Structural Representations and the Brain

Abstract: In *Representation Reconsidered*, William Ramsey suggests that the notion of structural representation is posited by classical theories of cognition, but not by the "newer accounts" (e.g., connectionist modeling). I challenge the assertion about the newer accounts. I argue that the newer accounts also posit structural representations; in fact, the notion plays a key theoretical role in the current computational approaches in cognitive neuroscience. The argument rests on a close examination of computational work on the oculomotor system.

1. Introduction

A working assumption in brain and cognitive sciences is that the brain is a representational, information-processing, system. My aim in this paper is to examine the claim that the brain employs *structural representations*. In a recent book, William Ramsey (2007) suggests that the notion of structural representation is posited by classical theories of cognition, but not by the "newer accounts" (e.g., connectionist modeling). In what follows I challenge the assertion about the newer accounts. I argue that the newer accounts posit structural representations, and that, in fact, the notion is deeply entrenched in the current computational approaches in neuroscience. My argument rests on a close examination of computational work on the oculomotor system.

Though my argument is made in the context of the classical/non-classical debate, my interest here is not in this debate *per se*. Neither is my goal to play down the dichotomy between classical and non-classical, nor to play it up. Also, I do not suggest that the "newer accounts" are superior in some way to classical theories of cognition. My interest is in the conceptual

foundations of the current computational approaches in cognitive science. I urge that, contrary to the impression cast by Ramsey, the notion of structural representation is not exclusive to classical theories of cognition. I advance three claims about the current computational approaches in cognitive neuroscience.[1] First, the notion of structural representation – in the sense of an internal model – is central and explicit in recurrent networks; my focus here is attractor neural networks. This is important since Ramsey grounds most of his arguments on the multi-layer, feed-forward, networks, whereas the bulk of the computational work in cognitive neuroscience today invokes recurrent networks. Second, Ramsey detaches structural representations from the "receptor notion" of representation, whereas the two are often linked together. Third, in addition to its explanatory role, the notion of structural representations also plays a methodological role, in discovering what the nervous system computes. I will thus conclude that the notion of structural representation is very much present, playing a key theoretical role in the current computational work in cognitive neuroscience.

The paper proceeds as follows. I start with a brief presentation of the notion of structural representation (Section 2) and then explicate Ramsey's claim that the newer accounts of cognition do not posit structural representations (Section 3). I next turn to challenge Ramsey's claim. After discussing in some detail one case-study, that of the computational work on the oculomotor memory (Section 4), I argue that the notion of structural representation is at the heart of current computational approaches in cognitive neuroscience (Section 5). In the final two sections I consider two objections to the memory network; one pertains to the claim that its states

[1] I say cognitive neuroscience as neuroscience has taken center stage within cognitive science; it would not be an exaggeration to say that the field of cognitive psychology is being absorbed into cognitive neuroscience (Quartz 2008).

are truly internal (Section 6), the other to the claim that its states are truly representational (Section 7).

2. Structural representation

The term "structural representation" refers to representations that preserve the structure of the (target) domain being represented, or, as Swoyer puts it, "the *pattern* of relations among the constituents of the represented phenomenon is mirrored by the pattern of relations among the constituents of the representation itself" (1991: 452). This notion encompasses such representational systems as family trees, maps, and various measurement schemes. It also applies to cognitive representational systems, which are our focus here. Thus Johnson Laird is famous for postulating mental models whose structure "is identical to the structure of the states of affairs … that the models represent" (1983: 419). And Shepard and Chipman talk about *second-order isomorphism*, which refers to some parallelism "between the relations among different internal representations and the relations among their corresponding external objects" (1970: l).[2]

Swoyer introduces a detailed analysis of the notion. The basic idea is that the representing domain and the represented domain have "the same structural or formal features" (1991:457). They share a high-order – formal or mathematical – structure, which Swoyer calls a *shared structure*. To be a little more precise, let *A* and *B* be two domains, each of which includes individuals and relations.[3] We would say that the two domains are of the same *similarity type* just in case there is a one-to-one and onto function, *c*, from the full domain of relations of *A* to

[2] See also Palmer (1978) and Edelman (1998).
[3] See Swoyer for the precise and detailed characterization that specifies the domains as intentional relational systems.

that of *B* which maps each *R*-relation of the first to some ~~*R*~~-relation of the same type in the second. The two domains have a *shared structure* – they are isomorphic – just in case this function is type-preserving; that is, for every *n*-ary relation *R* and *n*-tuple of individuals (in *A*) the following condition holds:

$<x_1,\ldots, x_n> \in R$ if and if $<c(x_1),\ldots , c(x_n)> \in c(R)$.[4]

Finally, we would say that *A* is a *structural representation* of *B* just in case *A* and (a subset of) *B* have a shared structure.[5]

Swoyer then introduces several relaxations, modifications, and distinctions that make the notion more applicable; these are necessary as the isomorphism requirement holds only as an ideal.[6] I will mention here a few points that are relevant to the discussion below: One is that the shared structure is often a matter of approximation, certainly when we discuss biological systems such as the nervous system. Another is that there are different kinds of representational systems. There are "static" systems, such as the measurement schemes we impose on nature, and there are "dynamical" systems, such as the nervous system. Thirdly, the function *c* is not always one-to-one, onto, or total; sometimes it may even be a many-to-one ("non-function") mapping. In short, we might view a system as a structural representation even when *c* does not respect all of the relations in the original system but only some of them.

Cummins (1989) introduces a special (relaxed) case of structural representation. This case is of dynamical physical systems whose input-output relations respect relations between the things that the inputs and outputs represent. He calls this kind of (structural) representation

---

[4] The condition can be extended to high-order relations.
[5] There is disagreement over whether structure sharing is sufficient and/or necessary for representation. I return to discuss this issue in section 7.
[6] See Swoyer pp. 470ff.; the modified definition is on page 474. Other less-than-isomorphism characterizations are in terms of partial isomorphism (French and Ladyman 1999; Da Costa and French 2003), homomorphism (Bartels 2006), and similarity (Giere 2004).

simulation representation. Cummins explicates the notion with a simple adding machine. The machine is presented in the form of (what Cummins calls) the London-Tower-Bridge scheme (fig. 1). The bottom span is an input-output function, *g*, which maps *physical* states of the system – e.g., button-pressing sequences – to other physical states of the system – e.g., display states. Cummins refers to this mapping by the term *satisfaction* (a system satisfies a function *g* if it produces *o* as its output on input *i* just in case *g*(*i*) = *o*, where 'produce' refers to a *causal* process that starts from an input *i*, which is a physical entity, and terminates with an output *o*, yet another physical entity).[7] The top span is the mathematical function *plus*, which maps pairs of numbers <*m*,*n*> into their values *m*+*n*. The vertical arrows stand for an interpretation function, *I*, which maps physical states of the system onto numbers. It maps the pair of physical buttons <*M*,*N*> onto the pair of numbers <*m*,*n*>, and the physical display *D* onto the value *m*+*n*.

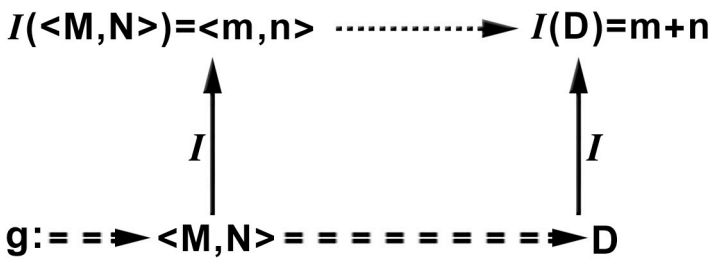

Figure 1: Cummins's London-Tower-Bridge. The bottom span is the input-output function satisfied (or computed), *g*; the double-dashed arrow is a causal process by which the system satisfies (computes) *g*. The top span is the function *plus*. The interpretation function, *I*, maps the inputs and outputs of *g* to the inputs and outputs of *plus*.

