferred disparities—when a motion stimulus moved in the cell's preferred direction, its responses were greatest at a particular disparity, and far lesser for other disparities. Disparity-selective cells were organized in a regular way across MT, suggesting that MT conveys meaningful information about disparity.[3] Moreover, many cells responded to disparity even in *stationary* stimuli. DeAngelis, Cumming, and Newsome (1998) showed that these responses were predictive of depth judgments, even for stationary stimuli, thus suggesting that MT could represent depth even in the absence of motion.

Subsequent studies showed that the vast majority of MT cells (93%) have disparity selectivity (Uka & DeAngelis, 2003)[4], and that over a third of these respond to disparity even in the absence of motion (Palanca & DeAngelis, 2003). A number of other results also further established the relevance of MT disparity signals for depth judgments (Dodd et al., 2001; Grunewald et al., 2002). In all, the consensus that has developed is that MT is as much a depth area as a motion area.[5]

The results regarding MT and depth parallel the kinds of concerns that motivate CA theorists, as well as contextualists, to deny TA. Depending on the perceptual and behavioral context, MT responses seem to represent multiple different types of information, and contribute to distinct tasks (for a full argument in favor of this interpretation, see Burnston, 2015, 2016). If CA is the right account of functional localization, then computational analysis of MT should be involved in positing computations that both explain the distinctive structure of MT and explain how MT will respond in any given context—including depth contexts. In the remainder of the paper, I argue that this view is not substantiated by projects attempting to describe MT's function in computational terms.

4. Models of MT Function

In perceptual neuroscience, computational models posit a type of computation that a brain area performs. The model is "successful" for a range of stimuli if, for those stimuli, the behavior of modeled cells is similar to the response of real cells in similar situations. For instance, if the model predicts a certain level of increase in activation for certain stimuli, this should be reflected in the firing rates of actual cells in those conditions. Often, in MT, computational models attempt to capture the kinds of motion-selectivity properties discussed above—e.g., direction selectivity, pattern motion, and opponency. The process of constructing a model involves positing a type of computation, constructing a model of a brain area involving particular anatomical features (such as the inputs to the modeled area, its outputs, and/or its internal connections), designing the stimuli, and fitting the parameters of the model (such as the level of excitation for a certain stimulus property) to physiological data. Successful models explain the data by showing how the computation, performed over the input, produces the right kind of

---

[3] Influences of disparity information on MT responses had in fact been noticed long before, but a variety of reasons were evinced for considering this influence to not be functionally related to depth perception (Maunsell & Van Essen, 1983). As a result, the disparity sensitivity of MT was relegated to footnotes and asides. I discuss this development in more detail in (Burnston, 2015).

[4] This speaks against the suggestion that feature-specificity, and thus absolutism, might be saved by subdividing MT into motion- and depth-selective parts, as had been done for color and form responses in V4; see note 1.

[5] Investigation of MT and depth has proceeded considerably beyond these early results, further distinguishing between different types of disparity responses and showing that MT responses are relevant to multiple kinds of depth perception. I discuss these results in detail elsewhere (Burnston, 2015, 2016).

responses. Thus, the posited computation is a candidate to be the working or signal transform proposed as a functional description by CA theorists.

4.1. Modeling Motion Processing

The models of MT function that I will focus on start from a description of the signal that MT receives from lower areas of the visual hierarchy, particularly V1. In Adelson and Bergen's (1985) classic "motion energy" model, the idea is that displacement-sensitive cells in the early visual system provide an initial description of the spatiotemporal frequency of the stimulus. Spatiotemporal frequency can be determined by Fourier decomposition, which breaks down the overall stimulus into a power spectrum of frequencies in space and time. In the frequency domain, a direction-selective filter responds to particular frequencies in the spatio-temporal decomposition—they thus represent a particular "component," of the decomposed "motion-energy" in the stimulus, which can be thought of as a combination of spatial frequency, direction, and speed (Bradley & Goyal, 2008). However, simple component filters will not match the response properties of MT—they will not, for instance, exhibit pattern motion responses or opponency. As such, the consensus surrounding Adelson and Bergen's model is that it is a good description of "early" vision processes, such as those present at V1 (Nowlan & Sejnowski, 1995), and many MT models use motion-energy filters as *inputs* to modeled MT cells. Modeled MT responses are then intended to capture the particular features of motion perception—including pattern motion responses for a variety of stimuli, as well as opponency—attributed to MT, as compared to V1. The hope is that MT's function can be characterized in terms of "simple operations in the spatiotemporal frequency domain" (Bradley & Goyal, 2008, p. 686).

4.2. Summation and Normalization Models

Simoncelli and Heeger's (1998) "summation and normalization" model began with V1 inputs to MT consisting of motion-energy filters of the type described by Adelson and Bergen. Each modeled MT cell received input from specific modeled V1 cells—excitatory input from "preferred-direction" cells and inhibitory from other cells. MT cell responses were determined by subsequent summation and normalization of these signals. That is, the summation and normalization steps were taken to be the computation performed by MT over the signal coming from V1. At the MT stage of processing, the signals leaving the modeled V1 cells were added together—the summation step. In the normalization step, each individual V1 output was divided by the sum total. The result of the computation was to provide the same modification to each V1 output, which was determined by the population of those cells, hence the "normalization." MT cell responses were due to the normalized values of the V1 cells that provided them with input. The summation and normalization procedure, then, purported to model the computation performed by MT—it modeled the transformation that occurs in a motion signal between the V1 and MT stage of the motion stream, which in turn produces the responses of particular MT cells. The "decision" made by MT was taken to be the peak activation of the population of MT cells—i.e., the most active cells were the ones whose direction and speed preferences determined the outcome.

