

Siren Call of Metaphor: subverting the proper task of System Neuroscience.

Gerhard Werner

Department of Biomedical Engineering

University of Texas

1 University Station C0800

Austin TX 78712-0238

e-mail: gwer1@mail.utexas.edu

Abstract:

Under the assumption that nervous systems form a distinct category among the objects in Nature, applying metaphors of psychological and behavioral science disciplines is flawed and invites confusion. Moreover, such practices obscure and detract from the primary task of Neurophysiology: to investigate the intrinsic properties of nervous systems, uncontaminated with concepts borrowed from other disciplines. A comprehensive fundamental theory of nervous systems is expected to have the character of high dimensional nonlinear systems in which state space transitions, set in motion by external influences, self-organize to dynamic state space configuration with consequences for behavior.

Key words: attractors, brain state space, cognitive neurophysiology,
conditioned motor behavior, naturalism, metaphor.

Running title: Siren Call of Metaphor

Introduction

The rise of Naturalism in Philosophy since the 1980s shifted its traditional task of establishing and questioning the foundational assumptions of the Sciences largely to the disciplines themselves. For the Philosopher N. W. Quine [29] Naturalism is “the recognition that it is within Science itself , and not in some prior Philosophy, that reality is to be identified and described”. This charter requires each empirical discipline to define explicitly to which objects in Nature its concepts refer: in the terminology inherited from pre-naturalistic Philosophy, what its “natural kind terms” designate in distinction from other discipline’s classes of objects, at least in practices of inquiry [5]. Based on insights gained in experimental investigations [42], Freeman [10] emphasized the distinctiveness of the “tissue formed by neurons in animal brains“ . Accordingly, he rejected the Machine Metaphor of brain function (and the various “crypto-Cartesian” forms it had taken since Descartes) as a “category error” in Ryle’s [32] definition: to construe objects of inquiry that belong to one category in terms of concepts of another. For the human brain, Edelman [9] affirmed emphatically its special status in the world, “like nothing Science has yet encountered”. Though not stated by these authors in these terms, their opinions accord with granting nervous systems the status of a ‘natural kind’ in the material world.

Disregard of this distinctiveness is prone to surface as conceptual confusion in the discourse practices of Neuroscientists when terms such as ‘coding’, ‘representation’, ‘information’ etc. are applied across category boundaries, and without consistent and uniformly adhered-to semantics: sometimes rooted in the one or the other category, or

some ill-defined hybrid [45,46,47]. Bennett and Hacker [4] have recently conducted a thoughtful and penetrating analysis of the conceptual entanglements of this kind. In this essay, I will focus on problems that arise from the ascription of mental and psychological attributes to neurophysiological observations in Cognitive Neuroscience, often under the guise of metaphorical language. I contend that such practices obscure and derail the proper task of Neuroscience to elucidate neural systems as a distinct category of physical systems, within its own conceptual and theoretical boundaries.

Ubiquity, assets and liabilities of Metaphors

Metaphors used in ordinary discourse reflect the models and theories by which we interpret and express thoughts and actions: “our ordinary conceptual system in which we think and act, is fundamentally metaphorical” [21]. In Science, metaphors have traditionally exerted considerable influence on the formation of scientific concepts and theories [15] and supplied evocative terms for their formulation [26]. They are often credited with triggering innovative thought or useful heuristics for experimental design. However, we also encounter the admonition “the price of metaphor is eternal vigilance” in a significant paper of which one of the authors (N. Wiener) was himself the creator of powerful metaphors [31]. Why the warning ?

The attractiveness of metaphors rests on the user’s familiarity with the source domain, and on providing convenient locutions for application in an as yet ill defined target domain [26]. Expedient as it is, this practice tends to obscure potential sources of confusion since concepts and reasoning pattern established for one domain (the source)

are applied to another domain (the target) under the tacit assumption that they transfer with at least approximate validity. Even if used only as shorthand and, perhaps, in a colloquial sense, we tend to fall into the pattern of thought in the more familiar source domain. For the experimenter, there lurks also the specter of circularity: experiments designed in the framework of the source domain tend to yield data for self affirming interpretation in the source's vocabulary. Think of the numerous terms Neurophysiologists import from both Scientific and Folk Psychology when designing experiments and interpreting observations in terms of 'deciding', 'choosing', 'selecting', etc. Once committed to an autonomous 'natural kind' status of Neural Systems, such crossing of categories is on first principles not only flawed but also detracts from investigating the functions and properties that are intrinsic to the neural domain in its own right.

The case of Neuro-Psychological vs. Naturalistic Neuroscience.

