Brains as analog-model computers

Oron Shagrir
Department of Philosophy and Program of Cognitive Science
The Hebrew University of Jerusalem
Jerusalem, 91905, Israel

Abstract: Computational neuroscientists not only employ computer models and simulations in studying brain functions. They also view the modeled nervous system itself as computing. What does it mean to say that the brain computes? And what is the utility of the 'brain-as-computer' assumption in studying brain functions? In previous work, I have argued that a structural conception of computation is not adequate to address these questions. Here I outline an alternative conception of computation, which I call the analog-model. The term 'analog-model' does *not* mean continuous, non-discrete or non-digital. It means that the functional performance of the system simulates mathematical relations in some other system, between what is being represented. The brain-as-computer view is invoked to demonstrate that the internal cellular activity is appropriate for the pertinent information-processing (often cognitive) task.

Keywords: Computation, computational neuroscience, analog computers, representation, simulation.

1. Introduction

The term 'computational neuroscience' often refers to two different enterprises. One is the extensive use of computer models and simulations in the study of brain functions. The other is the view that the modeled system itself, i.e., the brain, computes. With respect to the first, neuroscience is not very different from other sciences, where computer simulations and mathematical models are used to study such systems as

stomachs, planetary movements, tornadoes, and so on. With respect to the second, however, neuroscience *is* different. In most other sciences, no one perceives the modeled systems – stomachs, planetary systems, or tornadoes – as computing systems.

An example of this duality is illustrated in Stern and Travis's introduction to the *Science* 2006 special issue on computational neuroscience. They define computational neuroscience as the employment of models and simulations to study the brain:

> Computational neuroscience is now a mature field of research. In areas ranging from molecules to the highest brain functions, scientists use mathematical models and computer simulations to study and predict the behavior of the nervous system. Modeling has become so powerful these days that there is no longer a one-way flow of scientific information. There is considerable intellectual exchange between modelers and experimentalists. The results produced in the simulation lab often lead to testable predictions and thus challenge other researchers to design new experiments or reanalyze their data as they try to confirm or falsify the hypotheses put forward. (Stern & Travis, 2006:75)

Having said that, Stern and Travis immediately turn to the claim that the nervous system itself computes: "Understanding the dynamics and computations of single neurons and their role within larger neural networks is at the center of neuroscience. How do single-cell properties contribute to information processing and, ultimately, behavior?" (2006:75).

Travis and Stern are not alone in expressing the view that the brain computes. Christof Koch opens his *Biophysics of Computation: Information Processing in Single Neurons* with the statement: "The brain computes! This is accepted as a truism by the majority of neuroscientists engaged in discovering the principles employed in the

design and operation of nervous systems" (1999:1).[1] David Marr, in the introduction to *Vision*, writes that "the essence of the brain is … that it is a computer which is in the habit of performing some rather particular computations" (1982:5). And Shadmehr and Wise, in their *Computational Neurobiology of Reaching and Pointing* (2005), state that "according to the model presented in this book, in order to control a reaching movement, the CNS [central nervous system] computes the difference between the location of a target and the current location of the end effector" (p. 143).

But why describe the brain as a computer? There are two related queries here: (1) What, exactly, does it mean to say that an organ or a system such as the brain computes? What distinguishes brains, as computing systems, from other physical, chemical, and biological non-computing systems, such as stomachs, planetary systems, tornadoes, and washing machines? (2) What is the utility of the brain-as-computer assumption in studying brain functions, above and beyond the use of computer models and simulations? After all, computer models and simulations "challenge other researchers to design new experiments or reanalyze their data", regardless of whether or not we describe the brain itself as a computer. I refer to (1) as the meaning question, and to (2) as the utility question.

The "received view" is that there is something about the *structure* of computing mechanisms that distinguishes them from non-computing mechanisms. Computing mechanisms are often described as executing programs, following rules or algorithms, implementing automata or syntactic structures, or as having discrete, digital, or step-satisfaction architecture. The utility of computation is that it explains what the system does (computes) in these structural terms – in terms of the functions

[1] See also Sejnowski, Koch, and Churchland (1988), Churchland, Koch, and Sejnowski (1990), Churchland and Sejnowski (1992).

of its basic components and the relations between them. A detailed and forceful account along these lines has been recently advanced by Piccinini (2007, 2008).[2]

Elsewhere I argue that a structural approach cannot be a general account of computation in the context of physical systems (Shagrir 2006). There is no correlation between the cases in which we view (and explain) physical systems in structural terms and cases in which we view physical systems as computing. There is a discrepancy between these cases even when the viewed system is the nervous system.[3] If I am correct about this, there must be another notion of computation at play in neuroscience that accounts for at least some cases in which the nervous system is described as computing.

My aim here is to outline this alternative notion of computation. This alternative notion, which I label "analog-model," is named after the analog computer and means that the functional performance of the system models in some sense another system or state of affairs. In the analog-model notion, to view the brain as a computer is to take cellular activity that occurs in the brain as reflecting or simulating external mathematical relations between the entities that are being represented.

In the rest of this paper I will explicate this notion by looking at specific computational works in cognitive science and neuroscience. My focus is Marr's computational-level theory of edge detection (Marr 1982, Marr and Hildreth 1980) and the Zipser-Andersen model for locating a target in head-centered coordinates (Zipser and Andersen 1988, Shadmehr and Wise 2005). Marr's work is of special

---

[2] For precursors see Haugeland (1978), Fodor (1980), Cummins (1985).

[3] Piccinini (unpublished) himself notes that central works in neuroscience describe the nervous system as computing, even though these systems do not satisfy his structural constraints.

interest in the present context because he famously separates the computational level from algorithms and mechanisms.[4]

2. The meaning of computation: The analog-model notion

The term 'analog-model' goes back to the old-fashioned analog computer that was used to simulate the functional values of other physical or mathematical systems by performing (computing) its own function. The term, however, does *not* mean continuous, non-discrete, or non-digital. Computing in the analog-model sense (henceforth, AM-computing) can be continuous, non-discrete, or non-digital, but it does not have to be. Chess machines, expert systems and other digital machines are all instances of AM-computing. The term means that the functional performance of the machine simulates some abstract relation in some other dynamics, environment, or system.

