Beyond Componential Constitution in the Brain

Starburst Amacrine Cells and Enabling Constraints

Michael L. Anderson

Componential mechanism (Craver 2008) is an increasingly influential framework for understanding the norms of good explanation in neuroscience and beyond. Componential mechanism "construes explanation as a matter of decomposing systems into their parts and showing how those parts are organized together in such a way as to exhibit the explanandum phenomenon" (Craver 2008, p. 109). Although this clearly describes some instances of successful explanation, I argue here that as currently formulated the framework is too narrow to capture the full range of good mechanistic explanations in the neurosciences. The centerpiece of this essay is a case study of Starburst Amacrine Cells—a type of motion-sensitive cell in mammalian retina—for which function emerges from structure in a way that appears to violate the conditions specified by componential mechanism as currently conceived. I argue that the case of Starburst Amacrine Cells should move us to replace the notion of mechanistic componential constitution with a more general notion of enabling constraint. Introducing enabling constraints as a conceptual tool will allow us to capture and appropriately characterize a wider class of structure-function relationships in the brain and elsewhere.

Keywords
Componential constitution | Constitution | Constraint | Enabling constraint | Explanation | Functional levels | Levels | Mechanisms | Mechanistic explanation | Neuroscientific explanation | Spatial levels | Starburst amacrine cells | Structure function mapping

Author

Michael L. Anderson
michael.anderson@fandm.edu
Franklin & Marshall College
Lancaster, PA, U.S.A.

Commentator

Axel Kohler
axelkohler@web.de
Universität Osnabrück
Osnabrück, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

How, in the brain or any other system, does specific function arise from underlying structure? The question is a general one, and also in some sense a vague one, for it asks simultaneously about how structures shape events—generate causes—and also about what kinds of explanations one should aim for in neuroscience. Here I will focus on the second question in the hope of partially illuminating the first. One increasingly influential class of answers to this second question "construes explanation as a matter of decomposing systems into their parts and showing how those parts are organized together in such a way as to exhibit the explanandum phenomenon" (Craver 2008, p. 109; see also Craver this collection). This is an attractive idea as it is expressed, but what I hope to illustrate here is that the leading formalizations of this general idea (Craver 2008; Craver & Bechtel 2007) place overly restrictive conditions on good mechanistic explanation. In what follows, I lay out the norms of mechanistic explanation, as developed by Craver and Bechtel, and describe some cases that their model nicely cap-

tures. I then introduce the case of Starburst Amacrine Cells (SACs)—a type of motion-sensitive cell in mammalian retina. In SACs, and in the functionally coupled direction-selective ganglion cells, the function-structure relationship is hard to capture within the Craver/Bechtel mechanistic framework. I argue that we can better capture such cases by replacing the notion of mechanistic componential constitution with the more general notion of enabling constraints.

2 The requirements of mechanistic explanation

Craver (2008) sharply distinguishes between two traditions of understanding scientific explanation: reductive explanation and systems explanation. According to Craver, the first tradition accepts a version of the covering law model of explanation (Hempel 1965) whereby one explains regularities at a given level of organization by showing how these regularities (the laws describing events and their relations) can be derived from theories holding at lower levels. Put differently, one explains a phenomenon of interest by showing how it is to be *expected* based on the laws governing activity at lower levels of organization. This tradition is reductive because when such explanations are successful, one can strictly speaking *do without* the higher-level laws. However convenient they may be for understanding or predicting higher-level phenomena, the higher-level laws do not add, capture, or explain any facts that are not already contained in the lower-level laws. The lower-level laws are scientifically sufficient.

In contrast, in the systems tradition, a phenomenon of interest ψ exhibited by a system S is explained by identifying a set of component parts $\{X\}$ and showing how they are organized such that S ψ*s*. A systems explanation is similar to reductive explanation in that it too relies on the identification of levels of organization, since it requires identifying the parts of the system *S,* but, as I note below, it does not aim thereby at the reduction or explanatory absorption of one level by another. Craver & Bechtel write:

> In levels of mechanisms, an item X is at a lower level than an item S if and only if X is a component in the mechanism for some activity ψ of S. X is a component in a mechanism if and only if it is one of the entities or activities organized such that S ψ's. For that is what mechanisms are: they are entities and activities organized such that they exhibit a phenomenon. Scientists discover lower levels by decomposing the behavior of a mechanism into the behaviors of its component parts, decomposing the behaviors of the parts into the behaviors of their parts, and so on. (2007, pp. 548–549)[1]

As already noted, S is the system that ψs, or that exhibits phenomenon. It is, for instance, the car (S) that accelerates (ψ), and to explain car acceleration will require identifying the components $\{X\}$ that matter to S ψ-ing. To identify these components and their organization is to explicate the mechanism M that accounts for S ψ-ing. The target of mechanistic explanations of this sort is ψ: "mechanistic explanations are framed by the explanandum phenomenon" (Craver 2008, p. 121) and "[t]he explanandum of a mechanistic explanation is a phenomenon, typically some behavior of a mechanism as a whole" (Craver 2008, p. 139).

In mechanistic explanation, a given X is a component of the mechanism M if and only if it is one of the entities organized such that S exhibits some phenomenon ψ. So the engine, the accelerator, and the gas tank, but not the mudflaps or the windshield wipers are components of M that explain the car accelerating, even though these are *all* parts of the car S. In an

1 There is a terminological issue that needs to be raised at the outset to avoid confusion. Craver & Bechtel (2007; Craver 2008) usually, but not always, use S to refer to a *mechanism*. In contrast, I will always use S to refer to the *system* or entity exhibiting the explanandum phenomenon ψ, and I introduce the symbol M to refer to the responsible mechanism. I do this because M and S are clearly not identical. Moreover, they *are* (or at least appear to me) to be distinguished in this passage, at least on one reading. I think it is unfortunate that neither Craver nor Bechtel formally and consistently distinguish the system S and the mechanism M in their analysis, for reasons that will become clear at the end of this section. Here I'll attempt to faithfully capture the essence of the Craver–Bechtel mechanistic framework, were it to have included this important distinction.

ideal explanation, the mechanism defined by the parts $\{X\}$ will contain *all* and *only* the components relevant to S ψ-ing (see Craver 2008 for a discussion of constitutive relevance in this context). To identify the parts of M is thus to specify *both* a hierarchical and a functional relationship between M and its parts, and between M and S.

