Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Neurobiology

Foundations of computational neuroscience

Gualtiero Piccinini[1] and Oron Shagrir[2]

Most computational neuroscientists assume that nervous systems *compute* and *process information*. We discuss foundational issues such as what we mean by 'computation' and 'information processing' in nervous systems; whether computation and information processing are matters of objective fact or of conventional, observer-dependent description; and how computational descriptions and explanations are related to other levels of analysis and organization.

Addresses
[1] Philosophy Department, University of Missouri – St. Louis, St. Louis, MO 63121-4400, USA
[2] Philosophy Department and Cognitive Science Program, The Hebrew University, Jerusalem 91905, Israel

Corresponding authors: Shagrir, Oron (oron.shagrir@gmail.com)

Current Opinion in Neurobiology 2014, **25**:25–30

This review comes from a themed issue on **Theoretical and computational neuroscience**

Edited by **Adrienne Fairhall** and **Haim Sompolinsky**

For a complete overview see the Issue and the Editorial

Available online 10th December 2013

0959-4388/$ – see front matter, Published by Elsevier Ltd.

http://dx.doi.org/10.1016/j.conb.2013.10.005

Introduction

Computational neuroscience has two faces. On one hand, it builds computational models of neural phenomena, analogously to the way computational chemistry, climate science, and computational economics, among others, build computational models of their respective phenomena. On the other hand, computational neuroscience studies the way nervous systems *compute* and *process information*. Thus, unlike computational scientists in most other disciplines, computational neuroscientists often assume that *nervous systems* (in addition to the scientists who study them) perform computations and process information.

Consider for example the neural integrator that converts eye-velocity inputs to eye-position outputs, and thus enables the oculomotor system to move the eyes to the right position [1]. A variety of computational models have been offered for this network [2–6]. In addition, it is assumed that the integrator itself *processes information* about eye velocities and eye positions and produces eye-position codes by *computing* mathematical integration over these eye-velocity encoded inputs.

Is this assumption correct? That depends not only on what nervous systems do but also on what we mean by 'computation' and 'information processing'. This leads us into the foundations of computational neuroscience.

As to computation, there is a precise and powerful mathematical theory that defines which functions of a denumerable domain, such as the natural numbers or strings of letters from a finite alphabet, can be computed by following an algorithm. The same theory shows how to build machines that can compute any function that is computable by algorithm—that is, universal computers [7]. Our ordinary digital computers are universal in this sense until they run out of memory.

But the mathematical theory of computation does not tell us whether and how nervous systems perform computations, and in what sense. This is because the mathematical theory of computation was never intended to be and indeed is not a theory of physical computation, namely, of physical computing systems such as brains. Thus there might be hypothetical physical systems that compute functions that are not Turing machine computable [8,9]. Furthermore, there are many physical systems whose performance is described by computable functions even though we do not say that the systems compute the functions. A rock that is sitting still, for example, does not compute the identity function that describes some of its behavior (or lack thereof).

As to information, there is also a precise and powerful mathematical theory that defines information as the reduction of uncertainty about the state of a system. The same theory can be used to quantify the amount of information that can be transmitted over a communication channel [10]. Again, the mathematical theory of information does not tell us whether and how the brain processes information, and in what sense. So establishing the foundations of computational neuroscience requires more work.

Foundational discussion is important because it articulates the explanatory scope of computational descriptions, the relations between computational level and other levels of description (see Section 'Levels of organization and levels of analysis') and the metaphysical commitments carried by the terms 'information' and 'computation'. Take the oculomotor integrator. We say that it encodes information about eye velocities and positions and that it computes integration. Do we take this statement as a commitment to real, objective facts in the brain, or is it just a useful way to describe the brain used by scientist for heuristic or illustrative purposes? Churchland, Koch and Sejnowski [11], for example, state that "whether something is a computer has an interest-relative component, in the sense that it depends on whether someone has an interest in the device's abstract properties and in interpreting its states as representing states of something else" (p. 48). Others have replied that, on the contrary, whether something computes and processes information is an objective fact [12].