A paradigm case of AM-computing is a differential analyzer designed to solve differential equations by using wheel-and-disc mechanisms that performed integration. One example of such an equation is $m\mathrm{d}^2X/\mathrm{d}^2t + c\mathrm{d}X/\mathrm{d}t + k = F$, which describes the forced response of a single degree-of-freedom spring/mass/damper system. X is the displacement of mass m supported by a spring of stiffness k and a viscous damper of rate c, and F is an external force applied to the mass. Another example is the tide-predicting machine designed by Lord Kelvin and constructed in 1873. The machine determined the height of the tides by integrating of ten principal

[4] A detailed interpretation of Marr's conception of computational-level theories is provided elsewhere (Shagrir, unpublished a).

constituents. These constituents were made by means of teeth-wheels that simulated the motion of the sun, moon, earth, and other factors that influence tides.[5]

In the context of neuroscience, AM-computing is associated with the concept of information processing. Information processing roughly means that causal processes in the nervous systems are mappings from one brain state, B_1, which represents some feature, W_1, to another brain state, B_2, which represents W_2 (fig. 1). The concepts of information and representation are notoriously ambiguous and contentious. In the context of the models discussed below, representation and information are apparently associated with selective responses to stimuli, or with some "reliable causal correlation" between activity of cells and certain types of stimuli. In the context of artificial computers, such as the tide-predicting and chess machines, representation has some conventional element.[6]

Now, to say that the brain AM-computes is to say more than it is information processing. The brain computes only if functional relations between the representing states simulate relations between the features that are being represented. The simulation here refers to some similarity between abstract, often *mathematical*, relations among the representing states, and relations among the entities that are being represented. One way to understand this similarity is to say that the mathematical formula that describes the relations between the representing states also describes the relations between the things being represented. In other words, the mathematical formula is a true description of two different functional relations: at the level of the representing states; and at the level of the things that are being represented by these states.

---

[5] See Kelvin's presentation at: http://zapatopi.net/kelvin/papers/the_tides.html.
[6] I discuss this distinction at greater length in section 3.

Another way to put it is to say that the representation function is a sort of isomorphism with respect to the functional relation *f*. Let *f* be the functional relation between the representing states *x* and *y*, namely $f(x) = y$. Let *i* be the representation function, which maps a representing state to a represented feature. To say that a system computes in the analog sense is to state that functional relations between $i(x)$ and $i(y)$ is also *f*. We can formulate this condition as follows: $\boldsymbol{i(f(x)) = f(i(x))}$.

To summarize, a system AM-computes just in case it satisfies two conditions: (a) REPRESENTATION: it is engaged in some information processing, in the sense that it maps one set of representation to another, and (b) SIMULATION: The internal *f*-relations between its "input" state, B_1, and its "output" state, B_2, mirror the "external" relations between what is being represented by B_1 and B_2, meaning that $i(B_1)$ and $i(B_2)$ also stand in some abstract-mathematical *f*-relations. When these conditions are met we say that the system performs the information- processing task by computing *f*.

Let me explicate the notion of AM-computation with two examples from neuroscience.

Edge detection: Edge detection is the information-processing task of extracting (representations of) real physical edges, such as object boundaries, from the retinal image. Figure 2 schematizes the information-processing task. The retinal image, which is an array of activity (intensity values) of the photoreceptors, represents an array of light intensities in the visual field. Each receptor ("pixel") covers a spatial point (or region) that is sensitive to the light intensity in a certain location of the visual scene. The light intensities consist, mainly, in light reflectance, geometry,

illumination of the scene, and the viewpoint. These cells project to cells in V1 that are sensitive to oriented physical edges, and they often signify object boundaries.

In their "Theory of edge detection," Marr and Hildreth (1980, see also Marr 1982, chap. 2) describe this process in terms of co-located zero-crossings of different size-filters of the form $\nabla^2 G * I(x,y)$ (see fig. 3). The term *I* refers to the retinal image. The term $\nabla^2 G * I$ describes the activity of retinal ganglion cells, which perform a certain filtering of the intensity values. Here * is a convolution operator, $\nabla^2 G$ is a filtering operator, G is a Gaussian that blurs the image, and ∇^2 is the Laplacian operator ($\partial^2/\partial x^2 + \partial^2/\partial y^2$) that is sensitive to sudden intensity changes in the image. The zero-crossings of this formula are precisely those places in the image that have sharp intensity changes. Detecting zero-crossings is the task of cells in the early visual cortex (V1). This process takes place through several filters with different Gaussian distributions, and each produces a different set of zero-crossings (fig. 4). The co-located zero-crossings are the basis for the zero-crossing (edge) segments in the raw primal sketch.

What does it mean that the visual system *computes* the co-located zero-crossings of $\nabla^2 G * I(x,y)$ when performing edge detection? According to the analog-model notion, it means that the co-located zero-crossings of $\nabla^2 G * I(x,y)$ point to two different relations: They point to relations between the activity of the retinal photoreceptors and ganglion cells, and the activity of a set of cells in the primary visual cortex. But they *also* point to relations between what is being represented by these cells. The formula states that the relations between light intensity values, signified by *I*, and the light reflectance along physical edges, are those of derivation. I return to this example in the last section.

Locating targets in head-centered coordinates: It has been hypothesized that the PPC (area 7a) of macaque monkeys serves an important role in the information-processing task of locating a target in body- or head-centered coordinates. Experimental results (Andersen et al. 1985) show that the PPC is home to three classes of cells:

(1) Cells that respond to eye position only (15% of the sampled cells). The behavior of these cells can be described by the term $k_i^T e + b_i$, where e is a vector describing the eye orientation with respect to two angular components, and k_i and b_i are the cell's gain and bias parameters.

(2) Cells that are not sensitive to eye orientation (21%), but have an activity field in retinotopic coordinates. The behavior of these cells is described by the term $\exp(-(r-r_i)^T(r-r_i)/2\sigma^2)$, where r is a vector describing the location of the stimulus with respect to the fovea, r_i is the center of the activity field in retinotopic coordinates, and σ describes the width of the Gaussian.

(3) Cells that combine information from retinotopic coordinates with information about eye orientation (57%). The discharge rate of the cells is described by the equation $p_i = (k_i^T e + b_i)\exp(-(r-r_i)^T(r-r_i)/2\sigma^2)$, meaning that they can be seen as planar gain fields that result from linear modulation or multiplication of the terms.

Zipser and Andersen trained a neural network (fig. 5) with the aim of simulating the response properties of the PPC neurons. They used a three-layer network in which the two sets of input units model the behavior of the first two groups of cells. The input layer projects to a layer of hidden units, which aims to model the activity of the third group of cells. The output units encode the target's position in head-centered coordinates; cells with this property were not found in the PPC. Zipser and Andersen's impressive result is that the activity of the *hidden units*,

after the training period, turns out to be very similar to the response properties of the relevant PPC cells – namely, the third-group cells that combine information about eye orientation and the target's retinotopic location. Given this result, Zipser and Andersen hypothesized that there are head-centered target-location cells somewhere in the brain, cells that are the correspondents of the output units on the network model.