But although mechanistic explanation involves essential reference to hierarchical relationships between levels of organization, it is not thereby a species of *reductive* explanation because in a successful systems explanation nothing is rendered inessential or redundant. The phenomenon ψ is neither *derived* nor *derivable* from laws governing the parts of M; rather, the parts $\{X\}$ and their relationships simply *are* M, and together explain why S ψs. The explanatory relationship is not rational derivation, but functional composition: M is physically and functionally *constituted* by its parts, and S ψs in virtue of that constitution.

> Mechanistic explanations are constitutive or componential explanations: they explain the behavior of the mechanism as a whole in terms of the organized activities and interactions of its components. Components are the entities in a mechanism—what are commonly called 'parts'. (Craver 2008, p. 128)[2]

Given all this we can add one more criterion for a given X being a part of the mechanism M: each X must be not just a functional but also a *spatial* sub-part of M. As a component of M, X will be at a *lower level* than M, and *smaller than* M: "[b]ecause mechanisms are collections of components and their activities, no component can be larger than the mechanism as a whole, and so levels of mechanisms are ordered by size" (Craver & Bechtel 2007, pp. 549–550). Craver and Bechtel conclude: "[m]ost fundamentally, levels of mechanisms are a species of compositional, or part-whole relations" (Craver & Bechtel 2007, p. 550). In the overall framework developed by Craver and Bechtel, functional levels and spatial levels generally align.

Thus, although componential mechanistic explanations are not reductive, they generally *are* what one would call "bottom-up", or perhaps better in this context, "level-restricted": one explains the phenomenon ψ in S by reference to entities and relations at a lower level of organization, but never the reverse. In componential explanations of this sort, the intrinsic properties of and interactions between the mechanism's components account for a system's actions (where "intrinsic" means that such properties—such as the charge of an ion—are either basic to the entity or accounted for by reference to entities and properties at a still lower level of organization). Good mechanistic explanations on this view will not include references to unanalyzed properties of the whole S or M, its "shape" or overall organization, as the relations between the components $\{X\}$ at the lower level will already account for (in fact constitute) these.

This account of mechanistic explanation seems to me a clear and, indeed, compelling model of one kind of explanatory practice in the neurosciences. To satisfy the norms of mechanistic explanation, one must:

1. Identify the phenomenon of interest ψ
2. Identify the system S that ψs
3. Identify the relevant spatial sub-parts $\{X\}$ of M (and their relevant intrinsic properties)
4. Describe how the parts $\{X\}$ are organized such that S ψs

At least *prima facie*, a number of instances of successful (albeit incomplete) explanatory models in the neurosciences appear to neatly fit this description. Craver (2008) extensively discusses the mechanistic model of the action potential. Briefly, following the steps above:

1. The phenomenon ψ is the action potential, which consists of the rapid depolarization of neural cells from a resting membrane potential of approximately –70mV toward (and in many cases significantly exceeding) 0mV; an

2 Note that within this framework "componential mechanism", "constitutive mechanism", and "compositional mechanism" are synonymous.

equally rapid repolarization; a period of hyperpolarization, where the cell overshoots the normal resting potential; and a gradual return to the resting equilibrium (note that as even this simplified sketch illustrates, ψ will often be in and of itself complex, with many aspects that any adequate model must capture).
2. The system *S* that ψs is the neuron.
3. The parts in virtue of which *S* ψs include elements of the cell and its surrounding ionic milieu: positively charged K+ and Na+ ions; gated, ion-specific membrane channels; and the Na+/K+ pump.
4. Finally, the organization that explains ψ includes the following: The resting potential is in fact an equilibrium between two opposing forces: a chemical concentration gradient that pushes Na+ into the cell and K+ out of it, and an electrical gradient that pushes K+ into the cell, each maintained by the selective permeability of the cell to Na+ and K+. Na+ channels change their conformation in response to current flow (they are voltage-gated) such that they open to allow Na+ to flow into the cell. As Na+ flows into the cell this reduces the electrostatic pressure on K+, and opens voltage-gated K+ channels, allowing K+ to flow out of the cell. The net effect is to push the cell initially toward the electrochemical balance point for Na+, which is about +55mV. However, as the membrane potential drops, the Na+ channels close, thus slowing and eventually stopping the depolarization. The diffusion of K+ out of the cell combines with the activity of the Na+/K+ pump to repolarize the cell, which however overshoots the resting potential due to the fact that the K+ channels close later than the Na+ channels, thus allowing K+ to diffuse out of the cell for an extra millisecond or so during which the cell is hyper-polarized.

Obviously, this remains a sketch (see Craver 2008) or any basic neuroscience textbook for more detail), but it illustrates the main elements of a mechanistic explanation. The intrinsic properties, actions, and interactions of *M*'s spatial sub-parts together comprise the mechanism that allows *S* to ψ and thus explain how S ψs. One can likewise plausibly sketch the mechanisms that account for spatial long-term memory (e.g., the ability of an animal to return to some location in its environment) in terms of long-term potentiation of synapses in the hippocampus (Craver 2008), although it is worth noting that a more complete account of the functions of hippocampus will have some of the features I describe in 3 and 4 (Buckner 2010; Anderson 2015). Still, the fact that *some* explanations in neuroscience are like this is not under significant dispute.