When all is in place, we have simulation representation. The physical states of the device – button-pressing sequences and displays – are interpreted as representing numbers, and the causal

[7] According to Cummins, some systems (arguably, cognitive systems) satisfy functions by *computing* them. Computing, then, is a special kind of satisfaction; its (causal) processes have the form of step-satisfaction (pp. 91-92).

relations, *g*, between the input and output physical states "mirror" the *plus*-function relations between the numbers. Or, as Cummins puts it, the interpretation function, *I*, fulfills the following simulation condition:

$g(x) = y$ iff $+(I(x)) = I(y)$.

This condition amounts to saying that *g* is isomorphic to *plus*, and an interpretation function that fulfills this isomorphism condition is precisely simulation representation (pp. 96-97).[8] The term "simulation" highlights the fact that the *g*-relations, between the button-pressing inputs and the display output, model, as it were, the *plus*-relations between the input $<x,y>$ and output number $x+y$.

Cummins (1989: 108-111) then argues that a similar London-Tower-Bridge scheme is found in the context of cognitive systems: "The CTC [Computational Theory of Cognition] is just the thesis that what works for addition will work for cognition" (p. 111); the only difference is that what is being interpreted is a cognitive function. I argue elsewhere, however, that the shift from addition to cognition is not as smooth. The adding device is a very special case of simulation representation that does not automatically extend, without further refinements, to cognition.

3. Ramsey's argument

In *Representation Reconsidered*, Ramsey advances an extended argument whose conclusion is that the notion of inner representation does important explanatory work in classical theories of cognition, but not in the newer accounts of cognition (2007: 3). Ramsey identifies classical

[8] The affinities to Swoyer's definition are apparent, where *I* parallels the *c*-function, *g* the *R*-relation and *plus* the ~~*R*~~-relation.

computational theory of cognition (CCTC) with accounts that posit "internal symbolic representations" (p. 2), and associates it with a host of other labels, such as "Good-Old-Fashioned-Artificial-Intelligence", "Physical Symbol Hypothesis", "Digital Computational Theory of Mind".[9] The term "the newer accounts" refers to accounts of cognition that "radically depart" from CCTC; among them are "connectionist modeling, cognitive neuroscience, [and] embodied cognitive accounts" (p. 2). These accounts also characterize internal elements as "representations", but not as symbol-based ones (p. 3).

Ramsey's argument has the following structure. He contends that classical accounts posit two (genuine) notions of representation.[10] One is that of IO representation, which is not discussed here. The other is that of *S-representation*, which Ramsey identifies with what Swoyer "refers to as 'structural representation'" (pp. 78-79); he also mentions that "Cummins calls this notion of representation 'simulation representation'" (p. 80). Ramsey argues that both notions meet his "job description challenge" (roughly, that they do an explanatory work *qua* representation) and, hence, can be truly counted as notions of representation. These notions, he argues, are exclusive to classical accounts of cognition. The newer accounts typically posit two other notions of representation, both of which fail to meet the job description challenge. One is the "receptor notion" of representation; Ramsey says that "things described as representations in this sense are not really representations at all" (p. 118). The other is that of tacit representation, which is "metaphysically and explanatorily bankrupt" (p. 152). Our focus here is the key assumption that the newer accounts do not posit the notion of S-representation.

---

[9] The label "classical" was coined by Fodor and Pylyshyn (1988).

[10] Prior to discussing these notions, Ramsey (chapter 2) dismisses the "standard interpretation" of representation – the way philosophers often think about representations – in classical accounts.

At first glance, there is no reason to think that structural representations are exclusive to classical systems, at least when we think of simulation representation. Consider the example of the adding device, which maps an input representation of the number-argument $<x,y>$ to the output representation of the number-value $x+y$. This machine, according to Cummins, posits simulation representations: the function being satisfied, *g*, simulates the *plus* function, whose arguments and values are numbers. The architecture of the mechanism by which our machine satisfies *g* is irrelevant to simulation representation. Our machine can satisfy the function *g* by means of either a classical or a non-classical, e.g., connectionist, mechanism. The only thing that matters to simulation representation is that the mechanism maps the input representation of the number-argument $<x,y>$ to the output representation of the number-value $x+y$. Whether this mapping is executed by a Turing machine or by a neural network has no bearing on the notion of simulation representation. It is thus not surprising that Cummins himself thinks that connectionist theories typically posit simulation representations (1989, chapter 11).

Ramsey (2007) agrees that this picture of representation is found in both classical and connectionist architectures. But he distinguishes between systems whose input-output relations simulate certain relations, and systems whose inner workings also preserve structure. He argues that we find the explanatory force of structural representation only in the latter systems: "it is important not to confuse theory-neutral specifications of the explananda with the internal explanatory posits of particular cognitive theories. Since cognitive processes are defined with representational states as their end-points, it is a mistake to treat this notion of representation as belonging to the CCTC, or invoked by the CCTC. Since most theories treat types of input-output transformations as their starting points, the input and output themselves are not part of any particular theory's explanatory apparatus" (p. 71). Ramsey's point is that the notion of simulation

representation posits entities that are part of the explananda of cognitive theories. This notion is indeed found in many different cognitive theories, including the newer computational approaches, whose aim is to explain the input-output transformations. But the notion of structural representation that we are looking at should be part of the "theory's explanatory apparatus", and as such should essentially refer to *internal* entities as representations. Ramsey refers to this explanatory notion as S-representation and argues that it is pervasive in classical accounts but is not found in the newer approaches.

It turns out, then, that Ramsey's notion of S-representation is not exactly that of simulation representation. Both notions share the idea that the input-output relations simulate relations in the target domain, namely, those between the entities that the inputs and outputs represent. But S-representation goes beyond this in requiring that internal elements in the source system also mirror properties and relations in the target domain. The requirement, to be sure, is not that the representing domain and the represented domain are perfectly isomorphic; the demand is that "there is a significant type of isomorphism" (p. 80) between the domains. Ramsey associates this notion with that of an *internal model*, which is the source of its explanatory force.[11] He illustrates this idea of modeling with a family tree that posits "a representational model that invokes elements that serve to mirror the conditions and states of affairs and entailment relations" (p. 82). This family tree serves to explain how one finds out whether Jeff is or is not an uncle of Julie. The explanation rests on the fact that the tree models the familial relations, in that its internal nodes stand for members of the family and the various connecting arrows among the nodes stand for various familial relations among those members. Ramsey

[11] The association of structural representation with models is pervasive; see, e.g., Frigg and Hartmann (2006). See also Swoyer (1991), who connects this modeling feature with the ability to explain how a system solves problems and performs, more generally, "surrogative reasoning".

argues that we find this kind of representation (and explanation) in classical systems (SOAR, SHRDLU) including (classical) cognitive systems. According to these classical theories of cognition, "the mind/brain is claimed to be using a computational model or simulation, and the model/simulation is constructed out of symbols that are thereby serving as S-representations" (pp. 80-81).[12] Ramsey concludes that the notion of S-representation "serves as an important, distinct, and explanatorily valuable posit of classical computational accounts of cognition" (p. 79). The newer approaches, however, do not posit this notion of S-representation.