The model predicted a variety of response properties characteristic of MT. The aperture problem was overcome by having MT cells receive input from particular combinations of V1 cells. While one V1 cell is ambiguous for direction of motion, having a cell take input from multiple V1 cells can disambiguate the direction (called the "intersection of constraints" solution; Bradley & Goyal, 2008). The modeled MT cells could therefore represent pattern motion, which they did successfully for a range of drifting grating and dot stimuli. Finally, the cells showed opponency both due to normalization from the subtractive stage and from inhibition from V1 cells selective for other directions. These combined properties allowed the model to account for a variety of opponency effects, including the fact that motion opponency occurs most significantly for motion in the opposite direction to cells preferred directions, at approximately the cell's preferred speed (Snowden, Treue, Erickson, & Andersen, 1991).

4.2 Bayesian Models

Koechlin, Anton, and Bernard (1999) also modeled MT cells as operating over V1 inputs tuned for spatio-temporal frequency, but there were two key differences between their model and Simoncelli and Heeger's. First, they used a different computational principle, Bayesian inference. Second, they explicitly modeled lateral connections between MT cells: each cell received excitatory input from cells with similar direction preferences whose receptive fields were in adjacent parts of the visual field; each cell also received inhibitory input from cells covering the same part of the visual field but with distinct direction preferences. What made the model Bayesian was that it employed an activation equation for each cell that involved a "prior" probability for each particular velocity and direction (and for those in adjacent parts of the visual field, which set the excitatory connections). This was set by the training-history used to fit the model—the more a cell "saw" motion in a particular direction and velocity, the higher the prior probability it applied to that type of stimulus. The responses of cells to a test stimulus were determined by an application of Bayes' rule, which gave a posterior probability for a particular direction and velocity given the prior and the actual stimulus. Put simply: the cells fired most whose priors most closely matched the occurring stimulus. This allowed for the Bayesian model to capture direction and speed selectivity.

The lateral connections allowed for solving the aperture problem and inhibiting opponency. Excitatory connections from adjacent parts of the visual field could disambiguate which direction cells should be signaling. Similarly, motion in the opposite direction from a particular cell's preferred direction would inhibit it due to subtractive influence from the cells currently encoding the stimulus. These properties, while implemented differently, accounted for a similar set of responses as the Simoncelli and Heeger model. There was one major difference, however. Due to the lateral excitation, the motion signals "propagated" along the moving stimulus—e.g., with units in the middle of a moving plane or along the edge of a moving bar becoming more active—and did so with particular temporal patterns that matched psychophysical responses for estimating the velocity of bars of varying lengths. This result was taken to be important for understanding the perception of elongated moving objects, and was unique to the Koechlin et al. study.

4.3. Competition Models

Krekelberg and Albright (2005) set out to physiologically measure and model responses of MT cells to apparent motion stimuli, including "phi"-motion and "reverse-phi" motion. Apparent motion phenomena are motion percepts in which discrete spatio-temporal displacements of a stimulus are perceived as continuous motion. Phi- and reverse-phi stimuli are diagrammed below.

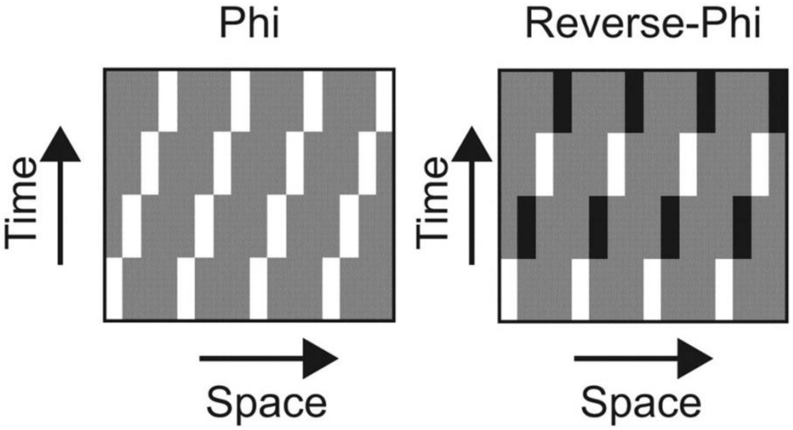

Figure 1. The phi- and reverse-phi stimuli from Krekelberg and Albright (2005).

