Studies in History and Philosophy of Science 41 (2010) 321–333

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science

William Bechtel [a], Adele Abrahamsen [b]

[a] Department of Philosophy, the Center for Chronobiology, and Interdisciplinary Programs in Science Studies and Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
[b] Center for Research in Language, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Keywords:
Circadian rhythms
Computational models
Cognitive science
Mechanistic models
Oscillators
Dynamic models
Explanation

ABSTRACT

We consider computational modeling in two fields: chronobiology and cognitive science. In circadian rhythm models, variables generally correspond to properties of parts and operations of the responsible mechanism. A computational model of this complex mechanism is grounded in empirical discoveries and contributes a more refined understanding of the dynamics of its behavior. In cognitive science, on the other hand, computational modelers typically advance *de novo* proposals for mechanisms to account for behavior. They offer indirect evidence that a proposed mechanism is adequate to produce particular behavioral data, but typically there is no direct empirical evidence for the hypothesized parts and operations. Models in these two fields differ in the extent of their empirical grounding, but they share the goal of achieving *dynamic mechanistic explanation*. That is, they augment a proposed mechanistic explanation with a computational model that enables exploration of the mechanism's dynamics. Using exemplars from circadian rhythm research, we extract six specific contributions provided by computational models. We then examine cognitive science models to determine how well they make the same types of contributions. We suggest that the modeling approach used in circadian research may prove useful in cognitive science as researchers develop procedures for experimentally decomposing cognitive mechanisms into parts and operations and begin to understand their nonlinear interactions.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

1. Introduction

Two widely accepted assumptions within cognitive science are that (1) the goal is to understand the mechanisms responsible for cognitive performances and (2) computational modeling is a major tool for understanding these mechanisms. The particular approaches to computational modeling adopted in cognitive science, moreover, have significantly affected the way in which cognitive mechanisms are understood. Unable to employ some of the more common methods for conducting research on mechanisms, cognitive scientists' guiding ideas about mechanism have developed in conjunction with their styles of modeling. In particular, mental operations often are conceptualized as comparable to the processes employed in classical symbolic AI or neural network models. These models, in turn, have been interpreted by some as themselves intelligent systems since they employ the same type of operations as does the mind. For this paper, what is significant about these approaches to modeling is that they are constructed specifically to account for behavior and are evaluated by how well they do so—not by independent evidence that they describe actual operations in mental mechanisms.

Cognitive modeling has both been fruitful and subject to certain limitations. A good way of exploring this is to contrast it with a different approach, one that involves more direct investigation into the internal parts and operations of the mechanism responsible for a phenomenon and tailors modeling to this mechanism. To do this we will focus on the phenomenon of circadian rhythms in animals: the ability of the nervous system to regulate activities, including human cognitive activities, on an approximately twenty-four hour cycle. Circadian effects on cognition generally

E-mail addresses: bill@mechanism.ucsd.edu (W. Bechtel), adele@crl.ucsd.edu (A. Abrahamsen)

0039-3681/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2010.07.003

have been ignored in cognitive science, but whether or not that is a desirable state of affairs is not relevant here. Rather, our goal is to use the increasingly prominent role of computational modeling in circadian rhythm research as a different type of exemplar against which to view cognitive modeling. In circadian research, the models are not proposals regarding the basic architecture of circadian mechanisms; rather, they are used to better understand the functioning of a mechanism whose parts, operations, and organization already have been independently determined. In particular, circadian modelers probe how the mechanism's organized parts and operations are orchestrated in real time to produce dynamic phenomena—what we have called *dynamic mechanistic explanation* (Bechtel & Abrahamsen, in press).

We begin with an overview of mechanistic explanation in general. We then develop the case of circadian rhythm research, where the architecture has been highly constrained by empirical inquiry into the physical mechanism and modeling is directed to understand the mechanism's dynamics. We do this by examining in turn six different exemplars from the research literature on computational modeling of circadian rhythms. In all of these cases computational modeling was needed to understand the behavior of a complex mechanism involving nonlinearly interacting components. In examining their particulars, though, we draw out six more specific contributions of computational modeling. We then go through these six contributions again, this time presenting for each a cognitive model and querying to what extent it might make the same kind of contribution. This review of models also brings to light certain differences between cognitive scientists and circadian modelers in how they approach computational modeling.