In addition to using a computer model to describe the neural activity in the PPC, Zipser and Andersen explicitly describe the modeled activity itself as some sort of computing. They open their article with this statement: "This article addresses the question of how the brain carries out computations such as coordinate transformations which translate sensory inputs to motor outputs" (p. 679). Rick Grush (2001), who analyzes this model, also refers to the computations by the third-group PPC cells: "We can suppose that the function computed by an idealized posterior parietal neuron is something like $f = (e\text{-}e_P)\sigma(r\text{-}r_i)$" (p. 161). And Shadmehr and Wise (2005), who rely on the Zipser-Andersen model in their computational theory of motor control, follow this trail. In a chapter entitled "Computing Target Location," they write that "the CNS appears to compute the location of the target with respect to the end effector" (p. 180), and that "Zipser and Andersen theorized that PPC computes $[R] + [xR] \rightarrow [Cr]$, that is, it combines retinotopic and extraretinal (eye-orientation) signals in order to compute target location in head-centered coordinates" (p. 194).

It is clear that these authors associate the idea of computing with the information-processing task of locating the target in a different coordinate frame. But computing is not just information processing. To say that the PPC cells "really are computing a function" (Grush 2001: 159) – namely, that they compute $(k_i^T e + b_i)\exp(-(r\text{-}r_i)^T(r\text{-}r_i)/2\sigma^2)$ when performing the task of transforming reference frames – is to say that this equation refers to mathematical relations at two different levels. The formula

refers to the mathematical relations, abstracted from the electric properties, between the two groups of "input" PPC cells and the group of "output" PPC cells. But, as Grush insightfully notes, it also refers to some complex mathematical relations between what is being represented – in our case, between eye orientation and stimulus retinotopic – and "stimulus distance from preferred direction relative to the head" (Grush 2001: 161). In other words, the relations between eye orientation and stimulus retinotopic location, and the stimulus distance from preferred direction relative to the head is (roughly) also that of multiplication. In effect, a cell fires most strongly when the distance from the preferred direction relative to the head is low.

3. Objections: abstraction, chauvinism, and liberalism

Three objections to my proposal might be raised. One is that it is not well formed, confusing physical and abstract categories; another that it is too demanding (or chauvinistic); finally, that it is too permissive (or liberal), making everything a computer.

Abstraction: One might argue that the proposal is not well formed.[7] The concern is that the function *f* operates on such mathematical entities as numbers, whereas the representation relates to physical properties of entities – e.g. electrical activity and distal objects in the environment. Thus SIMULATION is not well formed: On the one hand, the operation $\boldsymbol{i(f(x))}$ means that the representation function relates mathematical entities, whereas, on the other, the term $\boldsymbol{f(i(x))}$ means that the function *f* operates on physical properties.

[7] This point was raised by James Ladyman.

There are general philosophical issues about abstraction in the sciences that cannot be addressed here. Bearing this in mind, my response is that the function *f* is mathematical, in the sense that its domain and range are mathematical entities such as real numbers, geometrical relations, set-theoretic structures, and so forth. Thus the function *f* is defined over *numbers* (or other mathematical entities) and not physical entities. These entities are mathematical values as magnitudes that *abstract* from pertinent physical properties. At one level, the function relates numbers that abstract from the representations (e.g., electrical cellular activity). At another, it relates magnitudes that abstract representational contents (e.g., distal objects in the environment). One way to make this precise is to read the term *i*(x) as a relation *i*(<x>) between two mathematical magnitudes that abstract from physical properties. The value of *i*(<x>) is <y>, where the magnitude <y> abstracts from the physical property y, which is the value of *i*(x). Thus the term ***f(i*(x))** (should be read as ***f(i*(<x>))**) relates magnitudes that abstract from representational content – e.g., light intensities in the visual field – and other mathematical values. The term ***i(f*(x))** (should be read as ***i(f*(<x>))**) relates mathematical magnitudes that abstract from the representing property – e.g., the electrical activity of cells in V1 – and from representational content – e.g., physical edges.

A related concern is that the "inner" function and the "outer" function are often defined over *different* mathematical entities.[8] The worry is that the abstraction relation can be one when applied to the representing properties, but another when applied to representational content. One example is a machine in which “inner” relations are defined over “syntactic” properties, whereas the “outer” relations are defined over numbers. The Zipser-Andersen model is another example: The "inner" relations are

[8] This point was raised by Mark Sprevak.

defined over real numbers (abstracted from electric activities) and the "outer" relations are defined over geometrical properties.

I do not have a detailed account of the abstraction relation in such cases. I can only note that in many of these cases we see that there is a "higher-level" abstraction, under which the "inner" and "outer" relations are the same. In the Zipser-Andersen model, the geometrical properties are also described in terms of real numbers. In other cases there might be higher set-theoretic or additional structures, under which the "inner" and "outer" relations are similar.

Chauvinism: One might worry that the account is in some respects too demanding, or chauvinistic, in the sense that it excludes many computing systems.[9] Here is one phenomenon that raises chauvinism concerns: The neuroscience of PFC is largely concerned with planning processes – processes that formulate means for the future attainment of an agent's goals. Moreover, such processes are often characterized computationally.[10] But what are the relevant isomorphisms in such cases? The problem seems to be that in such cases, the "direction of fit" is the wrong way around. Plans don't represent how the world is. They represent how we seek to make it.

Another example: There is plenty of research on the neuroscience of arithmetic cognition.[11] Moreover, substantial amounts of this work are computational in character. But in this case, what is the relevant isomorphism between represented and representing? The obvious answer is that the isomorphism obtains between neural states/relations and *numbers*. But if this is so, then the characterization of AM-computation appears to have the weird consequence that computational

[9] This concern and the examples that follow were raised by Richard Samuels.
[10] See, for example, Norman and Shallice (1986).
[11] See, for example, Dehaene (1993), Butterworth (1999).

neuroscientists studying arithmetic incur commitments in the metaphysics of mathematics. E.g., they are committed to anti-fictionalism about mathematical entities; and if the relevant representation relations obtain between *physical* systems, then they are committed to anti-Platonism.

One final problem case: In the case of hallucinations and perceptual illusions, there is no isomorphism between representing system and represented system since there is no (actual) represented system at all. But if this is so, then, according to the analog-model notion, no computation has occurred. This seems highly counter-intuitive. Moreover, it fits poorly with the practices of neuroscience. Much vision science is concerned with illusions, and the processes that produce such illusions are often characterized computationally.