One could object to Ramsey's claim that the newer approaches do not posit the required notion of S-representation, i.e., that of an *inner* model whose structural elements encode structural features of the world. Indeed, critics of Ramsey have argued just that. Mark Sprevak (2011) contends that "it is far from clear why a receptor based notion cannot fulfill the role of a surrogate too. And the receptors posited in cognitive neuroscience typically do: a face detector or edge detector are understood, not just as detectors in isolation, but as part of our wider model of the world, and their activity is considered apt surrogates for the presence of distal faces or edges in reasoning".[13] Garzón and Rodríguez focus their discussion on connectionist, multi-layer, feed-forward networks (like NET talk) that seem to present structural isomorphism or similarity with their targets.[14] Ramsey discusses these networks in detail, arguing that while a cluster analysis might reveal that there is an isomorphism between the hidden unit layer and the target's domain,

---

[12] We can note the difference in terminology between Ramsey and Swoyer. Swoyer refers to the whole system, e.g., a family tree, as a structural representation. Ramsey refers to the whole system as a "model" and to its constituents as S-representations. I will mostly follow Swoyer's convention.

[13] Indeed, elsewhere I argue that Marr's theory of edge-detection posits the edge-detection process as modeling the structural features in the world, e.g., the mathematical (derivation) relation of light reflectance along object boundaries (Shagrir 2010a).

[14] See also Paul Churchland (2007) who stresses the similarity relations, or isomorphism, found in many networks between high-dimensional relations (e.g., geometrical congruence) in the state-space of the representing neural network and high-dimensional relations in the represented domain "in the world". See also O'Brien and Opie (2001; 2006; 2009).

these units do not constitute concrete inner states that the system can go “look up” for problem solving; instead, it is an abstract by-product of the fact that the system goes into specific states in response to specific inputs. He says that "mere clustering in the response profile doesn’t show this [i.e., that they are representations]" (2007: 145); we call these states representations only because we assume in advance that they serve as representations. Garzón and Rodríguez insist, however, that the isomorphism lies not in clustering, but in the architecture of such networks, i.e., in the "relations between the components of the modeling nets and the constituents of the target domain" (2009: 308). Lastly, Rick Grush (2008) argues that there is yet another paradigm of representation "that Ramsey doesn't directly address at all, namely modeling based on modern control theory"; Grush mentions forward models and his notion of emulator.[15] Referring to Grush's work and to the SINBAD neural network(s) (Ryder 2004), Ramsey says that "there are also a few connectionist-style theories that invoke model-based representations" (2007: 80, n. 5). But Ramsey apparently think that these are only very isolated examples.

I also argue that non-classical theories posit a notion of S-representation in the stronger sense of a model-based inner state. In this respect my claim is not very different from what others have claimed. But I want to highlight three further points. First, Ramsey does not seriously consider a central paradigm in neural network theory, that of recurrent neural networks – mainly attractor networks – and its descendants in control and dynamical theory. This paradigm has influenced cognitive science since the 1980s, and it encompasses more than the forward models mentioned by Grush. Ramsey mentions "few connectionist-style theories" that exemplify internal models, but he ignores a whole paradigm in neural network theory and its applications in cognitive and brain sciences. I will point out that these recurrent networks, which

[15] See Grush 2004; see also Eliasmith and Anderson (2003).

can hardly be called "connectionist-style", constitute much of the computational work in cognitive neuroscience today. In fact, computational cognitive neuroscientists invoke recurrent networks with the explicit aim of positing the notion of an internal model.

Second, Ramsey assumes that S-representation and the receptor notion are mutually exclusive. Ramsey is right of course in assuming that much of the experimental work, e.g., electrophysiological experiments, assumes one or another version of the receptor notion. But he is mistaken in concluding that current theoretical work in cognitive neuroscience, which relies on these results, does not posit S-representations. As I will indicate below, the notion of representation posited by current theories often satisfies the conditions of the receptor notion but also those of S-representation. Third, I will suggest that in addition to its explanatory role, the posited notion of S-representation plays an important methodological role in cognitive neuroscience. This notion helps in assessing, and even discovering, what function is being computed, on the basis of "external", e.g., ecological, constraints. More generally, this notion of representation links the work of experimentalists and theoreticians.

If I am right about this, then it appears that the structural, model-based, notion of representation plays a key theoretical role in the current computational approaches in cognitive neuroscience. I think that Ramsey went astray by focusing on the multi-level feed-forward networks of "higher-level" phenomena in which there is no explicit aim to introduce the notion of an internal model. I do not argue that the notion of an internal model is implicit in these networks; but I will briefly question Ramsey's claim that the *input-output* simulation relations are merely part of the explanandum; contrary to Ramsey, I suggest that the simulation relations do have explanatory value (Section 6). This will imply that the connectionist-style multi-layer networks posit, at the very least, an explanatory notion of (input-output) simulation-

representation. To demonstrate my claims, I survey and discuss some of the computational work on the oculomotor system.

4. **A brief review of computational work on oculomotor memory**

Oculomotor control has been studied extensively by experimentalists and theoreticians for decades.[16] Within oculomotor control, saccadic and inter-saccadic fixations have taken center stage. Saccadic eye movements rapidly shift the two eyes, with regard to the head, from one place in the visual world to another.[17] Between saccadic movements, the brain keeps the eyes still; experimental results show that normal humans are able to hold their eyes stationary at arbitrary positions for up to dozens of seconds, even in complete darkness (Becker and Klein, 1973; Hess et al., 1985). The brain can track the current eye position even after the stimulus has gone; in this respect, the brain employs a short-term memory of eye positions. When the memory is damaged there is a constant drift of the eyes to a null point. Our focus here is the memory for *horizontal* eye movements. In cats, monkeys, and goldfish, this system appears to be localized in two brainstem nuclei, the nucleus prepositus hypoglossi (NPH) and the medial vestibular nucleus (MVN) (Moschovakis 1997).

Using the language of state-space analysis, we can think of the memory as consisting of states, whereas each state, Si, is a representation of a different eye position, Ei. The dynamics in

[16] For reviews see Robinson 1989, Glimcher 1999, and Leigh and Zee 2006.

[17] See Glimcher 1999, who classifies eye-movements into two broad classes. Gaze stabilization movements stabilize the visual world on the retina when the head/body is moving. The *vestibulo-ocular reflex* (VOR) keeps the visual world stable on the retina when the head is moving. The *opto-kinetic reflex* stabilizes the visual world when the head is stationary (e.g., when one is looking from a train's window). Gaze-aligning movements include voluntary and reflexive saccades and smooth pursuit movements that allow one to track a moving target. Yet a third class (in binocular animals) includes vergence movements that allow the eyes to fixate the eye on the target at different distances.

the memory can be seen as a transition from one state, say S_1, which represents one eye position, E_1, to another stable state, S_2, which represents another eye position, E_2, by computing some function *g* that operates on inputs that indicate change in eye position (fig. 2).

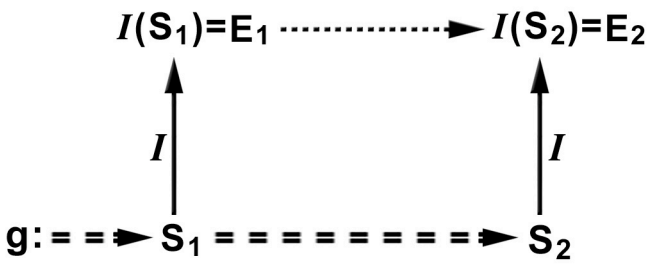

Figure 2: The oculomotor memory – an information-processing description. The memory system moves from state S_1 (which represents E_1) to state S_2 (which represents E_2), by satisfying (computing) the function *g* that operates on inputs that indicate change in eye position. The representation/interpretation function is *I*.

