Representations: Who needs them? *In: Third Conference, Brain Organization and Memory: Cells, Systems and Circuits. (Eds.) J.L. McGaugh JL, Weinberger NM, Lynch G (eds.). New York, Oxford, Guilford Press.. pp. 375-380. Freeman WJ, Skarda, C.A. (1990)*

Representations: Who Needs Them?

WALTER J. FREEMAN

CHRISTINE A. SKARDA

From: Brain Organization and Memory Cells, Systems, & Circuits

Edited by: James L. McGaugh, Norman Weinberger, Gary Lynch, page: 375–380

Biologists by tradition have seldom used the term *representation* to describe their findings. Instead they have relied on phrases such as "receptor field" on the sensory side and "command" or "corollary discharge" on the motor side when discussing neural control of sensation and motion in goal–directed behavior. Such words connote dynamic process rather than symbolic content. One might suppose that this neglect of a now common word reflects diffidence about discussing so–called higher functions of the brain, owing to a humbling lack of understanding of the brain's complexity. Inspection of biology textbooks belies this view. Biologists have shown no lack of hubris in pontificating about the properties of the brain supporting mental functions. On the contrary, they have always taken pride in being uniquely qualified to explain brain function to anyone willing to listen.

A turning point came in the 1940s with the popularization of digital and analog computers as "giant brains" and with the adoption of the Turing machine as a model for explanation. In this conception, which is lucidly illustrated by von der Malsburg (Chapter 18, this volume), the human ability to understand the world is likened to the procedure of incorporating information on a tape into a machine by means of symbols. Cognitive

operations are interpreted as the manipulation of these symbols according to certain semantic rules. At present we are all so accustomed to this metaphor that it seems self–evident (Goldman Rakic, 1987). The brain's job is to incorporate features of the outside world and make internal syntactical representations of these data, which together constitute a world model that serves to control motor output. Any other account appears to be "noncognitive" (Earle, 1987) and counterintuitive. In short, to question this commonsense notion seems quixotic, sophistic, and arbitrary.

We propose, however, that physiologists avoid this way of thinking for two reasons. One is that no one now understands how brains work, but the use of the term representation and its attached concepts tends to obscure this fact. The term gives us the illusion that we understand something that we do not. We suggest that the idea of representation is seductive and ennervating, promising good deals but delivering nothing new. When researchers refrain from using the term, knowledge of brain function is not significantly affected. We conclude that use of the term is unnecessary to describe brain dynamics.

The second reason is that the use of the metaphor points us in a direction that carries physiological research away from more profitable lines of inquiry. We have found that thinking of brain function in terms of representation seriously impedes progress toward genuine understanding.

An example is taken from our studies of the behavioral correlates of the electroencephalograms (EEGs) of the olfactory system under conditioning (Freeman & Skarda, 1985). The EEGs of the olfactory bulb and cortex show a brief oscillatory burst of potential that accompanies each inhalation. This can be likened to a burst of energy carried by a wave of neural activity at a common frequency. Each burst exists over the entire bulb or cortex with a spatial pattern of amplitude that varies from one burst to the next. We have shown that a stereotypical pattern recurs whenever a particular odorant is presented that the animal has been trained to respond to.

For more than 10 years we tried to say that each spatial pattern was like a snapshot, that each burst served to represent the odorant with which we correlated it, and that the pattern was like a search image that served to symbolize the presence or absence of the odorant that the system was looking for. But such interpretations were misleading. They encouraged us to view neural activity as a function of the features and causal impact of

stimuli on the organism and to look for a reflection of the environment within by correlating features of the stimuli with neural activity. This was a mistake. After years of sifting through our data, we identified the problem: it was the concept of representation.

Our research has now revealed the flaws in such interpretations of brain function. Neural activity patterns in the olfactory bulb cannot be equated with internal representations of particular odorants to the brain for several reasons. First, simply presenting an odorant to the system does not lead to any odor–specific activity patterns being formed. Only in motivated animals, that is, only when the odorant is reinforced leading to a behavioral change, do these stereotypical patterns of neural activity take shape. Second, odor–specific activity patterns are dependent on the behavioral response; when we change the reinforcement contingency of a CS we change the patterned activity. Third, patterned neural activity is context dependent: the introduction of a new reinforced odorant to the animal's repertoire leads to changes in the patterns associated with all previously learned odorants. Taken together these facts teach us that we who have looked at activity patterns as internal representations of events have misinterpreted the data. Our findings indicate that patterned neural activity correlates best with reliable forms of interaction in a context that is behaviorally and environmentally co–defined by what Steven Rose (1976) calls a dialectic. There is nothing intrinsically representational about this dynamic process until the observer intrudes. It is the experimenter who infers what the observed activity patterns represent to or in a subject, in order to explain his results to himself (Werner, 1988a, 1988b).