In response I want to emphasize that it might well be that different scientists have in mind different conceptions when describing cognitive and nervous systems as computing. My claim is not that that AM-computation covers all the cases in which the brain is viewed as a computer, but that it covers many of them. Having said this, I still insist that AM-computation can cover the cases of planning, arithmetic cognition, and illusions. AM-computation requires that there be some representations, but it does not require that the represented features be environmental or even physical features. The represented states can be future physical states, counterfactual situations, abstract states, or even other intentions.

In a forthcoming work I analyze a relatively simple system, known as the oculomotor integrator, the goal of which is keeping the eyes still between saccades. It is a sort of memory of eye positions: the inputs encode the velocity of the saccadic horizontal movement, and the outputs encode the current eye position (motor neurons then read out the encoded eye position). Usually this system is described in terms of a

line-attractor neural network – see, e.g., Seung (1996, 1998) – and falls out of the structural account of computing. Still the system arrives at the eye position by computing integration. According to the analog-model notion, the system computes integration because the mathematical relation between the represented entities, eye velocity and eye position, is also that of integration. In fact, the system was called an integrator in the first place because of this external relation of integration. The outer relations of integration motivated scientists to look for the parallel "inner" integration relation in the brain itself.

This example indicates that AM-computation applies not only to visual perception, but to other processes, such as memory and motor control tasks. But the example shows more. It shows that the represented entities do not have to be physical entities in the environment. They can also be physical states of the body – velocity and positions of the eyes (see also the Zipser-Andersen model). Moreover, Robinson (1989) and Goldman et al. (2002) argue that the same integrator is used for another task: keeping the eyes (say, on a target) when a head movement occurs. When employed for this task, the system computes the *desired* eye position, by integrating on inputs that encode the head's velocity. When doing this, the computation is not defined on representations of *actual* physical states, as these states are not yet there, but on representations of *future* or even *possible* states.

This example hopefully addresses the worries about planning. It indicates that there is no difficulty in describing the relation between the actual physical feature of target's velocity and a non-actual state of desired (planned?) eye position in terms of integration. In fact, describing this relation in terms of integration is *essential* for the explanation of the VOR as performing the task of keeping the eyes fixed on the same location. Without it, we cannot understand why performing integration – i.e., the

"inner" relations that are of integration – is appropriate for moving the eye to the right place (see the discussion in the last section). So perhaps not all cases of planning are instances of AM-computing, but this does not mean that AM-computing is not relevant for tasks of planning.

AM-computing does not exclude the representation of abstract entities. In fact, arithmetical operations, "implementation" of automata, and logical inferences are paradigm cases of AM-computing. In McCulloch and Pitts networks the AND, OR, XOR and other gates can be seen as representing and simulating the states of abstract automata.[12] Moreover, much of the work in arithmetical cognition assumes representations of numbers; debates in the field are often about *kind* of arithmetic representations – for example, whether representations of numbers are abstract or specific ("abstract" here means that cells respond to quantity, and "specific" refers to the medium of presentation, i.e., dots, numerals, etc.).[13] Thus AM-computation might exclude some views about the metaphysics of mathematics, but, as far as I can see, it is consistent with the current work in arithmetic cognition.

The case of illusions and hallucinations raises important questions about individuation, and mis-representation vis-à-vis mis-computation, which I cannot discuss here. Instead, I focus on one example in order to show that AM-computation is consistent with these phenomena. Assume that the early visual processes keep detecting the zero-crossings of second-derivatives. However, the system is removed to an environment that consists of surfaces that sharply change reflectance across their solid faces. In this environment what look like edges are actually solid faces. Does the system compute the second-derivatives and their zero-crossings?

---

[12] See Minsky (1967, chapter 3) for a detailed and clear presentation of the relations between McCulloch and Pitts neural networks and automata.

[13] See, e.g., Cohen-Kadosh et al. (2007) and Piazza et al. (2007)

There are at least two cases here. One is that the system does not represent at all. In this case, I insist, there is no computation; the system is no different from any other (non-computing) physical systems whose behavior is described in terms of first- (or second-) derivative. A second, more likely, case is that the system (mis-) represents solid faces as object boundaries. In this case, I would say, the system computes: not only does it (mis)-represent, but, in addition, the "outer" mathematical relations, between the things that are being represented, are also those of derivation. This example shows that a system might still compute even if it misrepresents.[14]

Liberalism: Is everything a computer? We considered the issue of whether or not the analog-model notion of computation is chauvinistic. But one might also think that it suffers from the converse concern: that the notion is too liberal, in that it implies that every physical system computes something, or even everything. For one thing, one might claim, every system can be interpreted as representing something, and so every system satisfies REPRESENTATION. The SIMULATION condition does not solve this problem, as every system could be used to simulate something, e.g., its twin system. Moreover, SIMULATION is symmetrical, and so the simulated phenomena, e.g., light intensities, compute as well. One could also claim that AM-computation is subject to Putnam's (1988) and Searle's (1992) claim that every physical system implements every finite-state automaton (FSA). On their construction, we can always take the states of the physical system to represent the states of an (any) automaton,

[14] A third case, and perhaps the most interesting one, is that the system acquires a new function, e.g., from visual to auditory (see Egan, 1995, and *this volume*). I argue elsewhere (Shagrir 2001) that in this case, the system might compute a different (mathematical) function, say, integration, which is the one underlying the auditory task (see also Sprevak, *this volume*). Note that in this case the system simultaneously implements, but does not compute, derivation. Cases in which the system fails to implement derivation (in normal environments) but computes integration can count as mis-computation (relative to the task of edge-detection).

and then the functional relations of the physical system also simulate the functional relations in the automaton. If this is correct, then every physical system computes every FSA-computable function!

The easy way out is to deny that everything represents and, hence, computes. My stomach perhaps indicates what I ate yesterday, but does not represent it. My washing machine perhaps indicates what I wore yesterday, but does not represent it. My brain, however, not only indicates light intensities, but really represents them. On the other hand, light intensities do not represent my brain states. Nor do my brain states really represent the states of every automaton, though they might represent the states of some of them. This reply, of course, is incomplete, pending some adequate account of representation. But this is hardly a problem peculiar to my account. After all, it is a presupposition of much philosophy, neuroscience, and cognitive science that some states are representational whilst others are not.[15]

I think that something like this is right. But the answer should be refined. The difficulty with this reply is that it assumes that computation requires a fairly strong notion of representation, whereas we all know that there are computations that are defined over fairly weak, derivative, sorts of representation. That my desktop represents the pieces on a chess board is a derivative matter, pending my interpretation of the states of my desktop. We thus have to say more on why the states of my stomach and the states of my washing machine could not be interpreted as representing and, then, as computing.