Each panel shows a moving grating over several time steps. In the left panel, the grating moves to the right, and this stimulus produces a phi-motion perception to the right. In the reverse-phi stimulus in the right panel, the displacement is the same, but the grating switches luminance contrast (i.e., going from white to black, or vice versa) at each time step. In this stimulus, perceived motion is generally to the left. To see this, focus on the far right bar in the first time step of the figure on the right. In the second time step, subjects perceive the stimulus as though this bar had moved to the left, and was now the black bar second from the right. This perception matches a shift in the direction of the predominant Fourier energy in the stimulus, although there are still some components exhibiting rightward motion. Interestingly, monkey MT cells *also* switch their direction preferences in the reverse-phi stimulus. This means that standard “subtraction" models of opponency are inadequate: if cells really responded only to motion energy in one direction, and opponency occurred to motion energy in the opposite direction simpliciter, then MT cells should not exhibit strong responses to the reverse-phi stimulus. Both Simoncelli and Heeger’s and Koechlin et al.’s models involves subtractive opponency, and therefore can’t account for this result.

As an alternative, Krekelberg and Albright propose a “competition model,” which includes an excitatory *and* an inhibitory weight for each Fourier component in the studied stimuli. Each cell was characterized by particular weights to each particular component. The combined positive and negative weights could account for the reverse-phi results, since the weights to multiple components could be combined in a variety of ratios. Consider the following graph, in which *P* and *p* are components with spatio-temporal energy moving in the cell’s preferred direction; *A* and *a* in the opposite direction, and *f* and *s* in other directions. If component *A* is paired with *P*, a significant inhibition will result, since the overall effect of inhibition is much greater. However, if *a* is combined with *P*, little change will occur. So, if in a reverse-phi stimulus the component that switches directions approximates the spatio-temporal frequency of *a*, the physiological response to the stimulus will remain strong.

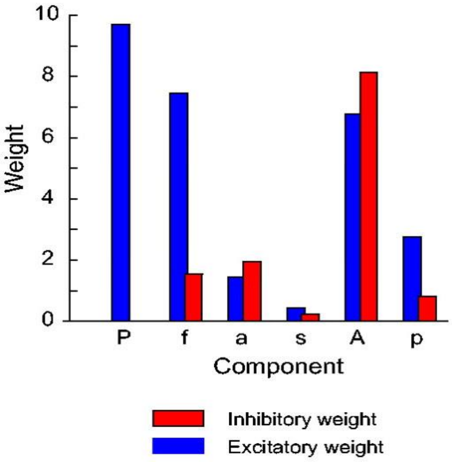

Figure 2. Competition model results for a single MT cell. From Krekelberg and Albright, 2005.

5. The Status of CA

CA is motivated both by the intuition that a successful model will explain the particular anatomical structure found in a particular part of the brain, and by the ideal of general predictive power. In this section I argue that neither of these claims is substantiated by the state of the field.

With regards to the anatomical motivation, the first thing to note is simply that no model discussed attempts to describe all of the explanatorily relevant features of MT's anatomy (defined, you'll recall, by the combination of its internal wiring patterns and its connections to other areas). Further, I can find no arguments in the literature suggesting that the models are explaining *the* distinctive anatomical structure of MT, as compared to other brain areas. While some structure is always modeled, the structure that is posited is what is required to understand why particular physiological behaviors come about, and not an attempt to explain the entire or the distinctive structure of the area. Thus, even if we grant the assumption that each brain area has a unique structure, it is highly questionable that computational models are in the business, primarily, of explaining that structure.

To flesh out this claim in more detail, let's return to the specifics. Simoncelli and Heeger's model explicitly builds in only relatively few anatomical facts, focusing mainly on input from V1 cells. It does not model the distinction between component and pattern motion cells in MT, nor does it model an output stage. This is in contradistinction to other models, which build in these architectural facts. Nowlan and Sejnowski's (1995) model, for instance, while also a summation/normalization model, does include these aspects, and in virtue of doing so can account for some motion phenomena that Simoncelli and Heeger's cannot (namely, stimuli for which both component and pattern cell responses are important). Koechlin et al.'s (1999) model builds in further architecture still, for instance explicitly modeling excitatory and inhibitory connections between MT cells, and use these properties to account for other phenomena. The Krekelberg and Albright model measures no MT architecture explicitly *at all*. It simply assumes that the different Fourier components in the stimulus are delivered as input to MT cells *somehow*.

The argument, then, is that modeling projects are tied to particular expananda, and it is not vital to their explanatory success to model some particular amount or specificity of structure. This is most clear in Krekelberg and Albright's model and its discoveries about inhibition in MT. The fact that each MT cell has both positive and negative weights for each Fourier component is a *major* revision to the traditional thinking about inhibition in motion-energy models. And the discovery was based on very minimal assumptions about architecture. But if this is the case, then the explanatory value of Krekelberg and Albright's model is *not* based on its modeling any specific structural aspects, but is instead a principle of MT function that holds independently of how the detailed structural/anatomical story comes out.

