Behavioural Processes 117 (2015) 4–11

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Mechanism, function, and computation in neural systems

Nathan Insel[a,b,*], Paul W. Frankland[b,c,d,e,*]

[a] Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada M1C1A4
[b] Department of Psychology, University of Toronto, Toronto, ON, Canada M5S3GM
[c] Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, ON, Canada M6G1X8
[d] Institute of Medical Science, University of Toronto, Toronto, ON, Canada M5S1A8
[e] Department of Physiology, University of Toronto, Toronto, ON, Canada M5S1A8

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form 21 August 2014
Accepted 10 September 2014
Available online 22 September 2014

Keywords:
Mechanism
Circuits
Levels of analysis
Hippocampus
Prefrontal cortex

ABSTRACT

What constitutes a "mechanism" of behavior? In this tribute to Jerry Hogan we examine how questions of behavioral mechanism can be reframed as causes and consequences of neural circuit activity. Drawing from our work on the hippocampus and the medial prefrontal cortex we discuss the inherent difficulties of characterizing the behavioral functions of circuits that are many synapses away from sensory reception and motor/visceral expression. We briefly review the advantages of reframing a region's functions according to its *computations*, while also distinguishing those computations from the algorithms by which they are achieved. As an example of how these ideas can be applied, we discuss why the hippocampus and medial prefrontal cortex may have overlapping roles in memory expression in spite of being very different circuits. The present analysis draws inspiration from David Marr, whose framework for describing neural systems can be compared with Aristotle's "causes." This article is part of a Special Issue entitled: In Honor of Jerry Hogan.

Contents

1. Introduction

A fundamental question in science is "Why?" But the question is ambiguous, as explanations can take any number of reference frames. Aristotle (350 BCE) had a scheme for classifying explanations that has endured history with a vitality customary amongst ancient Greeks. He contended that any given event can be explained in terms of (1) the physical substance of the system (material cause), (2) the way that the system or event is organized (formal cause), (3) the external forces that put the event in motion (efficient case), and (4) the purpose, goal, or function of the event (final cause). These divisions are not perfect. They are not entirely unambiguous and may fail to capture some of the most critical features of complex systems (meanwhile raising further metaphysical questions about the nature of contingency). But the framework provides a basic road map that, in spite of its simplicity, is often neglected.

It is easy to see how different types of explanations might become confused in the contemporary scientific setting, and even more so within the specific field of animal psychology. The field has expanded exponentially with the incorporation of neuroscience and genetics techniques, but merging with biology has also meant an intermingling of very different scientific goals. At the same

* Corresponding authors at: The Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8. Tel.: +1 416 813 7654x1823; fax: +1 416 813 7717.
E-mail addresses: nathan.insel@utoronto.ca (N. Insel), paul.frankland@sickkids.ca (P.W. Frankland).

http://dx.doi.org/10.1016/j.beproc.2014.09.015
0376-6357/

time, increased competition for "high impact" real estate offers incentives to interpret findings in as far-reaching ways as possible. Fortunately, the scientific community is a self-correcting system, one that depends on unique individuals like Jerry Hogan who, like Aristotle, has endured with a classical vitality. Jerry reintroduced Aristotle's "four causes" into ethology as a way of offering a common language for different types of inquiries (Hogan, 1994). Jerry has argued, for example, that learning about the behavioral role of specific brain regions (part of the material cause, or "matter") does not necessarily explain what stimuli motivate a behavior (the efficient causes; or simply "causation"). Similarly, he has emphasized the fundamental distinction between knowing the variables that motivate a behavior and understanding the implications of the behavior's consequences. Jerry's ideas are a reminder of the limits over which scientific findings can be interpreted, but they also provide a system for classifying and thereby promoting the diversity of scientific inquiry.

Our own research has focused less on animal behavior *per se* and more on the neural processes that support it. One way to put this is to say that we study "material causes" (or, arguably, "formal causes") of behavior, but at a certain point it becomes useful to depart from Aristotle's language. We can say instead that we study behavioral *mechanisms*; however, as we discuss in more detail below, this term is complicated by ambiguity. By definition, a mechanism is a description of a phenomenon in terms of the actions and interactions of a system's parts—in our case, the phenomenon of interest is the behavioral expression of memory. But establishing a mechanism requires addressing issues that do not have obvious or unanimously agreed-upon solutions, such as the right balance between detail, accuracy, and generalizability, the best ways to subdivide a system into parts, and defining which components are necessary in a complex system with inherent redundancies. It is probably safer to say that we study neural circuits, specifically a few rather deeply-buried circuits in the mammalian forebrain: the hippocampus and medial prefrontal cortex (mPFC). And what we want to know is: what causal interactions make these circuits so important for memory expression?