The impact of this insight on our research has been significant. Once we stopped looking at neural activity patterns as representations of odorants, we began to ask a new set of questions. Instead of focusing on pattern invariance and storage capacity, we began to ask how these patterns could be generated in the first place from less ordered initial conditions. What are the temporal dynamics of their development and evolution? What are their effects on the neurons to which they transmit? What kinds of structural changes in brains do they require and do they lead to? What neuromodulators do these pattern changes require? What principles of neural operations do these dynamical processes exemplify and instantiate? In short, we began to focus less on the outside world that is being put into the brain and more on what brains are doing.

Our efforts to answer these questions led us to develop mathematical, statistical, and electronic models that describe and explain the neural dynamics of pattern generation. These models have, in turn, caused radical changes in our views of how brains operate. In particular, we now see brains as physicochemical systems that largely organize themselves, rather than reacting to and determined by input. As Carew and co–workers (Chapter 2, this volume) showed, each brain has a history that begins with simple structures and that evolves through innumerable stages and phases of growth and development to increasing order and complexity. The patterns are formed from within and not imposed from outside, as is commonly supposed to occur in brains under sensory stimulation. We have found that an essential condition for these patterns to appear is the prior existence of unpatterned energy distributions which appear to be noise, but which in reality are chaos. New forms of order require that old forms of order collapse back into this chaotic state before they can appear. Therefore, in the EEG we see each burst appearing from chaotic basal activity and collapsing back into chaos, thereby clearing the way for the next burst of patterned activity (Skarda & Freeman, 1987).

These findings challenge two widely held assumptions concerning brain dynamics. First, conventional theory holds that full information is delivered into the system and that thereafter it is degraded by noise. This property is analogized to entropy. However, chaotic systems like the brain are open and, by virtue of energy throughput, operate far from equilibrium. They internally create new information and can be described as negentropic (Tsuda & Shimizu, 1985). The brain has immunity from the first and second laws of thermodynamics because its assured blood supply brings it more energy than it can use and carries off waste heat and entropy. As a result the formalisms of information theory that underlie the representation–based computational metaphor of brain dynamics do not apply to the neural networks of biological systems, because these formalisms make no sense in systems with positive information flow.

Second, the conventional description of signals embedded in noise is inappropriate. The same neural system that generates bursts (signals) also generates the background state of chaotic activity (often thought to be noise). When the system switches (bifurcates) from chaos to burst activity, the chaotic activity stops and the signal starts. Chaos operates up to the moment of bifurcation. It plays no role of "annealing" thereafter because response selection has already taken place, convergence is assured, and there is no

role for "noise." The metaphor of the "signal to noise ratio" is inappropriate for brain function, yet it is essential for representation in man–made systems.

These considerations are well illustrated by the preceding four chapters. Of these the report by Cooper, Bear, Ebner, and Scofield (Chapter 15) deals most directly with neurophysiological data, and they alone make no reference to representations. Their model describes the dynamics of modifiable synapses over the time scale of learning. It does not address the dynamic of stimulus–induced neural activity on the time scale of responding. It attempts to explain the dynamics in terms of membrane conductances and calcium fluxes and not the semantic content of the input. The strength of their model lies in the identification of global variables as important for consideration; the main weakness is their decision to "delay consideration of the global variables by assuming that they act to render cortical synapses modifiable or nonmodifiable by experience."

Their chapter emphasizes the great value of the "mean field" state variable for the analysis and understanding of physical systems composed of ensembles. They elect not to consider the basis for defining and using such variables in their work with neural networks, thereby depriving themselves and their readers of access to the existing literature, including well–reasoned approaches to brain systems from the classical standpoints of statistical mechanics (Amari, 1974; Wilson & Cowan, 1972) and of nonequilibrium thermodynamics from the school of Prigogine (Babloyantz & Kaczmarek, 1981; Freeman, 1975). Both approaches emphasize the importance of mutually excitatory (positive) feedback within laminar distributions of nerve cells by their recurrent collateral axons and the renewal process in cell firing, leading to the emergence of macroscopic state variables to represent the activity of local neighborhoods, that is, axonal pulse density and dendritic current density functions that are continuous in both time and the spatial dimensions of cerebral cortex. These activity states are readily observed in many parts of brains by use of electrodes, magnetic probes, and optical dyes and by computer–implemented spatial filtering, summing, and enhancement of the raw data for visual display (Freeman, 1987). By means of these and related well–documented procedures these investigators and others can test their models directly with respect to brain dynamics.