There is a similar diversity in terms of the actual phenomena being modeled. While there is significant overlap in terms of the MT responses that each model can account for, there are also differences. The Bayesian model, in virtue of its combined computational and anatomical assumptions, can explain pattern-spreading phenomena that the other models miss. Simoncelli and Heeger's model accounts for some opponency phenomena, not all. Krekelberg and Albright's model captures facts about opponency missed by other models, and can explain reverse-phi phenomena, but is geared specifically to opponency phenomena and not to all MT responses (see below). But this variation of successful model type with explanandum is a large problem for the CA theorist, since CA posits that a brain area performs *one* type of computation in all contexts, not different ones in different contexts. If the distinct models are taken as literally true descriptions of a brain area's function, then it seems that the computational function of the area does in fact change with context. But the CA theorist can also not retreat to a more pragmatic account of the role of these models. As Cartwright (1983) has argued, the presence of a variety of distinct theoretical descriptions, whose explanatory value depends on the situation being modeled, militates against taking one particular description as *the* correct one for the system being studied. Again, however, CA theorists are committed to giving precisely this sort of description. The current state of the field thus does not support the absolutist aims of CA.

For the remainder of this section, I will take the current variation in types of computations posited by successful MT models as read, and focus on possible responses that a CA theorist could offer. These responses rely, respectively, on modifying the level of description at which explanations are couched, or on stressing the need for conceptual unification in modeling.

5.1. Level of Description

The first response on behalf of CA is to suggest that positing mathematical functions is not the right level of description. Instead, we should focus on the level of syntax, the particular program implemented by a brain area. A given mathematical function, of course, can be implemented by a variety of syntactic programs, and it might be at this level of description that we can isolate a unique function for a specific area. Piccinini (2008), for instance, suggests that we should not individuate computations in terms of semantic notions, but instead purely in terms of their computational function—how they contribute to the overall syntactic processing of the computer in which they operate.

There are a variety of reasons to be skeptical about this move, at least in terms of neuroscientific explanation (Piccinini is primarily interested in how to *define* computation, which is not my concern here). As Shagrir (2001) notes, a physical/anatomical structure of any significant complexity is likely to be able to execute a large number of distinct syntactic programs at once—the more complex the structure, the more possible programs implemented. As the foregoing discussion contests, MT is pretty complex. Both its structure and its computational functioning can be and are described in different ways depending on the phenomenon being studied. Shagrir's own proposal is that when positing syntactic structures we index them according to the particular tasks they perform. In the current setting, we can read this as the view that we can only understand the syntactic programming of a particular part of the brain in terms of how we semantically describe the function it computes. But if this is the case, then our view of the syntax will change with context, exactly what a CA theorist cannot accept.

The situation is in fact no better, however, if we follow Piccinini's non-semantic proposal for how to individuate syntax. He suggests that we look at the *other* processes contributed to by a processor, and individuate the program that it runs according to those contributions. He calls this "wide functional individuation," and in the brain case it can be thought to correspond to the relationship between a syntactic program at one area and the programs at other areas to which it provides input. Alas, this is no good for the CA theorist either. Recall the quote from Anderson in section 1: the CA theorist wants to posit *one* computational working that makes the *same* contribution to all the "functional complexes" it serves. If we must re-describe the syntax when we talk about distinct contributions to further computational processes (viz., other parts of the brain), then CA's promise of a univocal description is undermined. Piccinini sums this point up well himself: "we cannot determine which computation is explanatory within a context without looking outside the mechanism" (2008, p. 231). To the extent, then, that we individuate programs in contexts, evidence of contextual variation in function can be construed as evidence for contextual variation in computation, even at the syntactic level.

Now, as Piccinini notes, there is at least a conceptual possibility that one and the same program could be given different mathematical/semantic interpretations in different contexts, so a CA-theorist might hope for this kind of theory of brain areas. This is really a slim reed to grasp at, however. As Piccinini points out, it may be "difficult or impossible" (2008, p. 216) in complex computers, to specify the syntax without first alluding to a semantic characterization of the operation being performed. One might just hope that, given a complete wiring diagram of (e.g.) MT, and a complete list of its physiological responses, we could describe its entire machine table. However, this would be unhelpful for a variety of reasons. First, part of the appeal of CA theories is that they have a "conceptual compactness" (Rathkopf, 2013, p. 10), one that is supposed to be lost by importing context. Simply listing the whole gamut of syntactic operations performed by a brain area, however, seems to offer little in the way of this compactness, compared with a description of a type of mathematical function. And lastly, the above discussion showed that computational explanation in neuroscience, as currently constructed anyway, is not primarily in the business of giving complete anatomical-computational accounts of each brain area. So, the levels-of-description response seems to offer little help to the CA theorist.

5.2. Unification

The most promising response for the CA theorist is to claim that, while no model *currently* accounts for all results, further model development may produce one that does. In particular, CA theorists might rest their hopes on theoretical unification: some model may prove to be the best at incorporating the results of other models, importing more architectural facts, and generalizing to the widest range of phenomena. For instance, while the summation/normalization model given by Simoncelli and Heeger doesn't account for all phenomena regarding MT function, it might be expanded to include competition between Fourier components (as in the Krekelberg and Albright model), to model the distinction between component and pattern cells (as in the Nowlan and Sejnowski model), and to include lateral connections between cells (as in the Koechlin et al. model). These kinds of development might allow the model to account for instances of MT functioning currently outside of its purview. Computational neuroscientists do sometimes express unificationist leanings—they often tout the generality and flexibility of their models, and express the hope that with development more architectural facts and phenomena can fall under their particular model's umbrella. The CA theorist might expand this sentiment to suggest that one computational description will prove to be the best at this kind of unification. As such, let's refer to it as the "unificationist gambit." The gambit is suggested by some statements made by CA theorists, particularly in Price and Friston's (2005) expressed desire for a single description that "encompasses" anything that an area might do in distinct contexts, as well as Friston's (2010) subsequent claim that specific functional accounts should fall under "global" theories.