For grounding the argument, let us look at the case of 'deciding to' [34] in studies of conditioned motor behavior in monkeys, on which there is a rich harvest of imaginative experimental work on scholarly reviews available. I write this in profound respect for the investigators who conduct this work with immense ingenuity and sophistication. However, I question the soundness of the conceptual framework on which such experiments are predicated, observations are interpreted, and conclusions are formulated. I contend that current practices tend to disregard genuine issues in Neurophysiology with its own definitions of what legitimate propositions and criteria of valid statements in this discipline are.

Here is the typical experimental protocol: the experimenter uses some measure of neural activity of his/her choice (usual neural spike discharges), recorded from a neural structure (selected by him/her on some criterion, and determines relations to behavior that he/she created as link between two events: an antecedent stimulus (chosen by him/her) and a consequent, arbitrary behavior, induced by the training protocol [49]. So far, the experimenter has done all the ‘deciding’, except leaving it up to the monkey to assign a “value” to complying with the experimental protocol.

Different investigators summarize their experimental objective in various ways (in the interest of brevity, I slightly paraphrase, though being careful to preserving the original sense): to characterize neural computations representing the formation of perceptual decision [12]; to investigate the neural basis of a decision process [37]; to examine the coupling of neural processes of stimulus selection with response preparation [34], reflecting connections between motor system and cognitive processes [38] ; to assess neural activity indicating probabilistic reward anticipation [22,27]. In Shadlen and Newsome’s [37] evocative analogy “it is a jury’s deliberation in which sensory signals are the evidence in open court, and motor signals the jury’s verdict”.

Helpful as metaphors and analogies can be as interim steps for making sense of the observation in familiar terms, they also import the conceptual burden of their source domain and lead us to attribute to the animal a decision and choice making capacity along principles for which Psychology has developed evidential and conceptual accounts in

humans under entirely different conditions, and based on different observational facts. Nevertheless, armed with the metaphors of choice and decision, we assert that the observed neural activity is a “correlate” [19] of a decision to emit the observed behavior. As the preceding citations indicate, the observed neural activity is variously attributed to perceptual discrimination between competing (or conflicting) stimuli, to motor planning, or to reward anticipation; the implication being that the neural activity stands for (“represents”) one or the other of these psychological categories.

On compliance

Having introduced compliance in the foregoing in its customary metaphorical use, it behooves me now to deconstruct it in the spirit of this essay. A series of astute experimental observation, stepwise refined since the mid 1980s, established some aspects of phasic neural activity in a class of dopaminergic neurons in the ventral tegmental region of monkeys, that is responsive to the sequence of experimental contingencies and the consistency of external events over time (for review: see [35]). The history of this work also reflects the entanglement of this basic neurophysiologic observation with an abundance of Behavioral Science and anthropomorphic vocabulary, itself couched in vaguely bounded categories like : “reward expectancy”, “incentive salience” , [23], “valuation” [24] and alike. These terms do not have intrinsic referents in the neural domain; rather they are descriptors of the investigator’s interpretation of the experimental observations, drawn from his/her own repertoire of psychological and behavioristic theories. Nor is it appropriate to impose the Temporal – Difference algorithm from Reinforcement Theory [44] as an explanatory model, elegant and imaginative as this may

seem. Dayan and Ballein [7] consider this a blinder to considering richer and faithful models, at the psychological and neural level. As in the case of “decision”, applying vocabulary from psychology and economics obscures the real question of intrinsic neural processes and mechanisms, under the veil of metaphors: characterizing the intrinsic neural processes of integrative functions in the system as a whole.

What is missing?

Did we not in the process “throw the brain out with the proverbial bathwater”? The experimenter set out with the intent to study the brain through subjecting it to external manipulations so that the brain’s response can be observed, but then accounts for the brain’s response in terms borrowed from Psychology, or colloquial language practice. In consequence, we become confined to interpret neurophysiologic observations in the terms of the metaphor’s source domain. I suggest that this is a fallacy, perhaps in a premature rush to reductionism. Instead, I submit, we should apply concepts and tools that are intrinsic to the brain’s own natural kind, i.e.: think and work at the level of genuine neuronal processes and mechanisms. Admittedly, these are the “hard questions” of Neurophysiology as an autonomous project of naturalistic inquiry, with conceptual foundations and methodological principles that are strictly commensurate to the brain’s own physical nature.

Taking System Neuroscience serious.