How does the nervous system implement the memory states? And what is the transition function *g* that is being computed? Let us start with the first question. Experimental findings show that when the eyes are still, the pattern of neural activity is constant in time, and that for every eye position, the pattern of activity is different and persistent. These findings have encouraged modelers to describe the memory system as a multi-stable recurrent network. The use of multi-stable attractor neural networks to implement memory is widespread (Amit 1989). The dynamics of these networks is often described in terms of an energy landscape whose minima are stable states. To implement an eye position, we can think of each state as coding a different eye position. However, the typical multi-stable networks do not seem appropriate for memory of the eye position. The reason is that the attractors are discrete, but the encoding of the eye position in the neural activity requires a continuous, analog-graded code. Theoreticians have thus suggested that the memory of eye position is implemented in a recurrent network with

continuous line attractor dynamics.[18] A new stimulus perturbs the state of the memory network away from the line of fixed points, and the network gradually relaxes on a new point along the attractor line; this point encodes the current eye position (fig. 3).

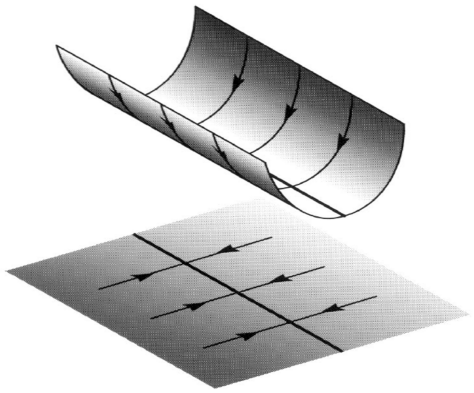

Figure 3: State-space portrait of the attractor network: All trajectories in the state-space of the memory network flow toward a line attractor. Each fixed point along the line is a persistent state which represents a different eye position. (From Seung 1996:13340)

The mathematical details are quite complex, but the crucial features of the network are easy to explain. First, the (attractor) network has no designated input and output units; all cells are interconnected to all other units, and each cell receives "external" inputs from outside (pulse saccadic stimuli). Second, a single cell does not represent an eye position; only a collective, "total", state of the network is a candidate for being such a representation. Respectively, each point in the state-space portrait signifies not the activity of a single cell, but the activity of a collective state; at each time the memory network is in one of the points in the landscape portrait, but it "aspires" to the line attractor; the points along the line attractor are precisely the collective states that encode eye positions.

[18] See Canon and Robinson (1985); Seung (1996, 1998). For a general framework see Eliasmith and Anderson (2003: 250ff.).

Third, there are no synaptic changes ("learning"), at least in the simplified case; the weights are fixed in advance. The important idea is that the weight matrix *Wij* has only positive feedbacks that are tuned to have a single unity eigenvalue (this produces an energy function with no curvature, as in fig. 3). The rest of the eigenvalues have real parts that are less than unity; this condition ensures stability, namely, that the bottom trough of the energy function is perfectly level (Seung 1996). In real biological systems, however, where these idealizations do not hold, the issues of robustness and stability become acute. There are suggestions to handle these with learning, i.e., synaptic plasticity, but their effectiveness and biological reality are questionable (Seung 1996, 1998). Yet we also have to remember that the biological system itself is not perfect, and that the memory of eye position is gradually corrupted over time. In human subjects, during gaze-holding in the dark, there is generally a slow, systematic drift, usually less than one degree per second in normal subjects (Becker and Klein, 1973; Hess et al., 1985).

Let us turn to the second question: What is the transition function from one stable state, which represents a pre-saccadic eye position, to another state, which represents the current eye position? Experimental studies show that the system converts transient eye-velocity-encoding inputs into persistent eye-position-encoding outputs. It was thus concluded that the network is a *neural integrator*.[19] The system moves from one stable state along the line attractor, Si, which represents one eye position, Ei, to another stable state, Sj, which represents another eye position, Ej, by performing integration on eye-velocity-encoding inputs *u* (fig. 4). The dynamics is of integration in that the system keeps accumulating the pulse inputs it has received. When new input pulses come in, they are added to or subtracted from (depending on direction of movement)

[19] See Robinson (1989). It is hypothesized that the neural integrator serves other systems, e.g., the *vestibulo-ocular system*. In this case, the integrator gets velocity-coded vestibular inputs that determine how quickly and in what direction the head is moving (Robinson 1989; Goldman et al. 2002).

the previous summation. These summations are just the stable states of the network; hence, the network is moving from one persistent point, Si, to another, Sj. The new summation is "memorized" by the network (i.e., staying in Sj) even when the velocity-coded stimulus is gone. The motor neurons "read" the current position and stabilize the eye in the new position by controlling the length-tension relationships of the muscles.

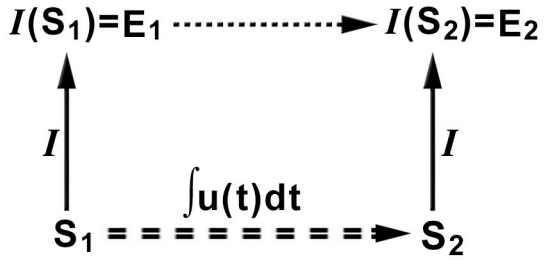

Figure 4: The dynamics of the memory system. The system moves from one stable state, say S_1, to another stable state, say S_2, by integrating on pulse saccadic inputs, *u(t)*. Each state, S_i, represents a different eye position E_i, and the transient pulse inputs encode eye velocity.

5. The memory network as structural representation

Scientists describe the oculomotor memory network as a representational system, where each persistent state of the network "is the *internal representation* of eye position" (Seung et al. 2000: 269; italics in the original); it "encodes the angular position of the eyes" (p. 259). But what do they mean by "representation"? At one point, Sueng et al. (2000) refer the reader to "Experimental Procedures for definition" (p. 263). The experimental procedures are those of the single-cell recording experiments that associate representational content with a selective response to external stimuli; in our case the pulse inputs are (causally) correlated with eye-velocity; whereas certain persistent collective activity in the neural network is correlated with a

certain eye position (and the current activity with current position). These procedures are the basis of the "receptor notion" representation.

What is perhaps less noticeable is that the memory network also constitutes a structural representation, in the sense of an internal model. In fact, theoreticians use this kind of network precisely to posit internal models.[20] The state-space of the network mirrors the space of eye positions. Each state, Si, along the line attractor encodes a different eye position, and the distance between two states, Si and Sj, corresponds to the distance between two eye-positions, Ei and Ej. We thus have a sort of isomorphism between the representing network and the eye: the function that maps the stable states, Si's, to the corresponding eye-position states, Ei's, is type-preserving, in that the distances between two states mirror the distances between eye-positions.[21] The state-space of the network could be seen as a *map* whose line attractor corresponds to the space of eye positions. By moving from one state, Si, to another, Sj, one could reflect a transition from one eye position, Ei, to another, Ej. Indeed, the oculomotor system employs this map in accomplishing various neuro-cognitive tasks. In the present context, the motor neurons "read out" the current state in order to move the eyes to a new eye position and keep the eyes in this current position. Thus the network is not just an abstract by-product of the fact that the system goes into specific states in response to specific inputs. It consists of concrete inner states that the oculomotor system can go “look up” for problem solving.