Two points before moving on to assess this idea. First, the gambit can't just be wishful thinking. We need to discuss whether this kind of unification is something we can expect given the nature and status of the field. Second, there is a limit to how abstractly and generally one can pitch one's computational descriptions while still succeeding at *localizing* function. Consider: a significant trend in computational theorizing about the brain involves the search for so-called "canonical computations"—computations that *every* functional unit (described at some level of grain) in the brain performs. In fact, several of the types of models discussed above are candidates for this kind of description. It has been contended in different quarters that perhaps *all* brain processes perform summation and normalization (Carandini & Heeger, 2009), or Bayesian inference (Friston, 2010). Functional decomposition, however, seeks to *differentiate* neural areas in terms of *distinct descriptions of their functions*. If, hypothetically, it turns out to be the case that *all* neural computations are best seen as implementing summation and normalization computations, then summation and normalization descriptions, as such, will fail to describe what (for instance) *MT* does, uniquely.[6]

This is only to say that highly general computational models *on their own* are unlikely to ground functional localization. It is not to say that they cannot in principle be helpful in differentiating functions. It might be that different areas employ canonical computations in distinct ways, or in distinct combinations—i.e., there might be something about how and when specifically MT implements (for instance) a summation/normalization computation that distinguishes MT from the other areas that do so. My point in the remainder of this section is that, viewed this way,

---

[6] Interestingly, Rathkopf (2013) raises this kind of consideration against abstract versions of TA, but does not consider the possibility that a similar concern might apply to his own version of CA. See Chirimuuta (2014) for a more general discussion of canonical computations.

even models that generalize over a wide range of cases fail to explain and predict what MT will do in all contexts.

Let's consider an example. Nishimoto and Gallant (2011) sought to generalize and extend the account of receptive field and pattern-motion properties of MT cells originally proposed by Simoncelli and Heeger. Their model similarly involved input to particular modeled MT cells from particular sets of V1 neurons, and also employed stages of summation and divisive normalization (although the details were slightly different). They had monkeys watch movies of naturalistic scenes (e.g., of animals moving across a prairie), and recorded MT responses across the entire stimulus set.[7] They thus hoped to explore MT responses to, if not the entire Fourier domain, at least an unbiased sample of it. Their model was fit to MT responses over a range of movies, and then tested for a variety of receptive field properties. In particular, Nishimoto and Gallant's MT cells showed a physiologically realistic range of pattern motion responses, suggesting that a summation/normalization model can capture physiological responses in both constrained experimental settings and naturalistic stimuli—the authors expressly phrased the result as a successful generalization of the notion of summation/normalization to a wider range of stimuli and receptive field properties (Nishimoto & Gallant, 2011, p. 14554).

This is undoubtedly an interesting result. When seen in the light of the discussion above, however, it is clear that the model is not successful in explaining and predicting what MT will do in all contexts. Even the "naturalistic" stimuli are two-dimensional, so the model does not say anything about the sorts of robust and functionally relevant responses to depth that have also been discovered in MT. Not only does Nishimoto and Gallant's model not predict any of these effects, but it is not intended to—it is a study of motion receptive fields and pattern selectivity for motion stimuli that attempts to extend models of *those* responses to cover naturalistic stimuli. Understanding MT function as a whole, however, involves not just understanding its responses in motion contexts, but also in depth contexts. As such, having a successful model for even a relatively unconstrained set of two-dimensional, black and white motion stimuli simply doesn't provide an explanation or prediction of what will happen in further contexts.[8]

The natural response on behalf of the CA theorist is to double down on the gambit by suggesting that with enough study and modification of the model, depth results can be incorporated. This, alas, gives the game away. What one would have to do in order to extend Nishimoto and Gallant's model to account for how MT represents depth is to pursue physiological analysis of the responses of MT in depth contexts, as well as the anatomical connections which underlie that

---

[7] The movies were not entirely naturalistic—they were "motion-enhanced", where the "enhancement" consisted of the random insertion of textured objects moving across the screen. The enhancement constrained the movies to meet certain statistics for spatial frequency, which I will not discuss in detail here. It should also be noted that the standards for interpreting model success for naturalistic stimuli are generally different—models for these stimuli generally account for less variance in responses than for more controlled stimuli. I will gloss over this detail as well.