A brief statement on Theoretical Neuroscience for setting the stage: since the inception of its mathematical orientation in the 1940’s, it has applied analytical

mathematical procedures, and in addition since the 1950s also computer simulations, to abstract mathematical objects, for mapping some sets of neuron-like properties and processes to observable neurophysiologic phenomena; conversely, also to trace a neurophysiologic phenomenon to a minimal set of assumptions for replicating it computationally. It shares this methodological stance with the standard procedures of the theoretical physical sciences. In 1988, Sejnowski et al. [36] introduced “Computational Neuroscience” with the explicit agenda of “explaining how electrical and chemical signals are used in the brain to represent and process information”. It is now becoming apparent that this program fails to effectively discriminate computational information processing from any other form of complex causal processes [13], but it supported and encouraged extant practices of applying informational Metaphors from human Psychology and Information Sciences. Nevertheless, the momentum of Theoretical Neuroscience persists, with new perspectives taking shape [28].

Taking System Neuroscience serious requires realizing that Nervous Systems are at a more global level of analysis one of the classes of physical systems with collective properties that result from the interactions among numerous constituent components. This is the stuff of Physics. In case the components are neurons, the term NeuroPhysics may be appropriate to reflect this outlook. The relevant conceptual frameworks provided by theoretical Physics are Statistical Mechanics, and non-linear Dynamics: the former examining density distributions of elements and their evolution in phase space [3,6,41]; the latter, tracking the motion of dynamical systems along trajectories, and the dynamics of attractors, in phase space. Considerable support has accrued in recent years for

Freeman's [11] interpretation of neurophysiologic phenomena in terms of the latter [1,14,18].

Computational simulations of attractor networks are taken to correspond to formations in physical reality. Under certain circumstances, they are accessible to direct experimental or naturalistic observation (e.g. chemical reaction-diffusion systems). Attractor networks for Neural System have been advocated by Amit [2] more than a decade ago, and since then successfully applied to interpretation of experimental observations and theoretical analyses: path integration for spatial behavior is one of several notable examples [43].

Conditioned motor behavior in a NeuroPhysics perspective.

Here is an illustration of a NeuroPhysics-type view on "decision making" in conditioned motor behavior. The experimental animal coming to the laboratory has been exposed in its life to many external stimuli and had developed certain response patterns. In the framework of nonlinear dynamics, we can view its brain as a high-dimensional state space in which external events have engendered a repertoire of attractors with intersecting basins, bifurcations points, singularities, etc. as triggers for the organism's responses. At this point, the animal enters the training sessions. We then say in the neuro-psychological jargon that the animal "learns" in a given experimental protocol to "decide" which stimulus from among a set is worth its effort to respond to, and which not. In the "brain language" of NeuroPhysics, we take the high dimensional state space of the brain as point of departure. We then say that plasticity of neural tissue enables the

experimenter in the training protocol to shape an attractor configuration in brain state space, with energy minima for behavior that complies with his/her agenda. Pictorially speaking: the experimenter ‘sculpts’ the landscape of troughs and valleys in the brain state space in accord with the actions he/she wanted the animal to make: to get the ball rolling downhill to fall into the “right” slot. This pictorial language must not detract from the physical situation they depict: consider, for instance, the Hopfield [17] networks that store information at local minima of energy functions; the distribution of local minima of energy profiles in analog networks [20]; or the many ways of “sculpting “ attractors in cellular automata [49].

Taking this stance motivates investigating the neuronal mechanisms which create, form and change brain state spaces on the brain’s own terms, unencumbered by metaphors which entice to complacency with terms borrowed from Psychology. In the scenario I sketched before, there is no room for the monkey to make decisions. What is erroneously conjectured to be the animal’s decision is in reality a particular configuration of the brain state space that the experimenter induced by the training protocol.

Conclusion

The case of conditioned motor behavior was intended to serve as an illustrative example. However, the thrust of my argument is more general. There are two ways of viewing the brain’s activity in interactive commerce with its environment. One, as an observer of transactions in the brain environment system: here, the observer looking the at the organism-environment interactions from the outside, as it were, to apply his/her

own linguistic and descriptive categories drawn from the repertoire of the Behavioral Sciences and Psychology. The other way is in terms of the brain's working in the form of neuronal transactions within its own domain: from the inside of the brain, as it were. In the former, one may speak of "decisions", "stimulus salience" or to use locutions like "...how brains think..." [16]. However, these are locutions in the observer's vocabulary, and not intrinsic to the brain's activity. Neglecting those distinctions reflects conceptual confusion.