Why are we are so worried about liberalism? There are two concerns that liberalism might raise. One concern, emphasized by Searle (1992), is that the claim

[15] This line of response was suggested to me by Frances Egan, Gualtiero Piccinini, and Richard Samuels.

that the brain computes is vacuous, in the sense that its truth-value is not a matter of facts about the brain ("computation is not discovered in the physics"; p. 225), but is a matter of our whims ("computation is not an intrinsic feature of the world. It is assigned relative to observers"; p. 212). The other concern is that my account draws the line in the wrong place. A conception of computing should draw a line between computing systems, such as brains and desktops, from non-computing systems, such as stomachs, tornadoes, and washing machines. But it now turns out that my account fails to do this, since it implies that everything AM-computes.

Let me start with the first concern, that the account is vacuous. One can note that there must be something wrong with this objection. There is a trivial sense in which everything represents, yet we (might) think that there is a fact of the matter to whether and what the brain represents. The same is true for computation: Everything is a computer in some trivial sense, and yet it does not follow that the claim that the brain computes is senseless. The computational properties of the brain are discovered, not assigned relative to an observer. We simply have to be more sensitive to facts about representation and simulation.

First, we have to note that there are different sorts of representations. One of my favorite accounts is that of Dretske (1988, chap. 3), who distinguishes between three types of representational systems. In a conventional, or derivative, system *type I*, the states of the system receive their representational powers from *us*, from the ways we *interpret* the states and *use* the system. Desktops are representational systems *type I*. The states of stomachs could be seen as a representational system *type I*, had we interpreted them as representing some states of an automaton. In a conventional system *type II*, the relation between the representing states and their representational content is rooted in some causal (and lawful) correlation between them. This causal

correlation makes the states *natural* signs or indicators.[16] What makes them representing states is that their *function* is to indicate what they do. The system is thus conventional in that this function is a matter of our decision, or the way we use the system. Examples of *type II* systems are certain gauges and some robots.

Finally, a natural system of representation is one in which the function to indicate is natural, or "intrinsic" in the sense that it is independent of our decision. Arguably, brains are natural systems of representations: photoreceptors are not only natural signs of light intensities, but, in addition, it is their "natural" function to indicate these features. Stomachs are not natural systems of representations. The states of my stomach really (naturally) indicate what I ate yesterday, but this indication is not its natural function. Perhaps we could use these indicative powers for one purpose or another, but even then stomachs would be conventional systems *type II* and not natural systems of representations.

The second issue is about simulation. I agree with Putnam and Searle that every physical system can be seen as implementing at least one automaton, and typically more than one. But this claim should be qualified. First, as many have already pointed out, Putnam's and Searle's arguments rest on too weak a notion of implementation. Their universality claim, of every system implementing *every* automaton, disappears if we introduce adequate constraints into the notion of implementation.[17] Second, we have to examine whether SIMULATION is satisfied or not with respect to a given representational system. Thus my stomach might satisfy SIMULATION, when its states are interpreted as representing, in the *type I* sense, the states of some automaton. In this sense the stomach computes. However, the states of

[16] Dretske follows Grice's distinction between natural and non-natural meaning (Grice 1957). Note, too, that Dretske associates the term 'information' with natural signs, whereas I associate 'information' with representation. For further discussion see Piccinini and Scarantino (*this volume*).
[17] See, e.g., Chalmers (1996), Scheutz (2001).

my stomach do not satisfy SIMULATION when its states are taken to represent, in the *type II* sense, the food I ate yesterday. The function that describes the digestive processes does not simulate the mathematical properties of the food. So, relative to these representational powers, my stomach does *not* compute.

The upshot of all this is that *the* claim that the brain computes, the one made in the context of cognitive and brain sciences, is highly contentious. It asserts that: (a) the brain is a natural system of representations; and (b) the functional relations among *these* representations simulate the functional relations between the things that are being represented. We know of no other systems that compute in this sense. True, we could view every system as computing in some sense. We could even view the brain, when we view its states as representational systems of *type I* or *II*, as computing many other things. But, then, so what? It simply does not follow that the claim that the brain computes has no empirical content.

We can now turn to the second objection, of failing to draw the line between computing and non-computing systems. The reply is this: My account implies that every system computes in the sense that it *could* be a computer, but it does not imply that it computes in the sense that it is one. According to my account, an AM-computer is either (i) a "natural" computer, in the sense that the simulation processes are defined over a natural system of representations, or (ii) a "conventional" computer, in the sense that its simulation processes are defined over a conventional system of representations. Brains belong to the first category. Desktops, defined over representational systems of *type I*, and robots, defined over representational systems of *type II,* are confined to the second. Stomachs, tornadoes, and washing machines are neither. Their processes are defined over neither natural nor conventional systems of representations. True, my account implies, as it should, that stomachs *could* be

(conventional) computers: We could interpret its states as representing the states of some automaton and simulating its behavior. But we do not, and so stomachs do not compute.

4. The utility of computation: Solving the appropriateness problem

I have so far addressed the meaning-question of what is computation. It is now time to turn to the utility-question of invoking the computational approach in studying brain function. In the context of conventional computers, the utility of computing is closely associated with the fact that the system is a computer. But in the context of natural computers there need not be such an association. That brains are natural computers does not entail that we must utilize their computational properties. Neurons have many natural properties – e.g., mass, size and shape – to which we seldom refer in the study of brain function. Why, then, refer to the computational properties of nerve cells?

The query consists of two questions that correspond to the two conditions of computing: What is the utility of referring to the representational properties of the brain? And, what is the utility of referring to its simulative properties when describing the brain as an information-processing system? I do not have much to say about the first question. My aim is to outline an answer to the second question, which is that we invoke the computational perspective in order to show how an information-processing system "tracks" the environment.

When we describe a system as an information-processing system (fig. 1), the following question arises: Why does the mapping process that starts from a representation of W_1, i.e. B_1, lead to a representation of W_2? Why does the output of

this mapping, B_2, represent W_2 and not something else, say W_3? After all, we start with one state B_1, which represents W_1, then proceed to another state, B_2, through a causal process that takes place inside the system. We should thus wonder why it is that the state we arrived at, B_2, encodes information about W_2, e.g., object boundaries. Why doesn't it lead to a representation of the target's color or shape? Why does it lead to anything meaningful at all?