But this brings us to the question of why I have distinguished *M* and *S* in my treatment. Because Craver (2008) does not formally distinguish these, he is never led to ask what the precise relationship is (or could be) between *M* and *S* (and between their respective parts). In fact, for Craver the symbol *S* usually (but not always) refers to what I have been calling *M*, and he frames his analysis of mechanistic composition entirely in terms of ψ and its mechanism. When he does mention the larger system it is generally to emphasize the fact that not every part of a system *S* is relevant to the mechanism in virtue of which it ψs. So what might the committed mechanist say about the relationships between *S*, *M* and {*X*}? One possibility is: all the parts {*X*} of *M* will be on a lower level than *S*. That would be in keeping with the level-restricted character of the framework, and its characteristic alignment between spatial and functional levels. It is certainly a feature of all the examples discussed in its support, including the model of the action potential outlined above. A slightly stronger possibility would be: all the parts {*X*} of *M* will be spatial sub-parts of *S*. I don't think anyone would or should endorse this stronger condition, but seeing why will be instructive, and will lead us to the reasons to reject the weaker formulation as well.[3]

3 On my reading, the framework developed in (Craver 2008) implicitly assumes the weaker condition, although most likely not the stronger one. But for my purposes here it is not crucial to pin this down. If the framework *does* assume the weaker condition, what follows should be read as arguing (contra this model) that there are systems for which functional

The immediate trouble with the stronger formulation is that it collides with a fact noted by Craver (2008), but not otherwise discussed: the mechanism that accounts for S ψ-ing may contain parts that are extrinsic to S (although not to M). For instance, in the mechanism for the action potential, the Na^+ and K^+ ions that are clearly part of M are (at least sometimes) extrinsic to S; and in embodied accounts of some cognitive processes like mathematics, the mechanism that accounts for a person (P) multiplying (ψ_m-ing) contains parts that are *always* extrinsic to P, such as pencil and paper (Clark 1997; see also this collection). These entities would arguably *not* be components of the systems that ψ, although they would be components of the mechanisms in virtue of which they ψ. At the very least, this suggests there are some details yet to be worked out about the necessary physical relationships between M and S that implement the hierarchical and functional relationships in virtue of which M can account for S ψ-ing. There will be (presumably rare) cases in which M and S are identical; cases such as the accelerating car where M contains only parts of S; and cases such as the action potential where M and S cross-cut one another, sharing some but not all of their parts.[4] There may also turn out to be cases in which they share no parts, perhaps because the parts of M and the parts of S are individuated by different criteria, or because S's ability to ψ is imposed by or inherited from an entirely extrinsic mechanism (indeed I'll discuss a potential instance of this class of cases later in the paper).

But distinguishing M and S in this way *also* allows one to ask whether all the parts of M need to be at a lower level than S. If not every X needs to be a spatial sub-part of S, then there is little reason to suppose that each X needs to be on a lower level than S, either. Indeed, I claim that in fact for some systems S the mechanism M will contain items that are neither intrinsic to *nor at a lower level than S*. For instance, I often use other people to help me remember things, in the easiest case by asking them to remind me at some future time. In such a case, this other individual is arguably part of the mechanism responsible for my remembering, but is certainly not for that reason on a lower ontological level than I am, qua remembering system. Moreover, as I will argue when looking at the case discussed below, some relevant parts of M (and certainly M itself) are at a *higher* organizational level than S. Now of course, Craver & Bechtel *define* the concept of lower level in terms of being a part of the mechanism: "an item X is at a lower level than an item S if and only if X is a component in the mechanism for some activity ψ of S" (2007, p. 548). I agree that this holds for the constitutive relationship between *mechanisms* and their parts. But it only holds for all systems S if we assume that all the parts of M are parts of S, and we have seen that this is not always the case. Thus although I think that Craver correctly analyzes the relationship between mechanisms and their parts in terms of constitution, I argue that the more capacious notion of *enabling constraint* better captures the relationship between mechanisms and the systems whose activities they enable.

In any case, with this as background, I now turn to the case of the SAC. In 3, I describe what we know about how the mechanisms in virtue of which the cell operates, and in 4 I discuss the implications of this case for componential mechanistic explanation.

3 Direction selectivity in SAC dendrites: Beyond componential constitution

Starburst Amacrine Cells are axonless neurons found in the retina of mammals and numerous non-mammalian species. Their morphology is planar, with multiple dendrites arrayed, as the

and spatial levels in fact dissociate. If it does not, then what follows should be read simply as offering an account of some of the possible functional relationships between mechanisms and systems, an issue not explored in the original analysis. Either path leads to the same recommended modification of the original model.

4 In the case of the action potential, one *might* mount the argument that the system that ψs is *strictly speaking* S + {the nominally non-S parts of M}, including the surrounding extracellular fluid. That would make M part of S in this case, but it is not clear to me that this move will be equally attractive in every such case, nor do I think the mechanist is *forced* to adopt this strategy.

name suggests, in a starburst pattern around the cell body (Figure 1).

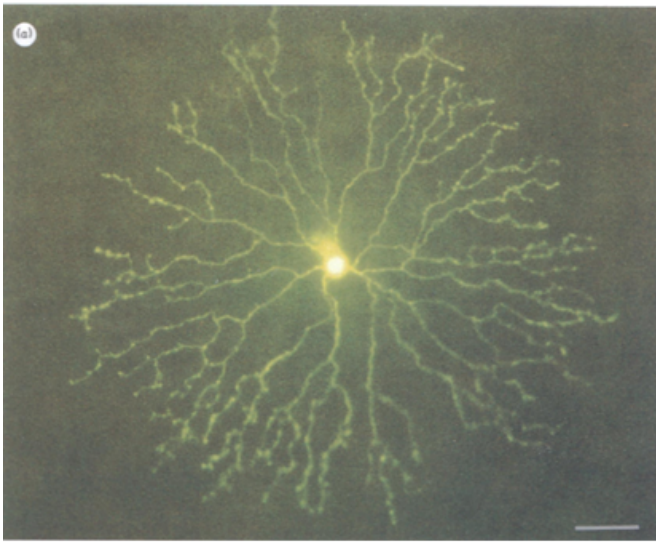

Figure 1: Micrograph of a Starburst Amacrine Cell. Calibration bar 50µm. Reprinted from Tauchi & Masland (1984).

SACs form dense, highly overlapping, co-fasciculating layers in the "on" and "off" levels of the inner synaptic layer of the retina, nestled physically and functionally between bipolar cells and direction-selective ganglion cells. Among the most numerous neural cells found in the mammalian retina, they represent a large proportion of the total neural volume in the eye; in the rabbit retina, for example, as much as six meters of SAC dendrites occupy each square millimeter of retinal surface—higher coverage than any other retinal cell by an order of magnitude (Masland 2005; Tauchi & Masland 1984; see Figure 2).