The task of the network is not just to mirror the space of eye-positions, but to fixate the eyes on a *specific* ("current") position. To accomplish this task, the network should get some

---

[20] Thus Seung writes that "according to modern computational theories, biological motor control is performed by an internal model… A wealth of experimental data indicates that the internal model used for maintaining eye position is the integrator, which has been localized to specific brainstem nuclei. The nature of the internal model is also known: it appears to be a recurrent network with a continuous attractor" (1988:1253-4).

[21] More formally: I(S-distance(Si,Sj)) = E-distance(I(Si),I(Sj)) = E-distance(Ei,Ej). S-distance is the distance in state-space, whereas E-distance is horizontal distance between angular eye positions.

indication that will enable it to compute this position. This is where integration kicks in. The network gets as inputs pulses that encode eye velocity; the integration of the coded velocity is just the distance between previous and current eye positions. The network is constructed such that it computes integration over the entering stimuli. Given that the network stabilizes on a new point ("state") along the line attractor, this state must be the one that encodes the current eye-position. In short, we have an isomorphism also at the level of dynamics. The distance between the two states is just the integration over the transient saccadic electrical inputs to the network, *u(t)*, with respect to time; these inputs encode the eye velocity *v*(*t*). The (horizontal) distance between the two eye-positions, Ei and Ej, is just the integration of the eye velocity, *v(t)*, with respect to time. Thus by computing integration the network mirrors the transition from previous to current eye position (fig. 5).[22]

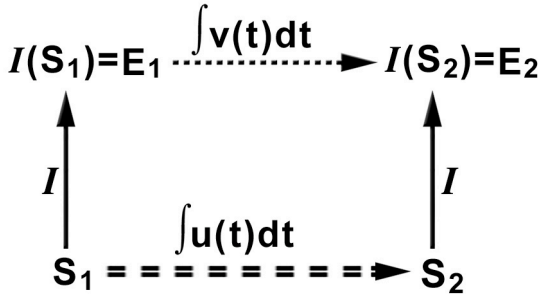

Figure 5: S-representation in the dynamics of the network. The representing domain and the represented domain share the structure of integration. The dynamics in the neural memory, from one state, S_1, to another, S_2, is described as integration over the pulse saccadic inputs, *u*(*t*). The relation between two (represented) eye positions, E_1 and E_2, is described in terms of integration over the eye velocity, *v*(*t*), encoded by *u*(*t*).

[22] More formally: $I(\text{S-distance}(S_{previous},S_{current})) = I(\int u(t)dt) = \int I(u(t))dt = \int v(t)dt = \text{E-distance}(E_{previous},E_{current}) = \text{E-distance}(I(S_{previous}),I(S_{current}))$. The isomorphism is achieved through the shared structure of integration expressed here by the term $I(\int u(t)) = \int I(u(t))$. We note that the integration is over real-number magnitudes abstracted from the electric properties of the nervous system and from the (represented) physical properties of the eyes (for further discussion of this point see Shagrir 2010b).

Explaining how the oculomotor system controls the eyes refers to the fact that the system implements an internal model in the form of a recurrent neural network with multiple stable states, one for each eye position. In this respect the memory meets the "job description challenge" (I elaborate on this in sections 6 and 7). What I want to emphasize now is that in addition to its explanatory role the notion of S-representation often plays a methodological one. The notion of S-representation guides the theoretical studies in discovering what the system computes. That is, the function that is being computed is revealed through the "external", overt, relations between the things that are being represented. To see this methodology at work, let us consider the course of investigation in the case of the memory system. The investigation started with the electrophysiological and anatomical experimental studies which have (roughly) shown that pre-synaptic stimuli encode eye velocity and that the post-synaptic activity (collectively) encodes eye position (fig. 2); these experiments assume in a way the "receptor notion" of representation. Next, looking at the target domain (i.e., the eye), theoreticians quickly realized that the external "top-span" relations between the represented velocity and positions are those of integration. On the basis of this observation, it was then *inferred* that the internal "bottom-span" *g*-relations (at a more abstract level) between the representing states must be of integration too. In fact, from the observation that the (high-order) *external* relations, between the represented entities, are those of integration, the system was called "an integrator" (Robinson 1989). In other words, the theoreticians *assume* that the oculomotor memory S-represents the external relations, in the target domain. They assume that if the high-order relations in the target domain are those of integration, the oculomotor memory itself must compute integration. Of course, after constructing the network model (for computing integration) one can conduct further experiments that confirm its validity. But the point is that the construction of the network model starts with

the premise that it computes integration, a premise whose evidence is the overt external relation between (the represented) position and velocity.[23]

The theoretical work describes the memory system as that of S-representation. Does it also describe the memory as a non-classical system? I doubt that anyone will insist that the memory network (as described above) is a classical system. But it is still worthwhile to mention how different it is from discrete-like systems, e.g., Turing machines, which are the paradigm cases of classical systems. First, the memory state-space does not consist of separable states, not even "in principle". The state-space is "dense", in that it is a continuous line attractor, whereas each fixed point along the line is, ideally, a different state. Second, the activity of the cells is not "symbolic", in that we cannot assign the symbol '1' to some given electrical activity and '0' to some other electrical activity. Nor can we easily assign to some streams of action potentials ("spikes") a string of such symbols. Rather, the activation of cells is a *continuous* function (described by differential equations; see, e.g., in the appendix). Lastly, the dynamics of the network does not consist, in any obvious sense, of discrete steps. It is a continuous relaxation toward the closest minimum point, where all the minimum points lie along a line (attractor). Even the language used to describe the dynamics of the system is not that of computation theory, but the language used in control and dynamical theory. Given all that, we can safely say that this

[23] I asked a theoretician: "But why do you assume that the internal relations must be those of integration? After all, the only evidence you have for this assumption is that the *external* relations are those of integration?" His reply was that this is just a truism: if the relations between velocity and position are those of integration, the internal relations must be that of integration too: "What else can it be?" The question we do ask, he added, is not whether the system computes integration (as it surely does), but *how* it computes integration. I certainly do not want to underestimate the theoretical work behind the "how", i.e., the theory of neural networks with a line attractor. But the reply not only indicates that the theoreticians assume that the memory system is one of S-representations; it indicates that they do not seriously consider other possibilities; they do not consider the possibility that the oculomotor memory attains the neuro-cognitive memory task by not positing S-representations.

memory network is not classical; if one insists that it is classical, then it is hard to comprehend what a *non*-classical system is.[24]

I have shown that at least one non-classical theory – that of the oculomotor memory – posits S-representation(s). But the example is by no means unique. Recurrent multi-stable networks have been used also for various tasks of problem solving, where the state-space S-represents the space problem, e.g., games (Hopfield and Tank 1985, Rumelhart, Smolensky, McClelland, and Hinton 1986; a more detailed example of problem-solving is discussed in the next section). The multi-stable recurrent neural networks, when applied in the context of cognition, often invoke S-representations. These networks are often applied in the context of (associative) memory whose state-space S-represents the target domain (Hopfield 1982, Amit 1989); the distinctive feature of the oculomotor memory is that the persistent states are arranged in a line attractor. This idea of internal model has become very widespread in motor control and is used to account for other motor systems.[25] Ramsey's discussion focuses on the multi-layer connectionist networks, and on some work in dynamical system theory (pp. 204 ff.). But the point is that much of the current computational work in cognitive neuroscience invokes the recurrent multi-stable networks.[26] This is perhaps due to the fact that much of the current computational work addresses specific neural mechanisms, and that cognition is rarely perceived in terms of one-directional perception-to-action processes.

