Response

The metaphorical brains

Michael Arbib [1]

USC Brain Project, University of Southern California, Los Angeles, CA 90089-2520, USA

Received 9 November 1997; received in revised form 26 January 1998

1. Introduction

In the late 1960s, I asked my "Brains, Machines, and Mathematics" class at Stanford to draw block diagrams identifying and relating the major functional subsystems of the brain. I took these diagrams home, redrew them on large sheets of brown paper, and attached them to the walls of my bedroom so that I could look at them in search of the best composite diagram. When I showed them to John McChesney, a friend from the English Department, he proclaimed, "So, these are your metaphors for the brain". And thus the title of my book *The Metaphorical Brain* was born. Deeper reflection showed that the title had two meanings—not only that our models of the brain can be seen as metaphors, but also that the operation of the brain is inherently metaphorical, building "models of the world" which provide metaphorical descriptions of fragments of the organism's world.

I am most grateful to the Book Review Editors of the journal *Artificial Intelligence* and to Messrs. Barnden, Hanson and Pribram for the Multiple Book Review of *The Metaphorical Brain 2: Neural Networks and Beyond* (TMB2; Arbib [3]) and the chance to respond in these pages. This response is being prepared in 1997, the 25th anniversary of the publication of *The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory* (TMB; Arbib [2]) to which TMB2 is the sequel. The subtitles tell something of the changing times of these books: In 1972, "symbolic AI" was dominant, and I sought to balance this dominance by looking back to the Cybernetics of the 1940s and to show that it was the source not only of AI in this limited sense but also of brain theory, as well as biological control theory, and cognitive psychology. In 1989, half the battle had been won and artificial neural networks (Hopfield networks, back-propagation, etc.) were enjoying a heyday, and so now my subtitle was designed to recognize the importance of such neural networks, but also to show that one must go "beyond", most especially with the development of schema theory, but also with a fuller appreciation of neuroscientific data, and of the immense subtlety of living neurons.

[1] Email: Arbib@pollux.usc.edu; http://www-hbp.usc.edu/HBP/.

Barnden has done a fine job of reviewing the contents of TMB2, and all three reviewers have offered some pointed criticisms of the book. I will not attempt to repeat Barnden's summary, nor answer every detailed comment, but I will take advantage of this opportunity to clarify some of my ideas and to give a sense of how my views have changed in the past decade, as reflected in part in the book *Neural Organization: Structure, Function, and Dynamics* (AESz; Arbib, Érdi and Szentágothai [9]).

Barnden notes that TMB2 "is not, and could not be, anywhere near encyclopedic on such fields; the emphasis is rather on aspects of those fields that serve in some way to illustrate, support or constrain the central theoretical ideas of the volume." For example, Barnden notes that TMB2 contains no description of "... Adaptive Resonance Theory... [nor of] automatic learning of topological maps", and so I am happy to note that these are treated in *The Handbook of Brain Theory and Neural Networks* (HBTNN; Arbib [5]), a book which *is* encyclopedic in its coverage of brain theory and neural networks.

Hanson suggests TMB "was great and there are probably diminishing returns on reading [TMB2] if you've already read [TMB]". However, I strongly disagree about the diminishing returns. There is so much new in TMB2. But I think the use of "2" may have made people think the returns would be diminishing. Anyway, in writing such a book there is always the issue of what needs to be reiterated and what can be omitted either because it is disparate or because it can now be assumed known. The main themes of this response are that (a) TMB2 is still very much worth reading eight or more years after publication because it still says important things not well covered elsewhere in the AI and neural net (NN) literature (including the first TMB); (b) it was not perfect and some pointers can add to a reading of it today, and (c) the field continues to progress, and some facets of that progress will be outlined below.