A related question concerns whether every physical object is a computer. Putnam [13] argues that every physical system satisfying minimal conditions implements every finite state automaton. Assuming that to compute is to satisfy Putnam's minimal conditions, this implies that every physical object, including rocks and chairs, computes practically everything! (see also [14]). Many have replied that Putnam assumes a much too liberal notion of implementation (e.g., [15,16]). Chalmers [17], for example, concedes that everything computes something, but insists that only few objects implement the kind of automata that suffice for minds (see [18] for further replies and discussion). Answering these questions depends on how we apply the notions of information and computation to physical systems.

What is information?

Let us begin with information. There is no doubt that nervous systems contain internal variables that correlate reliably with other variables, both internal and external to it. For instance, neuronal spike trains correlate reliably with other neuronal spike trains from other neurons and with aspects of the environment such as light, sound waves, pressure, and temperature.

This is enough to establish that nervous systems carry information in two senses [19]. First, they carry information in Shannon's sense — some of their variables reduce uncertainty about other variables. For example, certain spike trains in the oculomotor system correlate reliably with eye movements. Information in Shannon's sense has to do with the uncertainty that characterizes a process as a whole, including all of the possible alternative messages at once. The Shannon information generated by the selection of a particular message is a function of how many alternative messages may be selected instead and the probability with which any possible message is selected.

By contrast, semantic information has to do with what a particular signal stands for or means. To capture the semantics of a signal, it is not enough to know which other signals might have been selected instead and with what probabilities. We also need to know what a particular signal stands for. Different equiprobable messages carry the same amount of Shannon information, but they may well mean completely different things. We call 'semantic information' the information a signal carries by reducing uncertainty about a specific state of affairs. Nervous systems carry semantic information in the sense that specific states of some of their variables make it likely that other variables (which they reliably correlate with) are in certain specific states. For example, a certain spike train in the oculomotor integrator makes it likely that a specific eye movement is about to occur.

Our opinion is that at least some neural variables carry information in a third sense too—the sense in which neural variables *represent* the environment as being a certain way. Representation is something more than mere semantic information (which in turn is something more than Shannon information). This is because representation can be either correct or incorrect (in which case it is a *mis*representation), whereas mere semantic information, by itself, is neither correct nor incorrect (either a signal raises the probability of a state of affairs or it does not; there is nothing right or wrong either way). In this third sense of information, neural events are not merely correlated with a state of the world but *represent* such a state of the world, which means that they may be either correct or incorrect about how the world is. For instance, let us assume that there are neural events in every speaker's Wernicke's area corresponding to each utterance. Some neural events correspond to true utterances such as "The Moon is a satellite of the Earth." Those neural events truly represent a state of the world, for example, that the Moon is a satellite of the Earth. Other neural events correspond to false sentences such as "the Martians have invaded the Earth." Those neural events *mis*represent the world as different than the way it is.

There are those who think that *neural* representation, as neuroscientists understand it, is insufficient for genuine *mental* representation—that is, the kind of representation that we usually attribute to each other's minds (beliefs, desires, mental images, etc.) [20,21,22•]. Others think that neuroscience already assumes a notion of representation even stronger than the one we just mentioned, to be discussed below [23,24].

What is physical computation?

Let us turn to computation. Some philosophers have tried to explain what it takes for a physical system to perform computations by using notions found in logic and computability or automata theory. They describe computation as program execution [25], syntactic operations [26,27], automatic formal systems [28], or implementation of automata [17]. These notions might apply to digital computers. But, as many have noted, the brain is very different from the familiar digital computers [29••,30–34]. In nervous systems, the functional relevance of neural signals depends on non-digital aspects of the signals such as firing rates and spike timing. Therefore, there is a strong case to be made that typical neural signals are not strings of digits, and neural computation is not, in the general case, digital computation [35•].