SACs are interesting for multiple reasons. Despite lacking axons, they synthesize and release both excitatory and inhibitory neurotransmitters (ACh (acetylcholine) and GABA (-Aminobutyric acid)) from the distal regions of their dendrites. Both the role and relative proportion of excitatory and inhibitory synaptic connections change over time. Cholinergic synaptic connections between neighboring SACs disappear over development, and GABAergic connections between SACs begin as excitatory but later become inhibitory. However, excitatory cholinergic synapses between SACs and ganglion cells remain (Masland 2005).

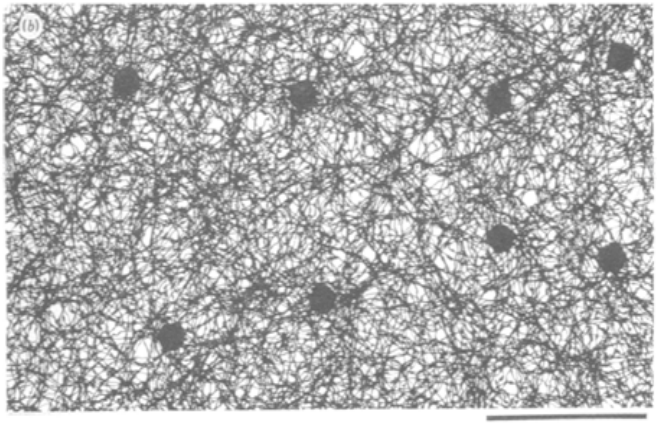

Figure 2: Depiction of the SAC network in peripheral retina. Calibration bar 50µm. Reprinted from Tauchi & Masland (1984).

Functionally, SACs play an important role in motion detection, and are part of the overall network for multiple uses including optokinetic eye movement and motion perception (Yoshida et al. 2001). In fact, each dendrite of the SAC acts independently of the others, and signals the presence of stimuli moving centrifugally, that is, from the cell body out in the direction of the signaling dendrite (Euler et al. 2002; see Figure 3). Put differently, each SAC dendrite is a directionally selective spatial sub-part of the overall cell, and this is the functional property that will interest us here. As with so much in the neurosciences, the mechanism that explains this function is complex and not fully understood. It is, however, possible to offer a sketch of it.

As mentioned above, SACs lie between bipolar cells and direction-selective ganglion cells. Bipolar cells thus mediate the initial stimulus such that a moving light causes them to fire in turn as the stimulus moves across the retina. The bipolar cells make excitatory synapses onto the SAC dendrites.[5] With these basic anatomical facts in view, we can turn to describing

5 In fact there are two classes of bipolar cells, "on" and "off", functionally differentiated by their disposition to respond to stimulus onset vs. stimulus offset—i.e., one responds to light and the other to dark —and anatomically distinguished by whether they synapse onto the "on" or "off" level of the inner synaptic layer (Figure 4). As the mechanisms for direction selectivity in SAC dendrites are the same regardless, I'll ignore this detail in what follows.

three different aspects of the overall mechanism for direction selectivity: wiring specificity between bipolar cells and the SAC dendrites; lateral inhibition between neighboring SACs; and active elements in the dendrites themselves.

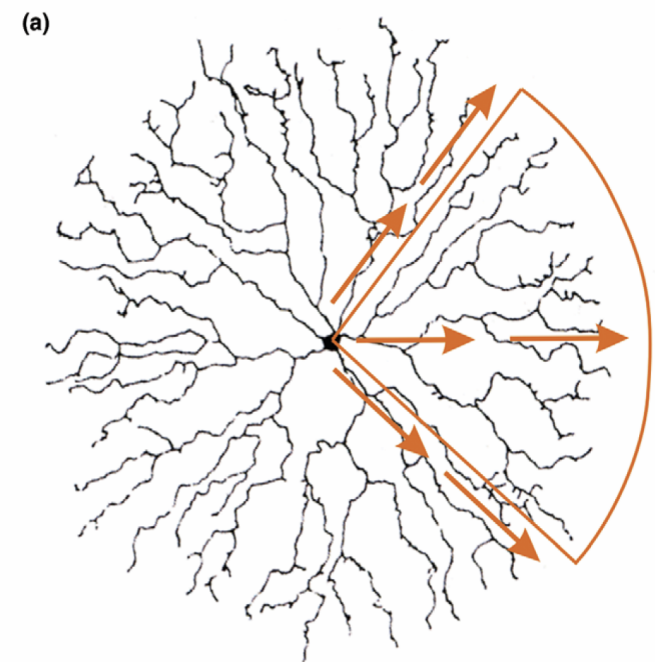

Figure 3: Depiction of direction selectivity in SAC dendrites. Reprinted from Masland (2005).

First, the axonal projections of bipolar cells largely preserve the topography of their inputs, such that neighboring axons come from cells with neighboring inputs, and make neighboring synapses onto post-synaptic cells. What this arrangement means for SACs is that neighboring synapses on the dendrite are likely to come from neighboring bipolar cells, so that when a moving stimulus activates one cell, and then another immediately to its left (say), this will tend to activate a given synapse, and then another immediately to *its* left. Thus, in the case where such a stimulus moves along the direction of a dendritic process, the successive excitatory inputs to that dendrite will tend to reinforce (Demb 2007; Lee & Zhou 2006). This is an important part of the overall mechanism, but is not sufficient by itself to produce the observed directional selectivity, as these inputs would tend to reinforce even during centripetal motion, although this would result in a weaker response at the *distal* process of the dendrite (Hausselt et al. 2007).

Another important part of the mechanism for directional selectivity involves mutual inhibition between neighboring SACs (Figure 5). As a stimulus moves so as to stimulate the centrifugal dendrite of SAC1 (in Figure 5A), reinforcing inputs will cause the release of GABA onto the centripetal dendrite of SAC0, such that even when the light stimuli begins to excite the centripetal dendrite of SAC0, the leading inhibition dominates the signal. Similarly, as the stimulus moves to the *centrifugal* dendrite of SAC0, the successive excitatory inputs from the bipolar cells reinforce, and any inhibitory inputs from the neighboring SAC2 come too late. Moreover, SAC0 will largely inhibit SAC2's response (Figure 5B; Lee & Zhou 2006). An important element of this mechanism involves the relative time-course of ACh and GABA: ACh response from the bipolar cells ramps up and decays fairly quickly, while GABA response is relatively delayed and prolonged (Demb 2007). This temporal asymmetry helps ensure that when inhibition leads it dominates, and vice-versa. The distance between SACs also plays a role. The likelihood of synaptic connections between the distal portion of the dendrites of two SACs—where inhibitory connections are most effective—depends on the distance between the cell bodies. Cells that are very close together or very far apart will thus not mutually inhibit one another (Figure 5C).