[8] Nor would it be helpful to model "totally naturalistic" stimuli, which contained all possible stimulus elements. Understanding cell responses to such stimuli would require knowing the statistics of the relevant aspects of the stimulus, and this is often difficult to discern in fully unconstrained settings. Drawing conclusions in such studies often requires having a model already in hand of what the cell is responding to (as was, in fact, the case for Nishimoto and Gallant's)—but, as already shown, a previous understanding of what cells do can be overturned in new contexts. Rust and Movshon (2005) discuss these points in an amusingly titled article, "In praise of artifice," which criticizes overly optimistic use of "naturalistic" stimuli in computational neuroscience.

influence, and then build parameters accounting for depth response into the model. That is, in order for the model to be both general *and* predictive, we would have to explore new contexts with physiological methods and modify the model to suit them. This fact is often noted by those who tout the potential generalizability of their models.[9]

In my view, this shows the unification gambit, and CA more generally, to offer only a rather empty promise. Recall, CA theorists take contextual variation to be inherently incompatible with functional localization, on largely pragmatic grounds—we cannot have good theories of localization, they argue, if we must continually explore distinct contexts and be willing to modify our functional accounts accordingly. But armed with only highly general descriptions of an area's function, this is what we will have to do in order to predict how a brain area will function in a given context. Morrison (2000) has made a similar point about highly general models: if we must build in new parameters to account for every fresh context, then there is no deep explanation or unification provided *by the model itself*. If this is correct, then the computational descriptions on offer in the unificationist gambit neither explain nor predict what brain areas do in all contexts.

It is worth pausing to note that this conclusion is reflected in Krekelberg and Albright's analysis of their competition model. The general lesson they draw from their study is that one cannot predict the responses of MT cells to multiple Fourier components in a stimulus just by knowing their responses to the individual ones. Fourier components need to be both measured individually and measured in combination. Krekelberg and Albright point out that they studied only a few Fourier components out of the entire set of possibilities, and that the cells they studied don't exhibit a clear pattern of inhibition and excitation in the components, instead being uniquely characterized. One of the things we want to know, then, is whether there *are* patterns of inhibition and excitation to different components across populations of MT cells, and how these relate to motion perception. The point is that the competition model, fit to the current data, *does not say anything about these patterns*. They write: "Sampling only 6 of the relevant points in [the Fourier domain] makes finding a common pattern of competitive interactions [shared amongst different cells] unlikely. To uncover a common pattern, we believe it to be necessary to measure interactions across a large part of Fourier space" (Krekelberg & Albright, 2005, p. 2920).[10] That is, to understand what MT cells do, neuroscientists have to look for a pattern in the *way their behavior tends to vary* as the stimulus changes: they need to look at contextual variation. Note that this is true *despite* the success of the model at fitting the data that was actually studied and its promise as a way of explaining reverse-phi phenomena. While the competition model provides a way of analyzing the responses to any particular combination of

---

[9] Both Koechlin et al. (1999, p. 40) and Simoncelli and Heeger (1998, p. 756) admit that this further investigation and explicit addition would be necessary for generalizing their models beyond motion contexts. Tellingly, despite expressly calling their model a generalization, Nishimoto and Gallant do not even mention depth in their discussion of motion-selectivity properties, other than to note that no disparity is present in their stimuli and therefore is absent from the modeled responses.

[10] Krekelberg and Albright note that there are pragmatic problems with this suggestion—namely the "combinatorial explosion" of needing to look at all of the Fourier components and their combination. Inevitably, this will involve some exploratory investigation. In other work (Burnston, 2015, 2016) I discuss how contextualist function ascriptions can be used to constrain search through very complex sets of possible contexts, and argue that this kind of search is generally systematic and intelligible despite not attempting a "complete" account of an area's function.

Fourier components, it does not *of itself* provide the explanatory principles for understanding MT function as a whole. Nor is it intended to.

To sum up: CA theorists contend that context is inimical to functional localization. The proposed advantage for CA is that it hopes to give accounts that explain and predict what will occur in all contexts. I have argued that this is a false hope—even given a highly general model, we must still investigate and describe detailed contextual variations in behavior before understanding the function of areas like MT. Alas, I can't prove the negative existential: there is no way to show that it is *impossible* that we could have one model that accounts for all and *only* everything that MT does, is predictive in the way that CA theorists demand, and is so in a way that does not require simply adding in new parameters to the model for each modeled situation. However, I think the situation is tenuous enough that we should be motivated to consider alternatives for understanding how computational description—especially highly generalized models—contribute to functional decomposition. In the final section, I sketch such a view.

6. Conclusion: Explanatory Unification versus Tools for Contextual Analysis

I have argued that CA is not substantiated by modeling projects analyzing MT. How, then, do we construe the role of computational modeling in functional localization? Denying computational descriptions pride-of-place in localization is not, or should not be, to decry their importance writ large. I have already noted some similarities between my characterization of neuroscientific models and Cartwright and Morrison's arguments about modeling in physics. Both carve out important places for general theoretical descriptions in giving explanations, but deny any explanatory primacy for those descriptions, as opposed to detailed descriptions of distinct contexts or phenomena. While their projects are different in focus and emphasis, I take both to be committed to the following idea: the usefulness of highly generalizable quantitative models in the sciences is not predicated on their providing an explanation of what individual physical systems do. Understanding their role requires dissociating generalization and unifying power from explaining what occurs in particular contexts. While I cannot flesh out the account in full, I want to close by suggesting that this perspective is fruitful if applied to the role of computational description in localizing function. I will unimaginatively call it the "CM account" of modeling projects in neuroscience.