2. Mechanisms and mechanistic explanation

Many philosophical presentations of cognitive science (and other sciences) continue to focus on laws as the explanatory vehicle. Laws are commonly construed as universal generalizations that have a modal status—they identify not just what has happened when particular conditions are met, but what must happen under those conditions. But cognitive scientists, and indeed life scientists generally, seldom propose laws. When they do (in psychology, typically referring to them as *effects*), they generally serve not to explain but to characterize the phenomenon to be explained (Cummins, 2000). When they advance explanations, life scientists commonly seek to uncover the mechanism responsible for the phenomenon of interest. Recently, a number of philosophers whose focus has been largely on biology have attempted to characterize what scientists mean by a mechanism and how they go about developing and evaluating mechanistic explanations (Bechtel & Abrahamsen, 2005; Bechtel & Richardson, 1993; Glennan, 1996, 2002; Machamer et al., 2000; Thagard, 2006). Our own 2005 characterization began:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. (Bechtel & Abrahamsen, 2005, p. 423)

Discovering the parts and operations of a mechanism requires decomposing it. This typically necessitates experimental techniques since in naturally occurring mechanisms, especially living systems, the parts and operations are so highly integrated that they cannot be identified directly. It is relatively easy to find ways to fracture a system into parts of some sort—the challenge is to identify the *working parts* that perform the operations producing the phenomenon of interest. In the case of the brain, a variety of approaches have been pursued. In the nineteenth century, the focus was on the sulci and gyri created by the folding of the cerebral cortex, and while these still are used as anatomical landmarks, they are not regarded as working parts. Once it was recognized that cortex comprised individual cells—neurons—neuroanatomists such as Brodmann (1994 [1909]) used the presence of neurons of specific types and especially differences in the thickness of the layers into which they were organized to differentiate regions in the cerebral cortex. His clear hope was that these areas had functional significance, but he lacked tools for determining this. Refined in later decades using such criteria as neural connectivity and topographical mapping, and studied functionally using such techniques as single-cell recording, it turned out that Brodmann's areas demarcated working parts of the brain so well that they still are in use (Mundale, 1998).

Identifying operations usually involves a very different set of experimental procedures than identifying parts. The goal is to identify operations that do not produce the phenomenon individually but only in collaboration with other operations performed by different parts of the mechanism (otherwise there is no explanatory gain from decomposing the mechanism). Detecting the effects on overall behavior from experimental manipulations of particular parts (e.g. ablating or stimulating them) often provides suggestive clues, as does recording specific internal effects of altering the inputs to the mechanism. Whatever technique is chosen, proposing operations on the basis of the outcome typically requires elaborate inferential schemes (Bechtel, 2008b) that can lead to blind alleys, overemphasis on particular operations to the exclusion of others, and additional sources of dispute. The challenges in identifying both parts and operations make mechanistic explanation a long and complex endeavor, but in numerous domains of biology well supported, enduring accounts have eventually been achieved, providing a foundation for more advanced research.

Discussions of mechanistic explanation often allude to the importance of how the components are organized, but this has been the least developed aspect both of philosophical accounts of mechanistic explanation and of mechanistic science itself. Much more attention has been paid to ways of *decomposing* a mechanism into component parts and operations than to ways of *recomposing* them into an appropriately organized system. Generally scientists use the simplest organizational scheme that will serve their immediate purpose. For example, since the 1930s and still today, the main backbone of reactions in glycolysis has been represented as a linear sequence (plus side reactions): Glucose → G6P → F6P and so forth—not unlike a diagram of a simple assembly line. Yet, as biological theorists from Claude Bernard to the present have recognized, there are distinctive modes of organization in organisms that enable them to exhibit such phenomena as maintaining themselves in a non-equilibrium relation to their environment. Recognition, first of negative feedback and later of positive feedback and self-organizing cycles, has offered biologists a more precise understanding of the key role of organization in living systems (Bechtel, 2006, 2007).

Such modes of organization orchestrate the parts and operations in real time. Thus, our 2005 characterization of mechanism continued as follows:

> The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel & Abrahamsen, 2005, p. 423)

Though this orchestration often is downplayed as investigators focus on identifying parts and operations, attending to it can reveal complex dynamics, ranging from periodic oscillations to chaos. Different tools than those employed in early investigations of a mechanism are required to pursue its dynamics: the tools of quantitative computational modeling. These tools have a long history and have been employed in a variety of ways, often disconnected from any sort of mechanistic project. For example, the