2. The hologram metaphor

Karl Pribram finds that "What is presented is valuable but seems to reflect accepted dogma in the neurosciences (as Horace Barlow [15] has called it) rather than a broader perspective of metaphors and their instantiation and substantiation in the neuroscience laboratory." In fact, Barlow asks whether "Single Units and Sensation" provide "A Neuron Doctrine for Perceptual Psychology", and I would say that TMB2 is far from accepting this as dogma. TMB2's emphasis on action-oriented perception shifts the frame of reference for perceptual psychology, and its emphasis on schemas and cooperative computation takes us far from single units and sensation. However, the real problem seems to be that I do not value Pribram's hologram metaphor: "What I [Pribram] miss most, of course, is a follow-through on what in the earlier book TMB was entitled 'The Hologram Metaphor' to which Arbib devoted a section of six pages. In *The Metaphorical Brain 2* he dismisses this particular metaphor with [a single] paragraph." Pribram then devotes 2600 words of his 3300 word review to that one paragraph!

He later asks "Why, with the exception of the development of PDP neural networks, does Arbib... exclude the work that has developed from the holographic metaphor? There must be some clear intellectual basis for this exclusion." The basis is both generic and specific:

(a) Even though TMB2 is longer than TMB, there were a number of items in TMB that were removed to stop TMB2 from being even longer.
(b) As my quoted paragraph made clear, I was not convinced of the validity of the hologram metaphor as Pribram viewed it; and the very success of work on Hopfield nets and other models of associative networks led me to devote space to these rather than to Pribram's writings.

However, since Pribram has set forth his work at such great length, I should respond to a number of the issues he raises.

I learned a great deal of neuropsychology from Pribram when I was a junior faculty colleague of his at Stanford, and I remain grateful for this experience and value what I learned. I also studied most of the papers by Pribram and others available at that time (1965–1970) and formed the conclusion then that a look at his new book (Pribram [33]) only confirms. We owe to Pribram many rich insights from his studies of the behavior of lesioned monkeys into the way in which distinctive psychological functions (schemas) are distributed over distinctive sets of brain regions (see, e.g., Pribram [31]). Such insights are too often ignored by people whose study of NNs is based on the adaptation of a single, initially unstructured network. However, his skill in the neuropsychological laboratory does not, in my opinion, transfer to theory. Pribram allows himself one low blow in his review when he says that "[the quoted] paragraph illuminates what can go wrong when one reports only what one reads and has no direct experience at the neurophysiological bench". I could equally say of Pribram's 1991 book that the Appendices illuminate what can go wrong when one relies on others for one's mathematics and has no direct experience of mathematical research. These appendices were written primarily by Kunio Yasue and Mari Jibu, and essentially recycle the type of mathematics familiar in quantum mechanics. It is claimed that the theory is relevant because the functions are defined on a "dendritic manifold", but the mathematics remains formal, and no evidence is given that it can be used to analyze behaviors of realistic sets of linked dendritic trees within a realistic glial environment receiving realistic patterns of synaptic input. My favorite is Appendix D which starts with a mention of amygdalectomy but then provides only a general formalism with no theorems deduced from it, and no mention of how one would tell a "normal" dendritic manifold from one appropriate to a brain from which the amygdala has been removed. But enough for low blows and counterpunches. Let me stress that Pribram draws our attention to a number of important phenomena, even while I disagree with his theoretical claims.

Pribram cites much interesting work on visual neurophysiology, but I still see "no evidence that the neural system has either the fine discrimination of spatial frequencies or the preservation of spatial phase information needed for such Fourier transformations to be computed with sufficient accuracy to be useful [for the reconstruction operations demanded by the original hologram theory]." However, it is true that Gabor functions and wavelets have been applied quite successfully in vision, though more for recognition than reconstruction. Anyway, no book can cover everything! Even HBTNN does not discuss Pribram's version of the hologram metaphor—but it does have articles on wavelets and on the use of holograms in photonic implementation of NNs, as well as three articles which mention other contributions by Pribram.

Pribram asks: "Why not give credit to the holographic metaphor for being one of the critical factors... helping to give birth to the PDP neural networks enterprise? David Willshaw['s ...] thesis and subsequent work with Buneman and Longuet–Higgins were inspired by holography. This work was already quoted in the chapter on Neural Holograms in my *Languages of the Brain* (Pribram [32])." I am happy to acknowledge that Pribram's hologram metaphor was one important motivation for work on associative networks (see TMB for more on this). However, the correspondence with Willshaw et al. reproduced by Pribram [32] explicitly rejects the "literal" form of the hologram metaphor espoused by Pribram, and offers a different metaphor more in tune with current associative networks.