The stuff of the brain is neurons, interconnected in networks of complex configuration, in dynamic interaction with perturbations reaching them via the senses. The terminology of the observer's Cogitation obscures the proper task of Neurophysiology which is the study of the brain's inner working in terms of the processes that must be described at a level intrinsic to its own functioning. A genuine brain theory must eschew contamination with terms and concepts from other categories, and the observer's external view of the organism – environment transactions. In the Physical Sciences, an analogous distinction of contrasting system characterizations "from the inside" and "from the outside" is gaining currency as the duality of Endo - and Exophysics [30].

Taking this route requires, on the theoretical side, to discover how experiences with low dimensional dynamic systems scales up to systems of high dimensions, probably in rapid switching modes and with multiple control parameters [8]; how macrostates of complex systems can be characterized by coarse grained description of microstates; how

different levels of description of the same system can coexist; how measurement partitions relate to different causal states [39, 40], and many more. The tools of Theoretical Physics hold some important concepts for addressing these questions in brain theory. Much work is also required on experimental methodology: what are appropriate measures of neuronal activity? Is the traditional reliance on sampling neuronal spike discharges, individually or in clusters, adequate?

Having established accounts of the brain's internal processes, we may then be able to say that such-and-such configuration in brain state space can be interpreted as corresponding to certain categories in the vocabulary of behavior or, perhaps, even introspection. But applying the Psychologist's concepts to neurophysiological observation without the intermediate step of a genuine, mechanistic Neurophysiology is an unwarranted shortcut, under the aegis of metaphor. Metaphors detract from situating Nervous Systems squarely in the "Natural Kinds" of basic Physics and its investigative practices.

References

- [1] Abarbanel H.D and Rabinovich M.I., *Neurodynamics: nonlinear dynamics and neurobiology*. Curr.Opin.Neurobiol. **11** (2001) pp.423-430.
- [2] Amit D.J., *Modeling brain function : the world of attractor neural networks*. (Cambridge University Press, 1989)
- [3] Bazso F., Szalisznoy K., Payrits S. and Erdi P., *A statistical approach to neural population dynamics: theory, algorithms, simulations*, Neural Computing **26-27** (1999) pp. 329-334.
- [4] Bennett M.R. and Hacker P.M.S., *Philosophical Foundations of Neuroscience*, (Blackwell, Oxford, 2003).
- [5] Churchland P.M., *Conceptual progress and Word/World relations: in search of the essence of Natural Kinds*. Canad. J. Philos. **15** (1985) pp.1-17.
- [6] Cowan J.D., *Statistical mechanics of nervous nets*. In: Neural Networks: Procc. of

- the School on Neural Networks, Ed. E.R. Caianiello, pp. 181-188, (Springer Berlin, 1968).
- [7] Dayan P. and Balleine B.W., *Reward, motivation and reinforcement learning*, *Neuron* **36** (2002) pp 285-298.
- [8] DeMaris D., *Dimension change, coarse grained coding, and pattern recognition in spatio-temporal nonlinear systems*. *J.Integr.Neurosci.* **2** (2003) pp 71-102.
- [9] Edelman G.M., *Building a picture of the brain*, *Daedalus* **127** (1998) n.2
- [10] Freeman W.J., *Three Centuries of category errors in studies of the neural basis of consciousness and intentionality*. *Neural Networks* **10** (1975) pp 1175-1183.
- [11] Freeman W.J., *Neurodynamics: an exploration in mesoscopic brain dynamics*. (Springer, Berlin, 2000).
- [12] Gold J.I. and Shadlen M.N., *Neural computations that underlie decisions about sensory stimuli*. *Trends.Cogn.Sci.* **5** (2001) pp. 10-16.
- [13] Grush, R., The semantic challenge to computational neuroscience., In: *Theory and Method in Neuroscience*, edits. Machamer P.K, Grush R. and McLaughlin P., (University of Pittsburgh Press, 2001).
- [14] Haken H., *Brain Dynamics* (Springer, Berlin, 2002).
- [15] Harree R. *Discursive psychology in practice*, (Sage Publications, N.Y., 1995).
- [16] Holden C., *How the brain thinks*, *Science* **303** (2004) p 1121.
- [17] Hopfield J.J., *Neural networks and physical systems with emergent collective Computational abilities*, *Poc.Natl.Acad.Sci. USA*, **79** (1982), pp 2554-2558.
- [18] Kaneko K. and Tsuda I., *Complex systems: chaos and beyond, a constructive approach with applications in life sciences* (Springer, Berlin, 2000).
- [19] Kim J-N and Shadlen M.N., *Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque*, *Nature Neurosci.* **2** (1999) pp 176-185.
- [20] Koch C. Marroquin J. and Yuille A., *Analog "neuronal" networks in early vision*, *Proc.Natl.Acad.Sci.USA* **83** (1986), pp 4263-4267.
- [21] Lakoff J. and Johnson M., *Metaphors we live by* (Chicago University Press, 1980).
- [22] Leon M.I. and Shadlen M.N., *Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque*. *Neuron* **24** (1999) pp 415-425.
- [23] McClure S.M. and Daw N.D. *et al.*, *A computational substrate for incentive salience*, *Trends Neurosci.* **26** (2003) pp 423-428.
- [24] Montague P.R. and Berns G.S., *Neural economics and the biological substrate of valuation*, *Neuron* **36** (2002) pp. 265-284.
- [25] Ortony A., *Metaphor and Thought* (Cambridge University Press, 1993).
- [26] Paton R., *Metaphors, models and bioinformation*, *BioSystems* **38** (1996) pp.155-162.
- [27] Platt M.L. and Glimcher P.W., *Neural correlates of decision variables in parietal cortex*, *Nature* **400** (1999) pp.233-238.
- [28] Poznanski R.R., Introduction to integrative Neuroscience, in: *Biophysical Neural Networks*, pp 1-21, (M.A. Liebert Inc. 2001),
- [29] Quine W.V.O., *Theories and Things* (Harvard University. Press 1981).
- [30] Roessler O.E., *Endophysics: the world as an interface*, (World Scientific, 1998).
- [31] Rosenblueth A. and Wiener N. *et al.*, *Behavior, Purpose and Teleology*. *Phil.Sci* **10** (1943) pp. 18-24.