Chapter 18 by von der Malsburg and part of Chapter 17 by Sejnowski and Tesauro are at the opposite pole and take representation for granted. They are also explicitly about machines and not about brains. "Representations"

that are selected and defined by the observer serve as the goals or end points of the evolution of the machines. Sejnowski and Tesauro present elements on both sides. The description of Gerald Westheimer's dynamics of spatial attractors, which in some ways recalls Wolfgang Kohler's field theories, is accounted for by the dynamics of mutually excitatory feedback. This is physiology (Freeman, 1975). When he asks "Who reads the population code?" he gives engineering answers: "find all possible depths and find which matches the closest with minimal error." The algorithms of back propagation and error correction by the observer instilled "teacher" and "correct answer" are machine processes that do not exist in biological brains. "NETtalk" can transduce optical characters to sounds that are recognizable by human observers, so it is a machine that can be shaped to read to the blind, but one cannot say that the machine has learned to read in the sense that a schoolchild has.

Kohonen (Chapter 16) most clearly addresses the nature of internal representations as they are needed and used by engineers and machines. Each representation has characteristics and attributes that are to be stored, matched, and retrieved by processes ultimately deriving from mappings. Our physiological data show that episodic storage of odor trials does not happen, that "retrieval" is not recovery but re–creation, always with differences, and that stimulus–bound patterns cannot coexist with re–created patterns to support matching procedures. We agree with Kohonen's statement that we are faced with semantic difficulties, and we conclude that they stem from deep incompatibilities between the dynamics respectively of biological and present–day artificial intelligence. The key words to look for are "best matching" and "error detection," because these refer to machine cognition and not neural cognition.

These considerations give an answer to our question about representations. Who needs them? Functionalist philosophers, computer scientists, and cognitive psychologists need them, often desperately, but physiologists do not, and those who wish to find and use biological brain algorithms should also avoid them. They are unnecessary for describing and understanding brain dynamics. They mislead by contributing the illusion that they add anything significant to our understanding of the brain. They impede further advances toward our goal of understanding brain function, because they deflect us from the hard problems of determining what neurons do and seduce us into concentrating instead on the relatively easy problems of determining what our computers can or might do. In a word, representations

are better left outside the laboratory when physiologists attempt to study the brain. Physiologists should welcome the ideas, concepts, and technologies brought to them by brain theorists and connectionists, but they should be aware that representation is like a dose of lithium chloride; it tastes good going down but it doesn't digest very well (Bureg, Chapter 1, this volume).

REFERENCES

Amari, S. (1974). A method of statistical neurodynamics. *Kybernetik* 14, 201–215.

Amari, S. (1987). *Statistical neurodynamics of associative memory.* (Tech. Rep. METR87–8). Tokyo: University of Tokyo, Department of Mathematics, Engineering, and Instrument Physics.

Babloyantz, A., & Kaczmarek, L. (198 1). Self–organization in biological systems with multiple cellular contacts. *Bulletin of Mathematical Biology,* 41, 193–20 1.

Earle, D. C. (1987). On the differences between cognitive and noncognitive systems.

Brain and Behavioral Science, 10, 177–178.

Freeman, W. J. (1975). *Mass action in the nervous system.* New York: Academic Press.

Freeman, W. J. (1987). Analytic techniques used in the search for the physiological basis of the EEG. In A. S. Gevins & A. Rémond (Eds.), *EEG Handbook* (pp. 583–664). Amsterdam: Elsevier.

Freeman, W. J., & Skarda, C. A. (198 5). Spatial EEG patterns, nonlinear dynamics and perception: The neo–Sherringtonian view. *Brain Research,* 10, 147–175.

Goldman–Rakic, P. (1987). Circuitry of primate prefrontal cortex and the regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of physiology: Sec. 1. The nervous system* (pp. 373–417). Bethesda Md.: American Physiological Society.

Gray, C. M., Freeman, W. J., & Skinner, J. E. (1986). Chemical dependencies of learning in the rabbit olfactory bulb: Acquisition of the transient spatial pattern change depends on norepinephrine. *Behavioral Neuroscience,* 100, 585–596.

Rose, S. P. R. (1976). *The conscious brain.* New York: Vintage Books.

Skarda, C. A., & Freeman, W. J. (1987). Brain makes chaos to make sense of the world. *Brain and Behavioral Science,* 10, 161–195.

Tsuda, I., & Shimizu, H. (1985). Self–organization of the dynamical channel. In H. Haken (Ed.), *Complex systems: Operational approaches in neurobiology, physics and computers* (pp. 240–251). Berlin: Springer–Verlag.

Wemer, G. (I 988a). Five decades on the path to naturalizing epistemology. In J. S. Lund (Ed.), *Sensory processing in the mammalian brain* (pp. 345–359). New York: Oxford University Press.

Werner, G. (1988b). The many faces of neuroreductionism. In E. Basar (Ed.), *Dynamics of sensory and cognitive processing by the brain* (pp. 241–257). Berlin: Springer–Verlag.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in biological populations of model neurons. *Biophysics Journal,* 12, 1–24.

END