The upshot is that the notion of S-representation is widespread in the newer computational approaches to cognition. It plays the explanatory role that is often assigned to

---

[24] This is not to claim, however, that there is a clear-cut distinction between classical systems and neural-network style systems; for a very useful discussion, see Piccinini (2008; Piccinini and Scarantino (2010)).
[25] See Seung (1998), and Shadmehr and Wise (2005).
[26] For some recent examples see Abbott (2008); Tkacik, Schneidman, Berry, and Bialek (2009); Moazzezi and Dayan (2010); and Rajan, Abbott, and Sompolinsky (2010).

models, which (apparently) is not captured by the receptor notion. It also often plays a methodological role in bridging the work of experimentalists who invoke the receptor notion and the theoreticians who use this information to discover what the system computes and, hence, how the system models its target. Thus, contrary to what Ramsey claims, S-representation is a central theoretical notion in the newer approaches in brain and cognitive sciences.

In the next two sections I consider two objections to this conclusion. The first is that the theory of the oculomotor memory does not posit real *inner* representations. The other is that it does not posit real *representations*.

6. Objection 1: The memory is not really an *internal* model

One might object that the memory does not really constitute an "internal" model of the represented eye. There are two closely related concerns here. One is that the internal states are sorts of "black boxes". Each box is a "collective" state that can be seen as a representation. Yet the "internal" ingredients of the state do not mirror any conditions in the target domain. They represent nothing. Thus the mirroring relation between the memory and the eye is partial: although the theory essentially refers to the ingredients of each box (state) in describing the computational structure of the memory, these ingredients do not mirror any states and conditions in the represented domain. Another concern pertains to the dynamics of the network. The worry here is that the internal dynamics, which is a relaxation on a new fixed point, does not mirror the dynamics of the eye movement, which is a continuous process that proceeds through other positions of the eye. It thus seems that each relaxation is really an input-output process, from one state to another, but the intermediate dynamics does not mirror any feature of the eyes. Thus, by

Ramsey's standards, we do not have a genuine case of an internal model, hence, of S-representation.

Both concerns can be addressed. Regarding the internal structure of the collective state, the point is that the memory is like a map; each point in the map is a collective internal state (collected from the activity of all cells) state that encodes a different eye-position. It is the space of collective states that lie along the line attractor. It is true that in this picture we can look at each collective state as a "black box", but it does not follow that the collective states are not internal; they are the states that constitute the "map". Furthermore, we can keep analyzing the internal information-processing structure of each state. Seung, for example, points out that each cell encodes a preference ("sensitivity") for the eye position (1996:13339). The fixed-point representation of an eye position consists of the representational contents of the ingredient cells that together form a model of this position. Each cell contributes its share to the collective representation, and the relations between the single cell-representations mirror the relations between the "sub-features" that are being represented.

Or take another attractor neural networks, for solving the n-queens problem (the problem is of locating n queens on a n×n chessboard, such that no two queens are on the same row, column or diagonal).[27] The state-space in this network stands for all possible assignments of queens on the board, and the global minima points are solutions to the problem. Each minimum is a collective state that represents one solution to the problem. Further analyzing this collective state, each cell represents the presence/absence of a queen in a square of the board, and the relations between them mirror the "spatial" relations on the board. So here not only does the state-space model the space of possible states of the board; in addition, each state of the network,

[27] See Shagrir (1992).

whether stable or not, models a state of the board in that its constituents and relations mirror ("explicit") features of the board.

As for the dynamics of the network, the reply is twofold. First, the task of the oculomotor system is not to mimic eye movement. Its task is to point to the current eye position in order to maintain a constant gaze of the eye. This memory task requires the mirroring of *some* relation among the eye positions, but not necessarily of eye movements. The memory achieves the task by employing another property of the eye, namely, its velocity: it tracks eye positions by mirroring the relation among them with respect to velocity. Second, it is incidental that some memory networks mirror the eye movement (which is perhaps not too surprising given the relation between velocity and movement). If the network is a linear one, the dynamics of the network takes place along the line attractor alone. In this case, the internal dynamics is a continuous movement along the line attractor from the point that encodes the previous eye position to the point representing the current eye position. Such a network can be calibrated to mirror the eye movement itself.

The two concerns raise more general issues about the amount of mirroring required from S-representations, given that full-scale isomorphism is impractical. Ramsey seems to think that modeling is explanatory only when "there is a significant type of isomorphism" (2007: 80). But what exactly amounts to "significant"? Should we require that every *computational* bit and operation mirror a state or a condition in the target domain? Ramsey is not too explicit about this. He does emphasize that the model should include internal elements, and not just input and output representations. We may recall that this claim plays a crucial role in his argument against connectionism, when he dismisses the explanatory powers of representation in connectionist machines. Ramsey concedes that connectionist machines simulate the target domain, but he

argues that this simulation occurs at input and output end-points and is related to function, not to mechanism. As such, it is part of the "theory-neutral specifications of the explananda" (p. 71) and has nothing to do with the "the internal explanatory posits of particular cognitive theories" (p. 71).

But can we dismiss the explanatory power of these input and output representations when they simulate some relation in the target domain? Let us assume, for the sake of argument, that the oculomotor memory provides only an input-output simulation; let us even grant, for the sake of argument, that the schema in fig. 5 describes an input-output modeling. To simplify a bit, the integration relation, over the incoming pulses, is just a transformation from one state (input) to another (output), and this transformation simulates the external (integration) relation between the represented eye positions (with respect to velocity). Can we say that this schema is just part of "theory-neutral specifications of the explananda"? I think not. I suspect that Ramsey confuses two different *non-mechanistic* input-output specifications. One specification refers to the information-processing, cognitive, task (fig. 2). In our case, the memory task is described as mapping from one state of the neural network, which represents one eye position, to another state, which represents another eye position; this mapping, *g*, is a function of eye-velocity encoding inputs. This specification, in terms of representations of eye velocity and eye positions, is indeed the part of the explanandum.

The other specification, however, refers to the shared formal (mathematical) structure between the satisfied neural function and the functional relation between the represented entities, i.e., between eye positions and eye velocity. Here, the specification refers not only to the representational content of the cells, but also to the shared formal structure of integration (fig. 5). This specification, I believe, is already part of the explanation. Showing that the system satisfies

(or computes) integration, and that this computation simulates the (integration) relation between velocity and positions, helps us understand how the memory system attains the task of keeping the eyes still (another part of the explanation, of course, is the specification of the internal mechanism by which the system computes integration). This contention requires elaboration and argumentation beyond the scope of this paper. But I think that we can at least say that it is far from trivial that simulation (input-output) representation, which refers to the shared formal structure, has no explanatory role.