Finally, direction selectivity depends upon properties of the dendrite itself. The dendrites are electrically isolated from one another, as a result of both overall cell morphology and the low impedance of the cell body. The uneven distribution of synaptic inputs and outputs also contributes: excitatory inputs from the bipolar cells are distributed along the length of the dendrite, but synaptic outputs are confined to the distal ends (as implied by the two aspects of the overall mechanism described above). A third, active aspect of the local dendritic portion of the mechanism appears to involve voltage-gated calcium channels. These channels lead to amplification of the ACh response beyond what the passive reinforcement caused by successive synaptic transmission from bipolar cells can account for (Hausselt et al. 2007).

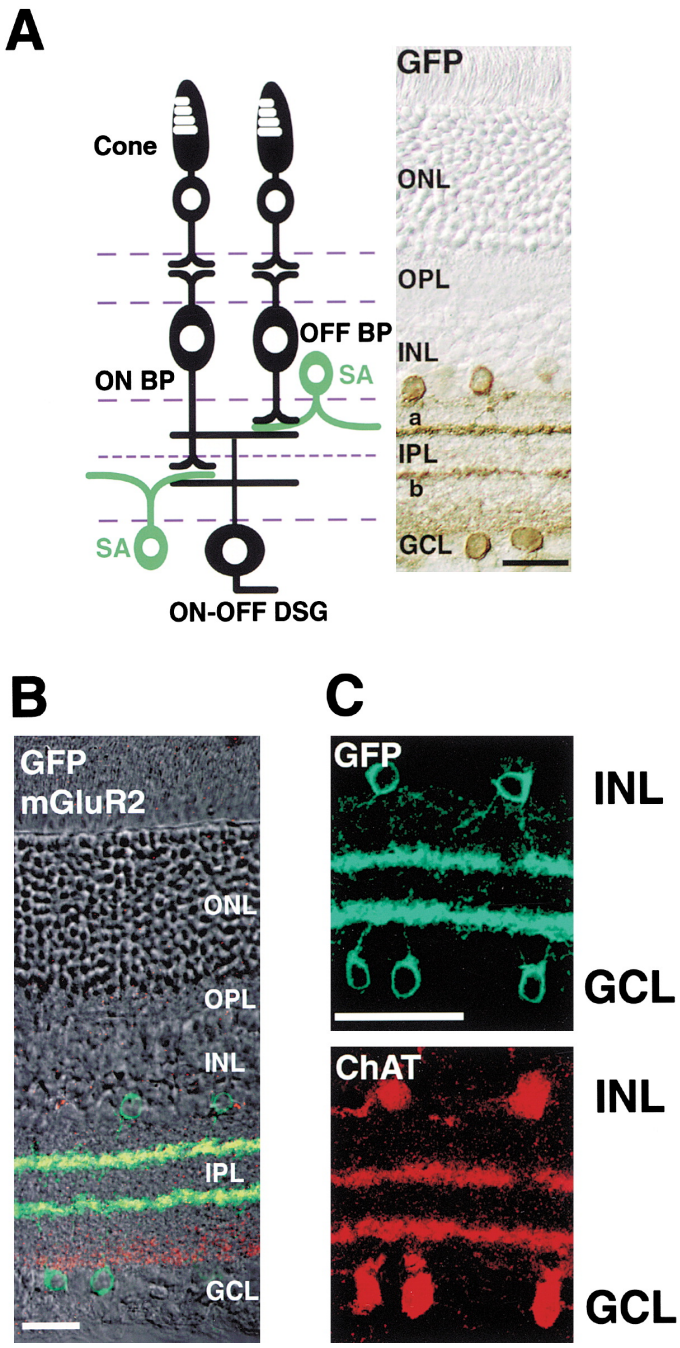

Figure 4: Schematic representation of the layered structure and synaptic relationships between bipolar cells and SACs. Reprinted from Yoshida et al. (2001).

All of these elements combine to produce the direction selectivity of the SAC dendrite. Bipolar cells successively synapse onto the dendritic process, resulting in passive reinforcement of excitatory input that preferentially promotes neurotransmitter release in response to motion in the centrifugal direction. Surrounding SACs selectively inhibit centripetal excitation, as a result of the different temporal activation profiles of GABA and ACh; the asymmetric distribution of input and output synapses; and the relative spatial placement of the SACs. And voltage-gated calcium channels in the dendrite actively amplify the centrifugal signal. Although this sketch leaves out many of the known details, and there remain many details still to be worked out, I believe it is sufficient to warrant the conclusion that this is (a) an instance of mechanistic explanation that (b) does not have the level-restricted character of the (canonical) mechanistic explanations laid out above. I spell out the reasons for this conclusion in the next section.

4 Constitution and constraint

We can most readily see why this case represents an interesting challenge for componential mechanism by fitting it to the four steps outlined in section 2, above.

1. Identify the phenomenon of interest ψ
2. Identify the system S that ψs
3. Identify the relevant spatial sub-parts $\{X\}$ of M (and their relevant intrinsic properties)
4. Describe how the parts $\{X\}$ are organized such that S ψs

The specific phenomenon of interest ψ_{ds} is direction selectivity or, more precisely, the release of neurotransmitter in and only in response to motion in a specific centrifugal direction. The system S_{ds} that exhibits ψ_{ds} is the dendrite of the SAC. It is also easy to say what the parts $\{X_{ds}\}$ of the mechanism M_{ds} are in virtue of which the dendrite ψ_{ds}s, and how they are organized. I have provided that sketch above. Finally, it seems right to say, following Craver (2008), that the relationship between M_{ds} and its parts $\{X_{ds}\}$ is one of componential constitution, such that all the parts $\{X_{ds}\}$ are at a lower level than M_{ds}, and together constitute M_{ds}. But now it gets interesting for componential mechanistic explanation as currently developed. For only some of the parts of M_{ds}—including the voltage gated calcium channels, and the input and output synapses—are at a lower (spatial) level than the

dendrite $S_{ds.}$ The inhibitory dendrites of the neighboring SACs are at the same level as S_{ds}, the bipolar cells and their spatial relations are arguably at a higher level than S_{ds} (although one might wish to screen these off as mere *inputs* to the mechanism), and the mechanism M as a whole in virtue of which S_{ds} $\psi_{\text{-ds}}$s is *certainly* at a higher level than, and is in no way a physical or functional component of S_{ds}.