As I see it, Marr himself sharply formulates this question. After demonstrating that *what* the visual system does is detect zero-crossings of $\nabla^2 G * I$, Marr turns to ask *why* the mapping from $I(x,y)$ to the co-located zero-values of $\nabla^2 G * I$ is *appropriate* to the task of edge detection. Here is what he says about this:

> Up to now I have studiously avoided using the word edge, preferring instead to discuss the detection of intensity changes and their representation by using oriented zero-crossing segments. The reason is that the term edge has a partly physical meaning – it makes us think of a real physical boundary, for example – and all we have discussed so far are the zero values of a set of roughly band-pass second derivative filters. We have no right to call these edges, or, if we do have a right, then we must say so and why. (1982:68)

It is clear from this passage that Marr is concerned with the veridical relation between the visual system and the visual scene – e.g., between co-located zero-crossings that are the basis of edge segments in the raw primal sketch, and physical edges such as object boundaries. Marr and Hildreth say that "the concept of 'edge' has a partly visual and partly physical meaning. One of our main purposes... is to make explicit this dual dependence" (1980:211). The questions more specifically are: (a) Why are the zero-crossings that result from different-sized filters related to the same physical feature? As Marr (p. 68) puts it, "there is no a priori reason why the zero-crossings obtained from the different-sized filters are related"; (b) Why is this physical feature often an edge in the physical sense, e.g., an object boundary? Again,

even if the co-located zero-values are related to a single physical feature, there is no a priori reason why this feature is a physical edge such as object boundary.

I put these questions under the heading of "the appropriateness problem," as the task is to explain why the mathematical function *f*, which describes the relation between B_1 and B_2, is appropriate for the information-processing task, which is defined in terms of W_1 and W_2. Why is it that the causal processes that take place within the PPC, and are described by a mathematical formula – say $(k_i^T e + b_i)\exp(-(r-r_i)^T(r-r_i)/2\sigma^2)$ – are relevant to locating the target in eye-position-independent coordinates? And why does detecting the zero-crossings of $\nabla^2 G * I$ have anything to do with edge detection: Why does it lead to a representation of boundaries and not, say, to a representation of color?

The appropriateness question is not a skeptical problem. We do not scrutinize our knowledge that B_2 represents W_2. That zero-crossing segments often represent oriented object boundaries is evident from electro-physiological cell-recording experiments. The question, rather, is why cells in V1 respond to object boundaries and not to colors, given that this response is mediated, as it were, through internal causal processes that start from *different* representations, i.e., cells that respond to light intensities.

A simple answer is that the inner B_1- B_2 relations are correlated with the outer W_1-W_2 relations, and that this correlation has been established by some evolutionary or learning adaptation-based process. The performance of the Zipser-Andersen network model was achieved by back-propagation learning techniques; the performance of the PPC is established by a different adaptation-based procedure.[18] The question, however, goes beyond evolution and learning. Evolution and learning

[18] See Zipser and Andersen, p. 683.

correlate things that can fit with each other. They cannot correlate the performance of a simple perceptron with non-linear structures. And they cannot correlate an automaton that consists of AND-gates and OR-gates with an XOR structure. We thus want to know what is it that facilitates the correlation between B_1- B_2 relations and W_1-W_2 relations? Why is it that the B_1- B_2 relations *can be* correlated with the W_1-W_2 relations?

One can answer *similarity*: B_1-B_2 relations are similar to W_1-W_2 relations, and this similarity underlies the possibility of correlation. But in the present context, the similarity cannot be at the level of physical properties. After all, the physiological properties of the brain are quite different from the physical and optical properties that make up our visual field. But there might be another kind of similarity. This similarity might be at the level of mathematical properties, namely, that the B_1- B_2 mathematical relations between the representing states are similar to the W_1-W_2 mathematical relations between the represented features. This is where the analog-model conception of computation kicks in. We say that performing the *f*-relations between B_1 and B_2 is appropriate to the information-processing task, which is defined in terms of W_1-W_2, because the nervous system computes *f*.

Let us inquire in greater detail why alluding to AM-computation serves to address the appropriateness problem. We start with a neural state B_1 that stimulates the activity of another neural state B_2. We see that B_1 selectively responds to some worldly feature W_1, and B_2 to W_2. But we still want to explain what is it about B_1 representing W_1 and the causal relation from B_1 to B_2 that makes B_2 represent W_2.

Ideally, the structure of the (computational) explanation is as follows:

- $i(B_1) = W_1$.
- $f(B_1) = B_2$.
- $f(W_1) = W_2$.

- $f(i(\mathrm{x})) = i(f(\mathrm{x}))$.

Therefore: $i(\mathrm{B}_2) = \mathrm{W}_2$.

The first premise states that B_1 represents W_1, which is something based on the electrophysiological single-cell recording experiments. The next task is formulating the mathematical relation (function), *f*, between B_1 and B_2, which in the case of edge-detection is given by the zero-crossings of $\nabla^2\mathrm{G}*I$. The mathematical relations between the pertinent groups of PPC cells is given by the equation $p_i = (k_i^{\mathrm{T}}e + b_i)\exp(-(r-r_i)^{\mathrm{T}}(r-r_i)/2\sigma^2)$.

Establishing the third premise is often trivial, but sometimes can take sophistication and effort. The Zipser-Andersen model is an example of the easy case. Given that the term $\exp(-(r-r_i)^{\mathrm{T}}(r-r_i)/2\sigma^2)$ refers to information about retinotopic location, and $k_i^{\mathrm{T}}e + b_i$ refers to eye orientation, it is obvious that the term $(k_i^{\mathrm{T}}e + b_i)\exp(-(r-r_i)^{\mathrm{T}}(r-r_i)/2\sigma^2)$ refers to some combined information of the two. Zipser and Andersen do not even bother to argue for this.

Marr is an example of the more complex case. It does not immediately follow that if the term $I(x,y)$ refers to the array of light intensities in the visual field, then the co-located zero-values of different-scale filtering of $\nabla^2\mathrm{G}*I$ stand for physical edges. That it does is a contingent fact about our visual environment, and something that should be argued for. This is where Marr appeals to "physical constraints", which are facts and assumptions about our physical world.

The first three premises do not yet entail the conclusion that B_2 represents W_2. To establish this conclusion, we have to allude to the fourth premise, namely, that the system is indeed a computer, in the analog sense. When doing this, we have an argument on the basis of which we can explain why B_2 represents W_2, and can

conclude that the system performs the pertinent information-processing task by computing *f*.

5. Summary

Two questions were asked about brains as computers. The first question was “What does it mean to say that an organ or a system such as the brain is a computer?” The second was “What is the utility of the 'brain-as-computer' assumption in studying brain functions?” My aim here has been to address these questions by means of an analog-model conception of computation. According to this conception we say that a brain is a computer when we take it that internal mathematical relations, between the representing features, reflect abstract external relations, between the represented features. The computational approach is applied in order to explain why the cellular activity of the representing cells is appropriate for the information-processing task that is defined, at least in part, by the features that are being represented by these cells.