The distinction between the two types of input-output specifications is often masked, e.g., in Cummins's adding device, when the represented domain and the shared structure are the same (i.e., the *plus* function). But it is highlighted in other descriptions of adding machines, as in Marr's example of the cash-register machine.[28] One specification of the cash-register is in information-processing terms. On this specification, this machine satisfies a mapping function *g* from input digits to output digits; these digits represent, respectively, the prices of purchased items in the store and the final bill. Another non-mechanistic specification refers to the shared structure between two domains, to the (simulating) input-output mapping and the (simulated) activity that these input and output digits represent; the shared structure in this example is the *plus* function. As Marr (1982) puts it, the simulating input-output mapping conforms to the rules of commutativity, associativity, zero, and inverse, which uniquely define *plus*. But so is the simulated activity in the store. It also conforms to the rules of zero, commutativity, associativity, and inverses.[29] Ramsey can say that the first, information-processing, specification is part of the explanandum; that what we want to explain is how the machine attains this information-

[28] Marr 1982: 22-23; I discuss this example in detail elsewhere (Shagrir 2010a).
[29] External associativity, for example amounts to "arranging the goods into two piles and paying for each pile separately should not affect the total amount you pay" (Marr 1982: 23).

processing task. I do not dispute that. My point, again, is that even if this information-processing specification belongs to the explananda, it does not follow that the second, shared-structure, specification is also part of the explananda and not part of the explanation. It might well be that the shared-structure specification explains (at least partly) why the cash-register, which satisfies the function *g*, is suitable for the information-processing task that is defined in terms of the activity in the store.[30]

7. Objection 2: The internal states are not real *representations*

One could object to the claim that the theory posits real *representations*. The objection here is not to the claim that the memory and the eye have a shared inner structure. The objection, rather, is that *all* I have shown is that the memory and the eye have a shared structure; I have not shown that the cells of the memory really have representational power. True, scientists refer to the memory system as a representational system, but, the objection goes, representational talk does not entail representational power. One can infer from shared structure to representational power *if* one assumes that representational power reduces to shared structure ("isomorphism"). But this assumption is highly contentious. As many have pointed out, one can have shared structure without having representation at all.[31]

[30] Ramsey grounds his claim about the inputs and outputs in the standard interpretation of Marr. Elsewhere I argue that this interpretation distorts Marr's conception of computational-level theory (Shagrir 2010a). In particular, I argue that one can hardly rely on Marr when dismissing the explanatory power of the input-output relations, claiming that they are part of the explananda.

[31] See, e.g., Goodman (1968), Frigg (2006), and Suarez (2010). Swoyer writes that it is not claimed that the notion of structural representation "fully captures our ordinary sense of representation - it doesn't, and it's not intended to" (1991: 452). Shared structure, Swoyer continues, is neither sufficient nor necessary for having representational power. It is not necessary since "with sufficient perseverance - or perversity - we can use anything to represent virtually anything else, and in many cases the two things won't have any interesting structural similarities at all"

In replying I want to make two pertinent comments. First, the notion of representation assumed in the memory system is richer than that of S-representation. For one thing, the representing neural states are causally related with the target;[32] as we saw above, these causal relations between the memory and its target underlies the notion of representation that guides the electrophysiological studies. It thus might well be that the content of the representations is partially grounded at least in part, in some sort of causal covariance.[33] Ramsey does not conceive of this option as he seems to assume that the S-representation notion and the receptor notion are mutually exclusive. But, as the memory example indicates, this assumption is dubious: shared structure and causal covariance can live together.[34] For another thing, these representations are *used* by the system to control the movement of the eyes. These representations enable the system to keep track of eye position; for example, the oculomotor system reads out the current eye-position in order to keep the eyes still. Thus taken together, the notion of representation assumed here is that of S-representation that is both used by the system to guide behavior and is also anchored by receptor-style cells. Thus while I cannot offer here a *theory* of representation, it seems to me that the notion of representation assumed in these studies is as strong as any notion of representation-in-a-cognitive-system we could hope for.

Second, the notion of representation assumed in the memory system satisfies Ramsey's requirements for being a representation. Ramsey does not provide a precise set of constraints that

---

(1991: 452). And it is not sufficient since "if you can find one structural representation of something, you can usually find many" (1991: 452).

[32] See Dretske (1988) for elaboration of this idea of co-variation into a forceful account of representation.

[33] This double-factor possibility is, of course, no novelty and quite widespread in the philosophy of mind; see, for example, Fodor (1987, 1990).

[34] See Sprevak (2011) for further criticism and discussion.

fully unpack his "job description challenge".[35] At some points, it seems that Ramsey upholds a reductive view of S-representations according to which shared structure *does entail* representational power and content (2007: 81-83). In particular, Ramsey relies in his characterization on Cummins, who clearly aims to provide a (reductive) theory of *content*; e.g., citing Cummins's statement that "Representation, in this context, is simply a way of talking about an aspect of more or less successful simulation" (2007: 83).[36] On this reductive view, *if* the memory system and the eye have a shared structure, then the memory system *is* a (structural) representational system; put differently, the memory system satisfies the job description challenge by virtue of having a shared structure with the eye. And having demonstrated that the antecedent of conditional is true, the consequent just follows.

At other points, Ramsey (pp. 83-87, 93-96) specifically addresses the claim that representational content is not automatically reduced to shared structure as a potential objection to his view. In this context he invokes the requirement that in addition to having shared structure with the target, the entity is also used to explain how the system solves a problem. Whether this strategy delivers real representations is an issue I put aside.[37] What I want to emphasize is that the memory network, as an internal model, satisfies the explanatory requirement. The fact that the memory network is an internal model of the eye plays a crucial role in explaining how the oculomotor system attains its neuro-cognitive tasks. The brain keeps the eyes still because it

---

[35] Ramsey identifies the "job description challenge" with the constraint that the posited representations has to do an explanatory work *qua* representation (2007; p. 12; see pertinent discussion on S-representation in pp. 81-83); for more general discussion see Frigg and Hartmann (2006).

[36] Cummins also says that "s-representation is simply a consequence of (one might almost say an artifact of) simulation" (Cummins 1989: 97). Discussing his adding machine, Cummins says that it is "simply the fact that *g* is isomorphic to + that makes the arguments and values of g representations of numbers for a system that satisfies *g* (p. 92). … There is no 'further fact' to representation in this case beyond *g*'s instantiating +… representation is just a name for the relation induced by the interpretation mapping between the elements of *g* and the elements of +" (p. 93). More recently, Cummins (1996) abandons this approach at least with respect to content.

[37] See Ramsey (2007: 96-102) for discussion. See also Sprevak (2011) who points out that Ramsey curiously assigns to the notion of S-representation a dual role, one explanatory and another as a theory of content.

stores a memory network which is an internal model of the eye. The stable states in the modeling network encode the space of eye position, whereas each stable state represents a different eye position. When the network is in one of its stable states, it memorizes the current eye position. When the eye changes positions the memory network computes integration over the pulse eye-velocity encoding input. This transition, from one state to another, mirrors the change in eye positions (which is just integration over eye velocity). Thus the oculomotor system, which supplies the transient saccadic inputs, reads out the eye-position encoding state of this internal model, using this information to keep the eyes still in the encoded position.[38]

The upshot, then, is that my argument stands, at least in the context of Ramsey's definitions. Ramsey argues that the newer accounts do not posit S-representations. My claim is that he is wrong about that: the newer accounts typically *do* posit S-representations, at least the way Ramsey understands the notion of S-representation.

8. Summary

My aim in this paper was to demonstrate that the notion of structural representation is a central theoretical notion in the current computational approaches in cognitive neuroscience. The argument rests on one case-study, that of some computational work on the oculomotor integrator. This work explains how the integrator system fixates the eyes by describing it in terms of an internal model. It also employs the (working) assumption of structural representation to

[38] The same goes for the network solving the n-queens problem. The explanation of how the network solves the problem crucially refers to the fact that its space-state is a model of all possible assignments of queens on the board, and each global-minimum in this state-space represents a possible solution to the problem. The network solves the queens problem – the explanation goes – due to the fact that its dynamics (at each trial) is a gradual relaxation on one a global minimum of this state-space model.

conclude, on the basis of external cues, that the system is an "integrator"; thus the notion of structural representation also plays a methodological role in the discovery of what the system computes. While my focus here was one case-study, I pointed out that it is not an isolated example. That the nervous system is an internal model of the world is a widespread assumption in current computational approaches in cognitive neuroscience.