I think this example demonstrates that not every mechanistic explanation will have the "bottom-up" or "level-restricted" character that the mechanism for the action potential has, where function is built entirely from the capacities of lower-level components and their interactions. In the SAC dendrite, we appear to have a case *not* of a system that ψs in virtue of the capacities and relations of its components (and that could in turn be thought of as a component supporting the activities of a larger functional system), but rather very nearly the reverse: a system that ψs in virtue of the properties of and interactions in the higher-level system of which it is a part. That is, the SAC dendrite is not functionally related to its surrounds as a component to a higher-level system; nor is the higher-level system related to the SAC dendrite as one of *its* components. Instead, I want to say that the higher-level mechanism M acts as an *enabling constraint* on S.

Before providing a bit more in the way of substantial analysis of the concept of an enabling constraint, let us pause to consider one way in which a supporter of componential mechanistic explanation might resist this conclusion by redefining the system S_{ds} to include the mechanism M_{ds}. I think this is not a viable option for a number of reasons. First, it would appear to violate standard usage: neuroscientists speak of direction-selective dendrites, and *not* of a directionally selective network spanning several retinal layers. The debate in the neuroscientific literature concerns *not* the definition of the direction-selective system, but the relative role of intrinsic and extrinsic mechanisms for dendritic direction selectivity in SACs (Hausselt et al. 2007; Lee & Zhou 2006).

Second, it appears that the mechanism as a whole is *not* direction selective. Any given SAC, for instance, and certainly the network as a whole, signals motion in *all* directions. Even if we restrict the definition of M_{ds} to the entities in virtue of which *one* particular SAC dendrite is directionally selective, the symmetry of the mechanism—the fact that SACs *mutually constrain* one another and the same bipolar cells synapse onto more than one SAC dendrite—strongly suggests that *very same mechanism* generates right direction selectivity in the rightward-reaching dendrite in SAC0, and left direction selectivity in the leftward-reaching dendrite in SAC2 (e.g., in Figure 4). The mechanism, that is, does not have the same direction selectivity as either of the dendrites. Rather, it's as if when you turn the crank one way (i.e., the stimulus moves one way) the mechanism produces one output; and when you turn it the other way, it produces the other output.

This suggests a different way to illustrate the limitations of componential mechanism as formulated. Craver writes that the explanandum phenomenon ψ is "typically some behavior of the mechanism as a whole" (Craver 2008, p. 139), and he thus might insist, contra my way of formulating his framework in 2, that it is the mechanism M and not the system S that exhibits ψ. In this case, because I have agreed that the parts $\{X\}$ in fact constitute M, any conflict between functional and spatial levels disappears. But in the case before us it seems that the mechanism *responsible* for, say, rightward direction selectivity does not in fact *exhibit* rightward direction selectivity. So the functional puzzle reasserts itself in a different guise.[6]

One might nevertheless insist on distinguishing these mechanisms in subtle ways—perhaps M_{ds0} includes these synapses from bipolar cells, but not those synapses, while M_{ds2} includes those synapses but not these. I doubt whether this can work, because explaining direction selectivity in *either* direction will require reference to the excitatory inputs from bipolar cells to the centrifugal dendrite, and the inhibitory inputs from the overlapping centripetal dendrite, which are in turn a result of the excitatory inputs from the *very same* bipolar cells synapsing

6 Thanks to an anonymous reviewer for pointing out this way of expressing the matter.

onto the centrifugal dendrite. But let us take the possibility as granted. Then one seems forced to say something along the following lines: the mechanism as a whole ψs, but *signals* ψ-ing with the dendrite.

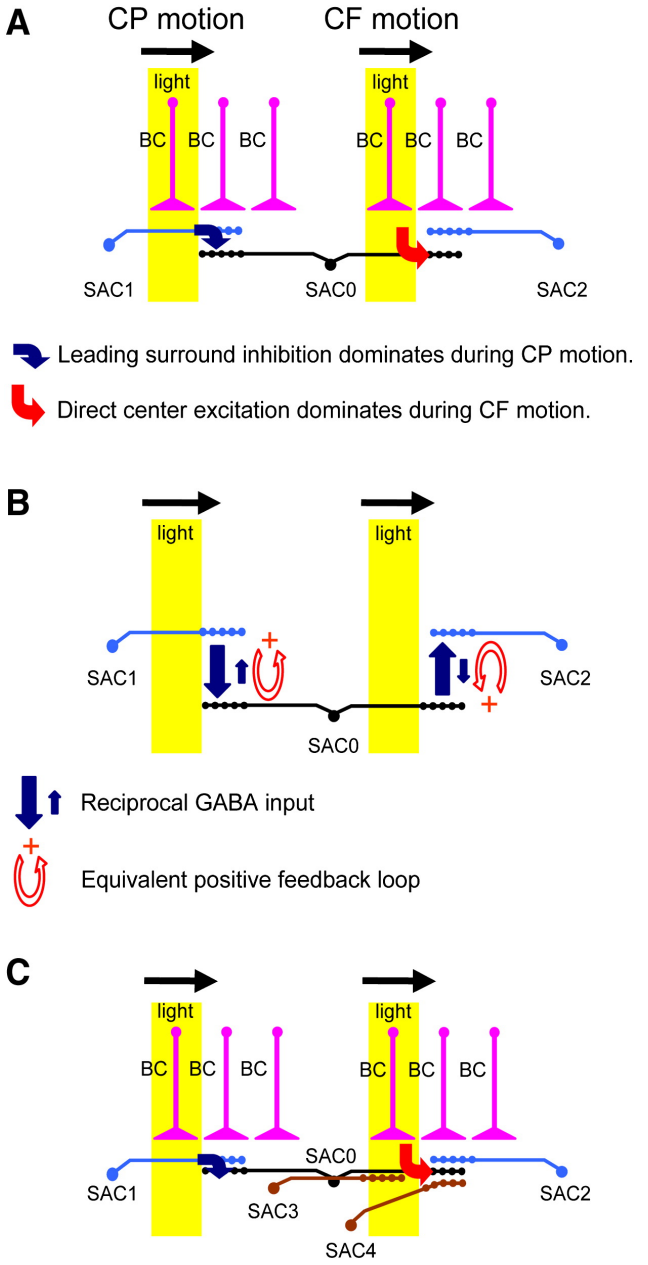

Figure 5: Lateral inhibition between neighbouring SACs contributes to direction selectivity in the dendrites. Reprinted from Lee & Zhou (2006).