Acknowledgements: I am grateful to the participants of the Workshop in Computation and Cognitive Science for discussion and comments. Special thanks are due to Frances Egan, Gualtiero Piccinini, Richard Samuels, and Mark Sprevak for detailed critical comments on early drafts of this article. This research was supported by The Israel Science Foundation, grant 725/08.

References:

Andersen, R.A., Essick, G.K., & Siegel, R.M (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230(4724), 456-458.

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Chalmers, D.J. (1996). Does a rock implement every finite-state automaton? *Synthese*, 108, 309-333.

Churchland, P.S., Koch, C., & Sejnowski T.J. (1990). What is computational neuroscience? In E.L. Schwartz (Ed.), *Computational Neuroscience* (pp. 46-55). Cambridge: MIT Press.

Churchland, P.S., & Sejnowski, T.J. (1992). *The Computational Brain*. Cambridge: MIT Press.

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron*, 53, 307–314.

Cummins, R. (1985). *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Dehaene, S. (1993). *Numerical Cognition*. Oxford: Blackwell.

Dretske, F. (1988) *Explaining Behavior*. Cambridge: MIT Press.

Egan, F. (1995). Computation and content. *Philosophical Review*, 104, 181-204.

-- (*This volume*). *A modest role for content*.

Fodor, J.A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, 3, 63-73.

Goldman, M.S., Kaneko, C.R.S., Major, G., Aksay, E., Tank, D.W., & Seung, H.S. (2002). Linear regression of eye velocity on eye position and head velocity suggests a common oculomotor neural integrator. *Journal of Neurophysiology*, 88, 659-665.

Grice, P. (1957). Meaning. *The Philosophical Review*, 66, 377-388.

Grush, R. (2001). The semantic challenge to computational neuroscience. In P. Machamer, R. Grush & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 155-172). Pittsburgh: University of Pittsburgh Press.

Haugeland, J. (1978). The Nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, 2, 215-226.

Koch, C. (1999). *The Biophysics of Computation*. New York: Oxford University Press.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 207(1167), 187-217.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice-Hall.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self Regulation: Advances in Research and Theory* (Volume 4, pp. 1-18). New York: Plenum.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, 53, 293-305.

Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science*, 74, 501-526.

– (2008). Computers. *Pacific Philosophical Quarterly*, 89, 32-73.
– (*Unpublished manuscript*). Digits, strings, and spikes: Empirical evidence against computationalism.

Piccinini, G., & Scarantino, A. (*This volume*). *Computation vs. information processing*.

Putnam, H. (1988). *Representations and Reality*. Cambridge: MIT Press.

Robinson, D.A. (1989). Integrating with neurons, *Ann. Rev. Neurosci.*, 12, 33-45.

Scheutz, M. (2001). Computational versus causal complexity. *Minds and Machines*, 11, 543-566.

Searle, J.R. (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press.

Sejnowski, T.J., Koch, C., & Churchland, P.S. (1988). Computational neuroscience. *Science*, 241(4871), 1299-1306.

Shadmehr, R., & Wise, S.P. (2005). *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. Cambridge: MIT Press.

Shagrir, O. (2001). Content, Computation and Externalism. *Mind*, 110, 369-400.
– (2006). Why we view the brain as a computer. *Synthese,* 153, 393-416.
– (*Unpublished manuscript a*). Marr on computational-level theories.
– (*Unpublished manuscript b*). Computation, San Diego style.

Sprevak, M. (*This volume*). *Computation, individuation, and the representation condition*.

Stern, P., & Travis, J. (2006). Of bytes and brains. *Science*, 314(5796), 75.

Zipser, D., & Andersen, R.A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679–684.

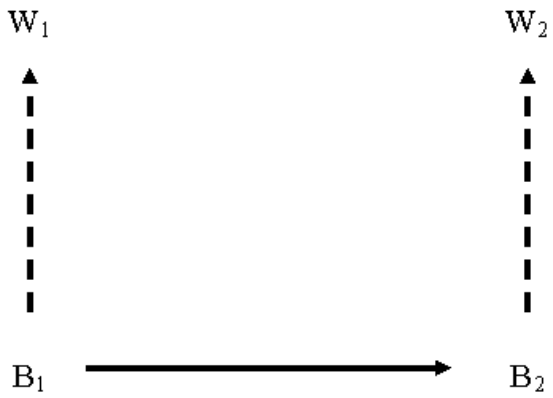

Figure 1: An information-processing task. A causal process (arrow) from a state/event B_1 to another state/event B_2, whereby B_1 represents (dashed arrow) an event/property W_1, in the visual scene, and B_2 represents object/property W_2.

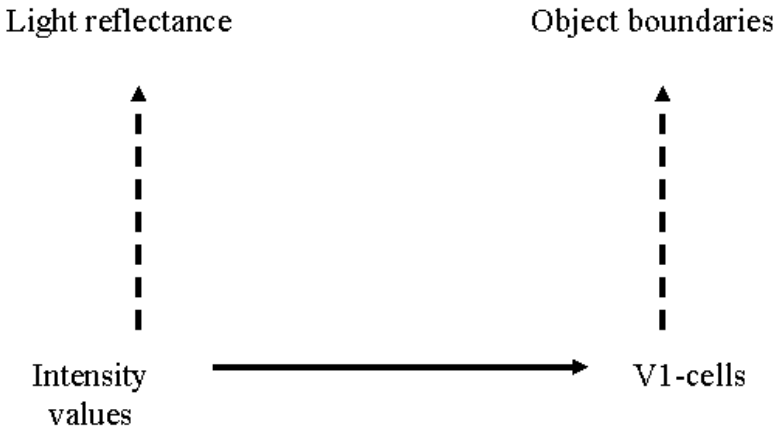

Figure 2: Edge detection as an information-processing task. The intensity values of the photoreceptors represent light intensities in the visual field that consist, among other things, in light reflectance. The activity of cells in the primary visual cortex (V1) respond to oriented "edges" such as object boundaries.