The purpose of the present article is to step back from what we are doing and think hard about what we do. What, specifically, are the challenges of explaining behavior according to its mechanisms? What angle of approach in studying brain regions like the hippocampus and mPFC will be most informative for a generalizable understanding of their respective roles in behavior? Our examination of these questions outlines the importance of articulating and testing theories about the computations a brain system performs. A theory of computation does not have to be a computer-based simulation or written in the form of differential equations, although we discuss some of the ways that mathematical and computer-based models can be valuable. Theories should also be consistent with, but not fully dependent upon, the region's proposed cognitive functions and the underlying algorithms by which the region is thought to work. After discussing these issues in the abstract sense, we see how they have played-out in studies of the hippocampus and mPFC. Although definitions will be provided throughout the text, we also offer a glossary (Table 1) that clarifies how we use some of the most relevant terms.

2. Challenges in identifying mechanisms of behavior

Ecosystems can be described as interacting communities, communities as interacting individuals, individuals as organs, and so on down to the apparently indivisible, "fundamental" particles and forces. These systems vary in spatio-temporal scale, but they also can be thought about as a hierarchy of embedded systems (or embedded *mereological* systems). The term "levels of analysis" is most often used, which is appealing in that it implies that the purpose of parsing systems into levels is to make them more accessible to our understanding or analysis.

Neuroscience is largely concerned with crossing levels of analysis, often by attempting to explain behavioral phenomena in the form of physiological processes. Claims about the mechanisms of behavior are common and elicit far more excitement – and are correspondingly more publishable – than more conservative "descriptions" of physiological or behavioral processes. When controversies arise, they are often based on disagreements over what empirical and computational evidence is sufficient to demonstrate a mechanism. Why is it so difficult to decide whether one set of processes is a mechanism of another? On the surface, it seems no harder than determining whether the process meets the definition. Craver (2007) has defined mechanism as a "set of entities and activities organized such that they exhibit the phenomenon to be explained". He has also noted that it is important to include all and only those components and actions that are relevant (Craver, 2008). This all sounds very sensible, in principle. In practice, mechanisms become very difficult to identify due to difficulties or disagreements in answering three main questions:

(1) How do we define the phenomenon we are attempting to explain? Specifically, what level of detail is necessary to describe a system's behavior, and what conditions must it generalize across before we can properly attempt to explain it? As an example, if we are interested in understanding the processes that underlie retrieval of memories for experiences, to what extent should the proposed mechanism be specific for, say, visual features of the experience, as compared with spatial? Another example commonly encountered when discussing the memory roles of the mPFC: to what extent are we concerned with memory for the experience as a whole, including its incidental details, versus only those associations that help an animal predict rewards and punishers?
(2) How much precision and accuracy is necessary for describing the parts of the system and their actions? Any useful understanding will necessarily require filtering out details (by including only the components and actions that are relevant), but the loss of details can also influence the accuracy of the described mechanism. Those who use simulations to study mechanism encounter this question all of the time. It is neither practical nor useful to model a network of 10,000 neurons using highly-detailed, multi-compartment simulations of each neuron, but by simplifying the cells as single-compartment, integrate-and-fire neurons a certain degree of accuracy is lost. In the words of Frances Skinner "We should neither ignore the details nor be consumed by them" (Skinner and Mulloney, 1998). A related question is what "level" of analysis is the right one to build a mechanistic account? In other words: what is the most optimal way to subdivide a complex system to understand and predict its behavior?
(3) How can descriptions of mechanism handle built-in redundancies in the functional components of the system? In Craver's formulation of mechanism, all and only those components that contribute to a phenomenon should be included; but when the components of a system have overlapping functions, the removal of any one may have no impact on the system's behavior. A specific example arose in a recent set of studies on molecules involved in memory (see Frankland and Josselyn, 2013). The question was whether the protein PKM-zeta was responsible for – meaning, part of the mechanism that explained – long-term maintenance of altered synaptic connections. While some manipulations suggested that PKM-zeta was involved, others which specifically isolated the protein suggested that it was not a necessary component. It may be the case