Let us consider this possibility carefully. As I intimated above, scientists debate the relative importance of intrinsic and extrinsic mechanisms for dendritic selectivity in SACs. Hausselt et al. (2007) note that direction selectivity in SAC dendrites persists in the presence of GABA and glycene receptor antagonists, which would deactivate the portions of the normal mechanism that involve mutual inhibition between neighboring SACs. In these circumstances, one might argue that *only* the portions of the original mechanism *intrinsic* to the dendrite matter in the explanation of direction-selectivity, and in such a case it is clearly the dendrite that ψs. What shall we say, then, when we remove the antagonists from the system and reapply the same directional stimulus, resulting in neurotransmitter release from this dendrite? One option is: whereas before the dendrite ψ'd, now it merely signals the ψ-ing of the larger mechanism. But it seems clear to me that, if the dendrite can ψ, then adding network interactions that *aid and enhance* (that is, do not in any sense prevent) ψ-ing can hardly cause it to *not* ψ, but only signal ψ. This points to a fourth and final reason to reject the general move to extend the neural system S to include the mechanism M whenever it is (or contains entities that are) on a higher level than S: one would apparently need the ability to rigorously distinguish between ψ-ing and *signaling* ψ in an overall system where to ψ is generally also to signal it—that is, where signaling and doing are deeply intertwined. Thus, I believe we must insist: the dendrite ψs.

For all these reasons, I do not think it is wise to hold onto level-restricted explanations and componential composition by fiat. Instead, it is time to expand the scope of mechanistic explanation by considering the various ways in which systems S relate to the mechanisms M that enable their activities. I think the case of SACs is especially important because it illustrates one way in which local selectivity in parts of a network can be the result of the interplay of excitation and mutual inhibition between non-selective parts of that network, which is clearly something that we need to understand better if we are to accurately characterize the functional mechanisms at work in both small and large-scale brain networks (Anderson et al. 2013). But other structure-function relationships appear to call equally for a broader account of mechanistic explanation. For instance,

the direction-selective ganglion cell DSGC (Direction-Selective Ganglion Cell), mentioned briefly above, responds to stimuli moving only in its preferred direction (which of course varies cell-to-cell). In this case, there do not appear to be *any* intrinsic mechanisms for the direction selectivity of the DSGC. Rather, SAC dendrites selectively synapse onto DSCGs with preferred stimuli antiparallel to the SAC dendrite preference (Briggman 2011) thus suppressing responses to motion in the non-preferred direction. DSCGs seem to simply *inherit* their selectivity via their synaptic contact with SACs—and, in fact, elimination of SACs from the retina abolishes direction selectivity in DSCGs (Yoshida et al. 2001). Here I just don't see any case for a compositional relationship between the mechanism (or its parts) and the selective system. Instead, the relevant mechanism synapses onto the relevant system, and by suppressing a sub-set of its response tendencies, induces selectivity.

This brings us finally back to the notion of "constraint", which I think may help us understand the full range of mechanism/system relationships in the brain. The term constraint has been used in myriad ways in the literature on scientific explanation. In evolutionary biology, scientists refer for instance to stability constraints (Schlosser 2007) and both universal and local developmental constraints on evolvability (Maynard Smith et al. 1985). There are also law-like constraints on the possible states of physical systems generally (Lange 2011). None of these capture the sense of "constraint" that will be most helpful to us here.

One notion that gets us close is the idea of a "capacity constraint", that is, a limitation on the capacity of a process that might take the form of changing the relative probabilities of the range of possible process outcomes (Sansom 2009). This certainly has the right flavor, for in the mechanism under discussion above it appears that the excitatory and inhibitory interactions between bipolar cells and neighboring SACs bias the outcome of the dendritic processing of the moving stimulus. But insofar as a capacity constraint is generally conceptualized in terms of the reduction of some pre-existing whole ability—in Sansom's (2009) example, being handcuffed limits one's ability to move one's hands—this does not offer quite the right organizing frame for explanation in neuroscience.

The reason is that in the neurosciences we want to understand not just the capacities of entities, but how the structured interactions between entities give rise to *functions*, which are, crucially, *differential* and *differentiating* processes (that is, they differ from one another, and they differentiate between stimuli). Capacities in the sense of general powers (the capacity to generate an action potential, say) are necessary conditions for functions, but they are not yet functions; the DSGC is strictly speaking *non-functional* in the absence of SACs, even though it will continue to exercise its capacity to fire action potentials in response to inputs from bipolar cells. Constraints of the sort under investigation here serve to limit capacities, but in so doing they enable functions; they result in an *enhancement* (not a reduction) of the abilities of the system (and the organism).

For this reason I propose to analyze the general functional (and, crucially, *non-hierarchical*) relationship between mechanisms and systems in the following way: an *enabling constraint* is a relationship between entities and/or mechanisms at a particular level of description and a functional system at the same or a different level, such that the entities/mechanisms bias (i.e., change the relative probabilities of) the outcomes of processing by the system. Such enabling constraints offer necessary but not sufficient conditions for the instantiation of differential function in neural systems. Because enabling constraints are synchronic rather than diachronic, the idea shares the same explanatory advantage that the relation of constitution has over the relation of "causation" (when understood, e.g., as an event involving the transmission of some property, power, or conserved quantity from one entity to another). As Craver & Bechtel (2007) point out, such a conception of causation does not accomodate interlevel functional relationships well, because these are often synchronic and symmetric, whereas causa-

tion of this sort is temporal and asymmetric.[7] In addition, enabling constraints can be *mutual*, which gives the idea an advantage over both causation and constitution as an analysis of functional relationships in the brain.