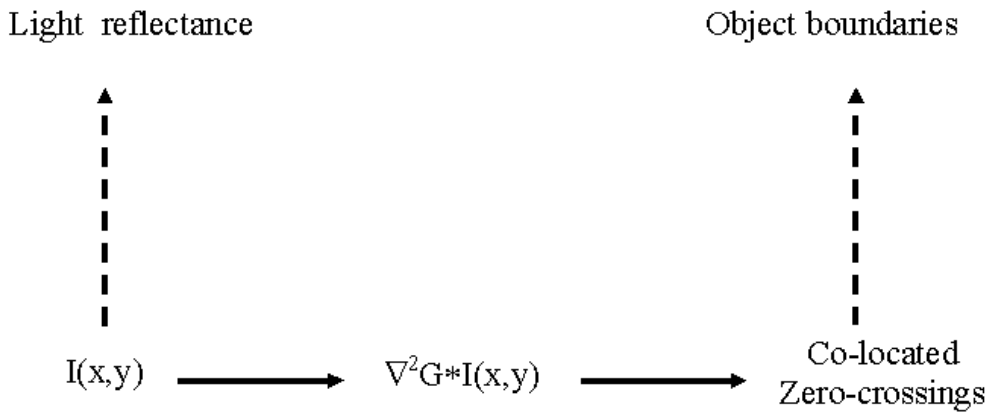

Figure 3: Edge detection: The retinal image *I*(x,y) is convoluted through the filtering operator ∇^2G, whereas G is a Gausian and ∇^2 is a second-derivative (Laplacian) operator. Early vision processes include several filters with different Gausian distributions, and each produces a different set of zero-crossings. The co-located zero-crossings often signify edges such as object boundaries.

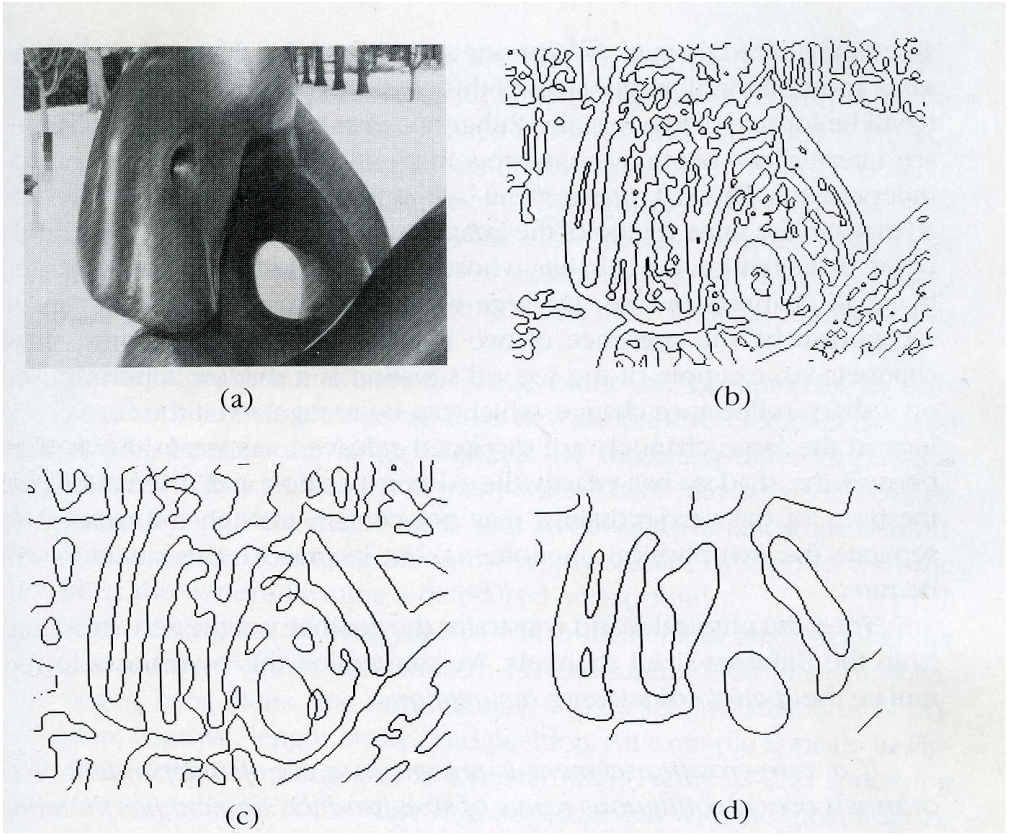

Figure 4: Different sets of zero-crossings. The image (a) has been convoluted with different-sized filters: (b), (c) and (d) show the zero-crossings thus obtained. Note that many of the fine details obtained through the smaller-size filter, (b), are not obtained with the larger-size filters, but that some of the zero-crossings obtained in a larger-size filter do not appear in the smaller-size filtered image. (Reprinted by permission of Royal Society Publishing, from D. Marr and E. Hildreth, "Theory of edge detection," *Proceedings of the Royal Society of London, Series B, Biological Sciences*, 207, 187-217 (1980)).

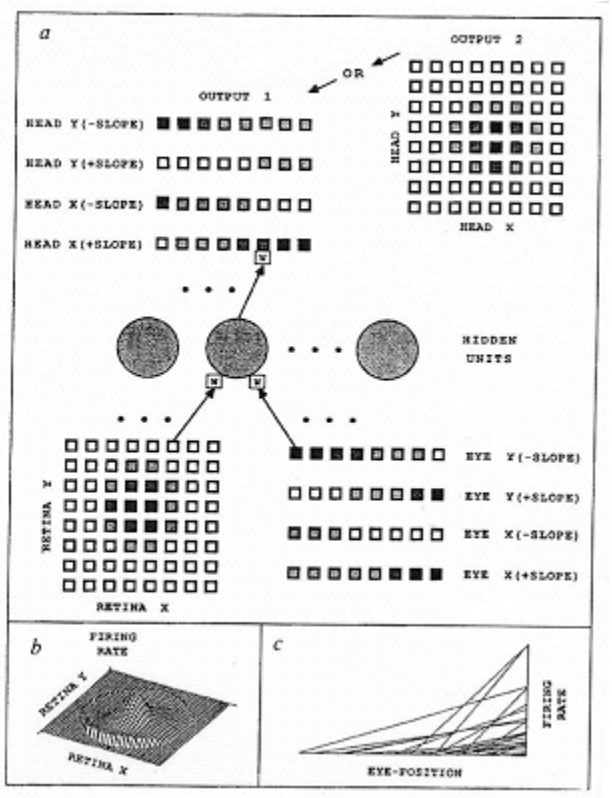

Figure 5: The Zipser-Andersen model. (a) The three-layer network, where the two sets of input units stand for the retinotopic location cells (bottom left) and eye-orientation cells (bottom right). The hidden units are meant to model the behavior of the third-group of PPC cells. The units of the output layer (two versions) stand for cells that encode the head-centered location. The network is trained through a supervised learning technique. (b) Area 7a visual neuron receptive field with a single peak near the fovea. (c) A composite of 30 area 7a-eye-position units, whose firing rates are plotted as a function of horizontal or vertical eye deviation (Reprinted by permission from Macmillan Publishers Ltd: D. Zipser & R.A., Andersen, "A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons," *Nature*, 331, 679–684, copyright (1988)).