The CM account considers computational descriptions to be highly general tools that are useful for modeling particular instances of neural function—useful in the sense that they provide frameworks for generating quantitative predictions for particular situations. Computational descriptions such as Bayesian inference, summation/normalization, and competition are descriptions that can, at least in principle, be applied to any number of instances in which a brain area processes information. Unlike CA, however, the CM view argues that these general descriptions only ever generate predictions for specific cases when wedded to specific knowledge about the kinds of informational functions being performed by that area—i.e., the kinds of functional descriptions that are likely to vary with context. This position gives *an* important role to computational description in generating explanations and predictions, but straightforwardly denies that a single functional description of a brain area can hold constant despite change in context. The reason is that a given type of computation, on its own, is insufficient to say much of anything about what a particular brain area does. In order to describe the function of a brain area, even in computational terms, we must explore, articulate, and model how the area processes information in distinct contexts.

I think this view has all kinds of advantages, primarily stemming from its ability to explain the state of the field in modeling MT. The problem for CA consists in the fact that there is a diversity of models accounting for different aspects of MT function, and that attempting to generalize beyond this removes the kind of explanatory power on which CA seeks to hang its hat. On the CM account, however, we should precisely expect a range of models each serving a variety of functions—sometimes attempting to generalize over a range of phenomena, sometimes to incorporate new anatomical facts, sometimes to account for physiological functions without detailed anatomical knowledge, and sometimes to model very specific phenomena. The CM account also incorporates perfectly a variety of features that should seem appealing in light of the above discussion. It captures the importance of computational models without placing an explanatory burden on them that (if the arguments above are correct) they cannot fulfill. It also captures neuroscientists' search for highly general models in a way that does not make them irrelevant to questions of localization, but also does not saddle them with the sole burden of making specific predictions in every case. It can hope to account for the range of attitudes modelers take to their models: sometimes touting their generalizability, sometimes their specific structure as used in describing a specific case, etc. Importantly, this includes attitudes in which modelers take their distinct approaches to be competing. There is certainly room in the field (as the CM account construes it) for debate about which particular model or set of models is best for saving some particular neural phenomenon. It seems to me that all of this is exactly as it should be, and that nothing is left out by the CM account.

Having this perspective on the table is useful, since it gives us an alternative to CA for thinking about other modeling projects. There are very successful models of other parts of the brain: for instance, the theory that V1 neurons incorporate a range of Gabor filters for performing local Fourier analysis on the information represented at the retina (Olshausen & Field, 2009). CM suggests the following general set of tendencies for modeling projects. Models that are constructed based on responses to particular stimuli are unlikely to generalize too far beyond those stimuli. Models that are very general are unlikely to explain and predict everything a brain area is capable of doing without adding a lot of extra context-specific information. In each case, we should expect neuroscientific progress both in seeking general models and in pursuing detailed, contextually-varying descriptions.

Gabor models of V1 are built to capture its responses to individually presented contrast gratings. As Chirimuuta and Gold (2009) note, Gabor models, with their assumption of fixed linear receptive fields for V1 neurons, fail to account for a wide range of results suggesting that receptive fields are actually stimulus and network-dependent—i.e., their functional responses change depending on context. While I can't undertake detailed discussion here, Chirimuuta and Gold lay out the dialectic surrounding this situation, and it takes a shape that is very predictable from a CM point of view. Some contend that we should abstract to a canonical computation, perhaps involving normalization, to account for non-classical receptive field properties, but these models have not succeeded at explaining a variety of effects, which might require greater extensions beyond the notion of a fixed receptive field. Others propose "piecemeal" solutions, suggesting that models must be restricted to particular contexts. Chirimuuta and Gold, however, suggest that this may miss facts about V1 neurons in general, and express hope for a unifying general model. On CM, this is not a choice that must be made one way or another. Both generalizing and piecemeal solutions are important, and it is unlikely that either will explain everything there is to know about V1's function.

Unifying perspectives are likely to be particularly valuable in cases where we know the function of an area to be very diverse. This is the case in Levy et al.'s (2005) discussion of the hippocampus, which Rathkopf (2013) proposes as implementing a CA type account. On Levy et al.'s view, the hippocampus serves as a "recoder," which searches for associations across spatially and temporally extended representations in other brain areas. While they tout the unifying power of their theory, they explicitly say that their view consists of a "family of models"—the recoding idea allows for the generation of a number of much more specific network models (sometimes involving different basic assumptions about the networks involved) engineered to account for specific behaviors. Moreover, the networks themselves involve large numbers of cells that are meant to encode "local context," and the function of these cells changes as tasks vary. The model generates predictions via positing random search defined over these particular elements. Thus, despite their touted unification, the authors do not seem to be positing a final explanation of the function of the hippocampus. Their view of unification "[does] not propose that the recoding perspective is the only correct one" (2005, p. 1256). Rather, they suggest that their family of models serves "as a unifying theme and as a compass to direct research questions" (2005, p. 1256).[11]