Acknowledgment: I am grateful to Yonatan Loewenstein who helped me understanding the computational principles underlying the oculomotor system. Thanks to Bill Bechtel, Eli Dresner, Frances Egan, Paul Humphries, Hilla Jacobson, Arnon Levy, Bob Matthews, Etye Steinberg, Eran Tal, and three anonymous referees of this *Journal*, for their comments, suggestions and corrections. This research was supported by the Israel Science Foundation, grant 725/08.

Oron Shagrir

Departments of Philosophy and Cognitive Science

The Hebrew University of Jerusalem

References:

Abbott, L.F. 2008: Theoretical neuroscience rising. *Neuron* 60:489-495.

Amit, D.J. 1989: *Modelling Brain Function*. Cambridge University Press.

Bartels, A. 2006: Defending the structural concept of representation. *Theoria* 55: 7–19.

Becker, W. & Klein, H.M. 1973: Accuracy of saccadic eye movements and maintenance of eccentric eye positions in the dark. *Vision Research* 13: 1021–1034.

Cannon, S.C. & Robinson, D.A. 1985: An improved neural-network model for the neural integrator of the oculomotor system: More realistic neuron behavior. *Biological Cybernetics* 53: 93-108.

Churchland, P.M. 2007: *Neurophilosophy at Work*. Cambridge University Press.

Cummins, R. 1989: *Meaning and Mental Representation*. MIT Press.
-- 1996: *Representations, Targets, and Attitudes*. MIT Press.

Da Costa, N.C.A. & French, S. 2003: *Science and partial truth: A unitary understanding of models and scientific reasoning*. Oxford University Press.

Dretske, F. 1988: *Explaining Behavior*. MIT Press.

Edelman, S. 1998: Representation is representation of similarities, *Behavioral and Brain Sciences* 21: 449-498.

Eliasmith, C. & Anderson C.H. 2003: *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*. MIT Press.

Fodor, J.A. 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
-- 1990: *A Theory of Content and Other Essays*. MIT Press.

Fodor, J.A. & Pylyshyn, Z.W. 1988: Connectionism and cognitive architecture: A critical analysis. *Cognition* 28: 3-71

French, S. & Ladyman, J. 1999: Reinflating the semantic approach. *International Studies in Philosophy of Science* 13: 99-117.

Frigg, R. 2006: Scientific representation and the semantic view of theories. *Theoria* 55: 49-65.

Frigg, R. & Hartmann, S. 2006: Models in science. In E.N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/spr2006/entries/models-science/>.

Garzón F.C. & Rodríguez Á. G. 2009: Where is cognitive science heading? *Minds and Machines* 19: 301-318.

Giere, R.N. 2004: How models are used to represent reality. *Philosophy of Science* 71: 742–752.

Glimcher, P.W. 1999: Oculomotor control. In Wilson and Kiel (eds.) *MIT Encyclopedia of Cognitive Science* (pp. 618-620). MIT Press.

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.

Goodman, N. 1968: *Languages of Art: An Approach to a Theory of Symbols*. Bobbs-Merrill.

Grush, R. 2004: The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences* 27: 377-442.
-- 2008: Review of William M. Ramsey, *Representation Reconsidered*. *Notre Dame Philosophical Reviews*.

Hess, R.F., Baker Jr., C.L., Verhoeve, J.N., Keesey, U.T. & France, T.D. 1985: The pattern evoked electroretinogram: its variability in normals and its relationship to amblyopia. *Investigative Ophthalmology & Visual Science* 26: 1610-1623.

Hopfield, J.J. 1982: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA* 79: 2554-2558.

Hopfield J.J. & Tank D.W. 1985: "Neural" computation of decisions in optimization problems. *Biological Cybernetics* 52: 141-152.

Johnson-Laird, P.N. 1983: *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*. Harvard University Press.

Leigh, R.J. & Zee, D.S. 2006: *The Neurology of Eye Movements* (4th edition). Oxford University Press.
Marr, D. 1982: *Vision*. Freeman.
Mensh, B. D., Aksay, E., Lee, D. D., Seung H. S. & Tank D. W. 2004: Spontaneous eye movements in goldfish: oculomotor integrator performance, plasticity, and dependence on visual feedback. *Vision Research*, 44: 711-726.
Moazzezi, R. & Dayan P. 2010: Change-based inference in attractor nets: Linear analysis. *Neural Computation* 22: 3036-3061.
Moschovakis, A.K. 1997: The neural integrators of the mammalian saccadic system. *Frontiers in Bioscience* 2: 552-577.
O'Brien, G. & Opie, J. 2001: Connectionist vehicles, structural resemblance, and the phenomenal mind. *Communication and Cognition* 34: 13-38.
-- How do connectionist networks compute? *Cognitive Processing* 7: 30-41.
-- The role of representation in computation. *Cognitive Processing* 10: 53-62.
Palmer, S.E. 1978: Fundamental aspects of cognitive representation. In E. Rosch and B. B. Lloyd (eds.), *Cognition and Categorization* (pp. 259-303). Erlbaum.
Quartz, S.R. 2008: From cognitive science to cognitive neuroscience to neuroeconomics. *Economics and Philosophy* 24: 459-471.
Piccinini, G. 2008: Some neural networks compute, others don't. *Neural Networks* 21: 311-321.
Piccinini, G. & Scarantino, A. 2010: Computation vs. information processing: How they are different and why it matters. *Studies in History and Philosophy of Science*, forthcoming.
Rajan, K., Abbott, L.F. & Sompolinsky, H. 2010: Stimulus-dependent suppression of chaos in recurrent neural networks. *Phys. Rev. E* 82: 01193.
Ramsey, W. 2007: *Representation Reconsidered*. Cambridge University Press.
Robinson, D.A. 1989: Integrating with Neurons, *Annual Review of Neuroscience* 12: 33-45.
Rumelhart, D.E., Smolensky, P. McClelland, J.L. & Hinton, G.E. 1986: Schemata and sequential thought processes in PDP models. In McClelland, J.L., Rumelhart, D.E., and the PDP group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models* (pp. 7-57). MIT Press.
Ryder, D. 2004: *SINBAD* Neurosemantics: a theory of mental representation. *Mind & Language* 19: 211-240.
Seung, S.H. 1996: How the brain keeps the eyes still. *Proceedings of the National Academy of Science USA* 93: 13339-13344.
-- 1998: Continuous attractors and oculomotor control. *Neural Networks* 11: 1253-1258.
Seung S.H., Lee, D.D., Reis, B.Y. & Tank, D.W. 2000: Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26: 259-271.
Shadmehr, R. & Wise, S.P. 2005: *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. MIT Press.
Shagrir, O. 1992: A Neural Net with self-inhibiting units for the n-queens problem. *International Journal of Neural Systems*, 3: 249-252.
-- 2010a: Marr on computational-level theories. *Philosophy of Science* 77: 477-500.
-- 2010b: Brains as analog-model computers. *Studies in the History and Philosophy of Science* 41: 271–279.
Shepard, R.N. & Chipman, S. 1970: Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology* 1: 1-17.
Sprevak, M. 2011: Review of W. M. Ramsey, *Representation Reconsidered*. *British Journal for the Philosophy of Science*, forthcoming.
Swoyer, C. 1991: Structural representation and surrogative reasoning. *Synthese* 87: 449-508.
Suárez, M. 2010: Scientific representation. *Philosophy Compass* 5: 91-101.
Tkacik, G., Schneidman, E., Berry, M.J. & Bialek, W. 2009: Spin glass models for a network of real neurons. *arXiv:0912.5409 [q–bio.NC]*.