Studies in History and Philosophy of Science 41 (2010) 313–320

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

How we ought to describe computation in the brain

Chris Eliasmith

Department of Philosophy, Department of Systems Design Engineering, University of Waterloo, Ontario, N2L 3G1, Canada

ARTICLE INFO

Keywords:
Neural computation
Control theory
Neural representation
Cognitive architecture
Cognitive models
Theoretical neuroscience

ABSTRACT

I argue that of the four kinds of quantitative description relevant for understanding brain function, a control theoretic approach is most appealing. This argument proceeds by comparing computational, dynamical, statistical and control theoretic approaches, and identifying criteria for a good description of brain function. These criteria include providing useful decompositions, simple state mappings, and the ability to account for variability. The criteria are justified by their importance in providing unified accounts of multi-level mechanisms that support intervention. Evaluation of the four kinds of description with respect to these criteria supports the claim that control theoretic characterizations of brain function are the kind of quantitative description we ought to provide.

1. Introduction

This essay is structured such that each heading is a specific claim related to quantitative descriptions of brain function. Any subheadings under a given heading are intended to provide additional considerations or details in support of the heading. While this does not provide for typical, smooth, reading of the paper, it serves to make the argument clearer and can shorten reading time, as the content of any ''obviously true'' heading can be skipped.

The word 'computation' is used in a liberal and definitional sense. I am using the liberal sense in the title (the sense typical of cognitive science usage, which means something like a 'transformation of representations'). However, I am using the definitional sense, from computational theory (i.e. Turing Machine equivalence) in the remainder of the essay. I will generally replace 'computation in the brain' in the first sense with 'a quantitative description of brain function' for clarity.

In brief, the argument I present here is:

1. There are four relevant kinds of quantitative description of brain function: computational, dynamical, statistical, and control theoretic
2. We ought to provide the best quantitative description of brain function
3. A good description of brain function provides for simple state mappings, and useful decompositions that account for variability
4. A good description in the brain sciences explains by positing mechanisms that support interventions
5. Computation theoretic descriptions do not meet these criteria well
6. Conclusion 1: therefore, computation theoretic descriptions are not good descriptions (from 3–5)
7. Control theoretic descriptions meet these criteria better than any of the other alternatives
8. Therefore, control theoretic descriptions are the best descriptions (from 1, 7)
9. Conclusion 2: therefore, control theoretic descriptions are the kind of quantitative description we ought to provide (from 2, 8)

One clarification is important: conclusion 2 does not rule out the other descriptions as useful. Rather, it suggests that other descriptions are essentially heuristics for temporarily stating the description. That is, ultimately, other descriptions should be

E-mail address: celiasmith@uwaterloo.ca

0039-3681/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2010.07.001

translated into a unifying description of brain function stated with control theoretic constructs.

2. There are four kinds of quantitative description of brain function

I begin with some considerations regarding how quantitative descriptions relate to physical systems in general, and then turn to which quantitative descriptions are relevant for understanding brain function.

2.1. Different quantitative descriptions are better for different classes of phenomena

I do not worry about how quantitative descriptions are individuated (i.e. why statistical descriptions are different from dynamical descriptions).

2.1.1. Physical systems can have multiple quantitative descriptions

In most cases, what we identify as a physical system (e.g. a gas, a computer chip) can be described using different quantitative descriptions (e.g. statistical or Newtonian mechanics, computational theory or circuit theory). If we are trying to argue for *the best* description of some physical system, we must have a means of picking between these possible descriptions.

2.1.2. Quantitative descriptions have a natural class of physical phenomena that they describe

Notably, many descriptions are of the same mathematical class (e.g. both computational and circuit descriptions are algebraic), so it is not their mathematical properties that distinguish them. Instead, it is the mapping between the mathematics and the physical world that classifies the different kinds of quantitative descriptions. So, in circuit theory, variables are measurable properties like resistance, current, and voltage, whereas in computational theory variables are easily distinguished system states, like low/high voltage, or open/closed (mechanical) gates.

In essence, this is why such descriptions are quantitative descriptions *of* something: there is a defined mapping from the description to physical states. Mappings are natural (i.e. simple, straightforward, easy for us to understand) for the class of phenomena that they are explicitly defined over (and to the extent those definitions are specific). For instance, circuit descriptions are natural over the class of voltages, currents, and so on—they are neither overly specific (i.e. picking out material properties) nor overly abstract (i.e. picking out non-electrical properties like fluid flow).

These considerations result in the unsurprising conclusion that quantitative descriptions are natural for the class of physical systems that they are explicitly defined to be descriptions of.

2.1.3. Quantitative descriptions are implementation independent, but to differing degrees

As is again evident from the computation versus circuit descriptions, some quantitative descriptions (e.g. circuit theory) apply only to a subclass of others (e.g. computational theory). As a result, computational theory is more implementation independent than circuit theory. Notice also that circuit theory is independent of many specific material properties of potential circuit elements, for which chemical descriptions may be most natural.

2.1.4. The goodness of a description varies depending on the phenomena of interest

I have more to say on what constitutes a good description in Section 3. These considerations can be preliminary given an agreed characterization of goodness.

If the agreed notion of goodness is partly psychological (e.g. relies on simplicity), and the natural class for a description is too (e.g. also relying on simplicity), then the goodness of a description will vary depending on the natural class of phenomena in question. A description will be best for the phenomena which fall most directly in its natural class.

Just to be clear, this principle does not result in unbridled relativism: so long as we have a consistent measure of goodness across all phenomena, there will be one description which is best for a given class.

2.2. There are four kinds of quantitative description relevant to brain function

Here, I briefly describe each approach, indicate the class of systems it is most natural for, and describe its type of implementation independence.

2.2.1. Computational

Computational descriptions adopt computational theory, which characterizes systems using Turing languages. Such languages are able to describe any Turing Machine (TM) computable function. I take this to have historically been the dominant approach in cognitive science.

2.2.1.1. The natural physical phenomena for computational descriptions are those that are easily discretizable. What I have called Turing languages assume a mapping between the description in the language and distinct physical states. The paradigm case of this is the high/low voltages of silicon transistors mapped to 1s and 0s in the description. In general, any physical system that has easily distinguished (i.e. discrete in both space and time) states can be well-described by such languages. Often such systems are engineered.

2.2.1.2. Computational descriptions are highly implementation independent. Turing Machines are a powerful computational description precisely because they are completely implementation independent. Much has been made of this by functionalists in cognitive science. Notably, this independence means that certainty of the state value is generally assumed (i.e. that it is either 1 or 0). In short, randomness or noise is typically ignored.

2.2.2. Dynamical

Dynamical systems theory, as a mathematical theory, is extremely general (and arguably equivalent to control theory). However, in the context of cognitive systems, a number of researchers have championed the 'dynamical systems theory of mind', which I refer to as DST. DST uses the mathematical theory but adds additional assumptions when applying it to cognitive systems. Given the equivalence between the mathematical theory of dynamical systems and control descriptions, I will discuss DST unless otherwise noted.

2.2.2.1. The natural physical phenomena for DST dynamical descriptions are simple phenomena governed by physical laws. Simplicity is a stated assumption of DST theorists in cognitive science: van Gelder & Port (1995) argue that DST theorists must 'provide a *low-dimensional* model that provides a scientifically tractable description of the same qualitative dynamics as is exhibited by the high-dimensional system (the brain)' (ibid., p. 35). This constraint of low-dimensionality is a severe one, and limits the complexity of such descriptions to simple systems. However, such systems, being continuous, are strictly speaking more computationally powerful than TMs.