Much more would need to be said to give full descriptions of these particular cases, and there are of course many other modeling projects in neuroscience that I've not discussed. But combined with the discussion of MT, I think the foregoing helps to show that CM offers a cogent, productive view of computational neuroscience, which is more descriptively adequate than CA. The burden, at this point, should fall heavily on the CA theorist to give us a reason to adopt the weighty responsibility given to computational descriptions on the CA view, rather than the more modest tenets of the CM one. Thus, while the CM perspective I have advanced makes ample room for computational modeling in explanation, it is not meant to be particularly conciliatory towards CA. I contend that absolutism, in any form, is false. CA theorists will be left unhappy here, because I have not said anything to dissuade them from their absolutist fears of contextual variation. Elsewhere, however, I argue that these fears are unfounded. One can have a contextualist theory of localization that succeeds in meeting the aims of functional decomposition in the brain. For now, I merely hope to have pushed the question towards that larger platform.

References

---

[11] I would like to point out that while pragmatic virtue is central to the CM view of models, the view is not *necessarily* instrumentalist, or at least not purely so. For instance, even if a canonical computation were agreed upon tomorrow (which I find doubtful, but not impossible), this would not change the point about explanation in specific cases at all. Put briefly, even if it were "true" in some deep sense that all brain areas fundamentally perform (say) summation and normalization, the point about needing to incorporate contextual knowledge to explain what MT specifically does would hold unadulterated. This requires much more discussion than I can give it here, however (c.f. Cartwright, 1999).

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*(2), 284-299.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *The Behavioral and Brain Sciences, 33*(4), 245-266; discussion 266-313. doi: 10.1017/S0140525X10000853

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*: MIT Press.

Bergeron, V. (2007). Anatomical and functional modularity in cognitive science: Shifting the focus. *Philosophical Psychology, 20*(2), 175-195.

Bradley, D. C., & Goyal, M. S. (2008). Velocity computation in the primate visual system. *Nature Reviews Neuroscience, 9*(9), 686-695.

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience, 13*, 87-100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience, 12*(12), 4745-4765.

Burnston, D. C. (2015). *Perceptual context and the nature of neural function.* (PhD Dissertation), University of California, San Diego.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. *Biology & Philosophy, 31*(4), 527-550.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51-62. doi: 10.1038/nrn3136

Cartwright, N. (1983). *How the laws of physics lie*. Cambridge University Press.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*: Cambridge University Press.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese, 191*(2), 127-153. doi: 10.1007/s11229-013-0369-y

Chirimuuta, M., & Gold, I. (2009). The embedded neuron, the enactive field? In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience*: Oxford University Press.

Cummins, R. C. (1975). Functional analysis. *Journal of Philosophy, 72*(20), 741-765.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394*(6694), 677-680.

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *The Journal of Neuroscience, 19*(4), 1398-1415.

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, a. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 21*(13), 4809-4821.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Grunewald, A., Bradley, D. C., & Andersen, R. A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *The Journal of Neuroscience, 22*(14), 6195-6207.

Klein, C. (2012). Cognitive ontology and region- versus network-oriented analyses. *Philosophy of Science, 79*(5), 952-960.

Koechlin, E., Anton, J. L., & Burnod, Y. (1999). Bayesian inference in populations of cortical neurons: a model of motion integration and segmentation in area MT. *Biological Cybernetics, 80*(1), 25-44. doi: 10.1007/s004220050502

Krekelberg, B., & Albright, T. D. (2005). Motion mechanisms in macaque MT. *Journal of Neurophysiology*, 93(5), 2908-2921. doi: 10.1152/jn.00473.2004.

Levy, W. B., Hocking, A. B., & Wu, X. (2005). Interpreting hippocampal function as recoding and forecasting. *Neural Networks, 18*(9), 1242-1264.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240*(4853), 740-749.

Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49*(5), 1148-1167.

McIntosh, A. R. (2004). Contexts and catalysts: A resolution of the localization and integration of function in the brain. *Neuroinformatics, 2*(2), 175-182.

Morrison, M. (2000). *Unifying scientific theories: Physical concepts and mathematical structures*. Cambridge University Press.

Nishimoto, S., & Gallant, J. L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *The Journal of Neuroscience, 31*(41), 14551-14564.

Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *The Journal of Neuroscience, 15*(2), 1195-1214.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*(23), 3311-3325.

Palanca, B. J. A., & DeAngelis, G. C. (2003). Macaque middle temporal neurons signal depth in the absence of motion. *The Journal of Neuroscience, 23*(20), 7647-7658.

Piccinini, G. (2008). Computation without representation. *Philosophical Studies, 137*(2), 205-241.

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

Rathkopf, C. A. (2013). Localization and intrinsic function. *Philosophy of Science, 80*(1), 1-21.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8*(12), 1647-1650. doi: 10.1038/nn1606

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience, 16*(4), 1486-1510.

Shagrir, O. (2001). Content, computation and externalism. *Mind, 110*(438), 369-400.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research, 38*(5), 743-761.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *The Journal of Neuroscience, 11*(9), 2768-2785.

Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *The Journal of Neuroscience, 23*(8), 3515-3530.

Van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13*(1), 1-10.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature, 274*(5670), 423-428.