> Enabling constraint $=_{\text{Df}}$ A physical relationship between a functional system S and entities {X} (and/or mechanism M), at the same or different level of description, such that {X} (and/or M) changes the relative probabilities of various possible functional outcomes of activity in S.

To understand function not just in systems like SAC dendrites and DSGCs, but also in the large scale networks that are partially constituted by the Transiently Assembled Local Neural Subsystems TALoNS (Transiently Assembled Local Neural Subsystems) crucial to the functioning of a dynamic brain (Anderson 2015), we need to accept that there is a broader range of relationships that mechanisms can have to functional systems, beyond componential constitution. Function in TALoNS results not from structured interactions between stable, autonomous low-level components, but rather from the interplay between the capacities of lower-level entities and higher-level network dynamics. That interplay, I argue, is best analyzed in terms of the mutual constraint that exists between bottom-up and top-down, feed-forward and feed-back mechanisms in the brain.

5 Conclusion

Although mechanistic explanation as developed by Craver & Bechtel (2007; Craver 2008) does seem to accurately characterize one kind of explanation in neuroscience, and one kind of functional arrangement in neural systems, I've argued here that the formulation is not wide enough to capture the variety of mechanisms in the brain. When we formally distinguish the system S from the mechanism M in virtue of which S exhibits the explanandum phenomenon Ψ, we see that although it seems correct to describe the relationship between M and its parts $\{X\}$ in terms of constitution, it will only sometimes be the case that S is (partially) constituted by $\{X\}$.

As an alternative to the relationship of componential constitution, I have offered the notion of an *enabling constraint* that can exist between a system and the mechanism(s) in virtue of which it has its various functions. SAC dendrites appear to have their function in virtue of the enabling constraints imposed by entities at the same and higher levels of organization; and DSGC function is enabled by the constraints imposed by the SAC dendrites. In neither case is it appropriate to describe the relationship between the mechanism M and the relevant system S in terms of constitution, nor are all (or, in the case of DSGCs arguably any) of the parts $\{X\}$ of M components of S.

Overall, I hope to have made the case that moving beyond level-restricted mechanistic explanation will allow us to better capture the variety of neural systems that emerge from the constant, constraining, biasing interplay between feed-forward, feedback, bottom-up, and top-down processes in the dynamic brain.

7 For instance, what explains why a neuron has a particular functional property cannot be an event involving the transmission of some property, power or conserved quantity from the parts of the neuron to the whole, because if causes must precede their effects, this would appear require that there be a time prior to which the neuron did not have the functional property conferred by its parts. Interlevel functional relationships do not generally appear to be temporal in this way. Rather, for Craver and Bechtel, what explains the functional property of the neuron is the way it is *constituted* by its parts. Enabling constraints are also synchronic in the relevant way, and so the view I am advocating here is also able to accommodate such cases of interlevel functional relationships.

References

Anderson, M. L. (2015). *After phrenology: Neural reuse and the interactive brain.* Cambridge, MA: MIT Press.

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

Briggman, K. L., Helmstaedter, M. & Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, *471* (7337), 183-188. 10.1038/nature09818

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology*, *61*, 27-48. 10.1146/annurev.psych.60.110707.163508

Clark, A. (1997). *Being there: Putting brain, body and world together again.* Cambridge, MA: MIT Press.

——— (Ed.) (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Craver, C. F. (2008). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience.* Oxford, UK: Oxford University Press.

——— (Ed.) (2015). Levels. In T. Metzinger & J. M. Windt (Eds.) *Open MIND.* Frankfurt a. M., GER: MIND Group.

Craver, C. F. & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy*, *22* (4), 547-563. 10.1007/s10539-006-9028-8

Demb, J. B. (2007). Cellular mechanisms for direction selectivity in the retina. *Neuron*, *55* (2), 179-186. 10.1016/j.neuron.2007.07.001

Euler, T., Detwiler, P. B. & Denk, W. (2002). Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, *418* (6900), 845-852. 10.1038/nature00931

Hausselt, S. E., Euler, T., Detwiler, P. B. & Denk, W. (2007). A dendrite-autonomous mechanism for direction selectivity in retinal starburst amacrine cells. *PLoS Biology*, *5* (7), e185. 10.1371/journal.pbio.0050185

Hempel, C. G. (1965). *Aspects of scientific explanation.* New York, NY: Free Press.

Lange, M. (2011). Conservation laws in scientific explanations: Constraints or coincidences? *Philosophy of Science*, *78* (3), 333-352. 10.1086/660299

Lee, S. & Zhou, Z. J. (2006). The synaptic mechanism of direction selectivity in distal processes of starburst amacrine cells. *Neuron*, *51* (6), 787-799. 10.1016/j.neuron.2006.08.007

Masland, R. H. (2005). The many roles of starburst amacrine cells. *Trends in Neurosciences*, *28* (8), 395-396. 10.1016/j.tins.2005.06.002

Maynard Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D. & Wolpert, L. (1985). Developmental constraints and evolution. *Quarterly Review of Biology*, *60* (3), 265-287.

Sansom, R. (2009). The nature of developmental constraints and the difference maker argument for externalism. *Biology & Philosophy*, *24* (4), 441-59. 10.1007/s10539-008-9121-2

Schlosser, G. (2007). Functional and developmental constraints on life cycle evolution: An attempt on the architecture of constraints. In R. Sansom & R. Brandon (Eds.) *Integrating evolution and development: from theory to practice* (pp. 113-173). Cambridge, MA: MIT Press.

Tauchi, M. & Masland, R. H. (1984). The shape and arrangement of the cholinergic neurons in the rabbit retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *223* (1230), 101-119. 10.1098/rspb.1984.0085

Yoshida, K., Watanabe, D., Ishikane, H., Tachibana, M., Pastan, I. & Nakanishi, S. (2001). A key role of starburst amacrine cells in originating retinal directional selectivity and optokinetic eye movement. *Neuron*, *30* (3), 771-780. 10.1016/S0896-6273(01)00316-6