- [32] Ryle G., *The Concept of Mind* (Barnes and Noble, N.Y., 1949).
- [33] Schall J.D., *Neural basis of deciding, choosing and acting*, Nature Rev.Neurosci. **2** (2001) pp. 33-42.
- [34] Schall J.D., *The neural selection and control of saccades by the frontal eye field*, pp. 173-189 In *The Physiology of Cognitive Processes*, Eds. Parker A., Derrington A. et al., (Oxford Univ. Press 2003).
- [35] Schultz W., *Getting formal with dopamine and reward*, Neuron **36** (2002) pp.241-263.
- [36] Sejnowski T.J. Koch C. and Churchland P.S., Computational Neuroscience, Science **241** (1988), pp.1299-1306
- [37] Shadlen M.N. and Newsome W.T., *Motion perception: Seeing and deciding*. Proc.Nat.Acad.Sci USA **93** (1996) pp. 628-633.
- [38] Shadlen M.N., *Pursuing commitments*, Nature Neurosci. **5** (2002) pp. 819-821.
- [39] Shalizi C.R. and Crutchfield J.P., *Computational mechanics: patterns and prediction, structure and simplicity*, J.Statist.Phys. **104** (2001) pp. 817-879.
- [40] Shalizi, C.R. and Moore C., *What is a macrostate ? Subjective observations and objective dynamics*, (2003) arXiv:cond-mat/0303625 v 1.
- [41] Sirovich L., Omurtag A. and Knight B.W., *Dynamics of neuronal population: the Equilibrium solution*, SIAM J.Appl.Math. **60** (2000), pp. 2009-2028.
- [42] Skarda C.A. and Freeman W.J., *How brains make chaos in order to make sense of of the world*, Behavioral and Brain Sciences **10** (1987) pp. 161-195.
- [43] Stringer S.M. Roll E.T. Trappenberg T.P., and de Araujo,I.E.T., *Self-organizing continuous attractor networks and motor function*, Neural Networks **16** (2003), pp. 161-182.
- [44] Sutton R.S. and Barto A.G., *Reinforcement learning: an Introduction*, (MIT Press, 1998).
- [45] Werner G., *The Many faces of Neuroreductionism*, pp.242-257, In *Dynamics of sensory and cognitive processing by the brain*, Ed. Basar E. (Springer, Berlin, 1985).
- [46] Werner G., *Knowledge and natural selection: five decades on the path to naturalizing Epistemology*, pp.345-359, In *Sensory processing in the mammalian brain*, Ed. Lund J.S. (Oxford Univ. Press 1989).
- [47] Werner G. (2000) Computation in Nervous Systems
http://www.ece.utexas.edu/~werner/Neural_computation.html
- [48] Wise S.P. and Murray E.A., *Arbitrary associations between antecedents and actions*, Trends Neurosci. **23** (2000) pp271-276
- [49] Wuensche, A., *Discrete Dynamcis Lab: tools for investigating cellular automata and Discrete dynamical networks*.
<http://www.santafe.edu/sfi/education/csss/files02/wuensche1.pdf>