Pribram notes that Figure 2.3.9 of TMB2 includes a panel at a magnification "displayed in such a way that one can readily imagine the neurochemical field effects that must be operating as processing takes place." He then argues that "It is at the synapto-dendritic level that neural computation occurs. The neuron is not the unit of processing. The synapse is". There is no unique "the" unit of processing, but I think Pribram is right to insist that we pay attention to this level of detail.

(i) Pribram notes that "simulations of [synapto-dendritic] processes... have been made by Shepherd... and by Perkel.... These simulations deserve a hearing...". Yes, it is good work, but Wilfrid Rall's seminal work on dendritic processing is based not on holonomy but on extensions of the Hodgkin–Huxley equations (see the articles by Rall and by Segev in HBTNN for more details).

(ii) People studying synaptic plasticity are now paying attention to the possible role of the diffusion of nitric oxide in the intracellular medium.

(iii) There is increasing work on glia, though the role of glia seems more concerned with neuron maintenance and repair than with "computations" (though this may change).

(iv) Certainly, field effects deserve and are receiving study.

My approach is to seek the simplest model that does the job, and in TMB2 I found that much could be explained at the level of synaptic interaction of neurons without attention to finer levels of detail or to field effects. However, I am prepared to accept complexities when I have to do so, and recent findings push me below the neuron to consider neuromodulation, eligibility, and dendritic processing. Indeed, although in TMB2 I chose to use the neuron as the cutoff level, flags for more detail were raised in Sections 6.1 (Neuromodulation) and 8.1 (Synaptic Complexities). An important virtue of HBTNN is that it draws the attention of the reader to these complexities. Roadmap II.5 (pp. 45–46) of HBTNN introduces the reader to the relevant articles. However, it is true that topics like field effects, volume conduction, and glial processing were mainly ignored, and should not be in the next edition.

In summary: There is no doubt that the work of Gabor and Shannon provides tools that are being used in Brain Theory today (HBTNN provides a rich set of examples), and Pribram draws our attention to important styles of neural processing of which Brain Theory currently takes little heed. His latest book, *Brain and Perception: Holonomy and Structure in Figural Processing*, contains a wealth of noteworthy empirical data—but it has not convinced me that neural holograms or the holonomic brain theory provides the right framework for analysis of these data.

3. Frogs

Hanson notes that there are "10 references on frogs peppered throughout the text" and adds that "I don't really mean to disparage frogs here. I even think the examples of the frog's visual system are presented in a relevant useful way. Nonetheless,... although the frog's tectum gets fair representation in the book, it is hard to struggle back to the schemas and computational backdrop that the book promotes."

My response has two trivial parts:

(a) TMB2 reflected my then current publications and enthusiasms. In hindsight, I would say that the material in Section 5.1, Schemas for Frog and Toad, is still appropriate, while the exposition in Section 7.3, Modeling the Frog Tectum, was too detailed and that today's reader may well skip the latter section.

(b) I think the richness of topics in TMB2 is such that there would be discontinuities in switching from one important topic to another, no matter what pruning or reordering one attempts.

Now to the important part, namely to demonstrate why an interest in frogs is not an authorial indulgence, but an important ingredient in a general understanding of the brain:

(i) The best-known study of visual coding, the work of Hubel and Wiesel on the cortex of cat and monkey, emphasizes "general" features like bars and edges that are neutral to the life of the animal. Lettvin et al. (1959), in studying "What the Frog's Eye Tells the Frog's Brain", not only provided equally important results on neural coding per se, but also related this visual coding to the behavior of the animal, with talk of "bug detectors" and "enemy detectors". As TMB2 shows, much has been done to show how further neural circuitry is required to build on this initial processing. Nonetheless, we have here a fundamental insight into "action-oriented perception" missing in the mammalian literature of that time.

(ii) Much classic literature of vision focused on "object recognition" as mediated by visual cortex. It was the work on frog vision that inspired the insight into subcortical mechanisms of vision that could subserve visually-guided action in the absence of visual cortex, and provides an evolutionary basis for the later recognition of separate cortical pathways for "vision for action" and "vision for recognition" (Milner and Goodale [30]; Jeannerod [26]).

(iii) Discussion of monkey superior colliculus (the mammalian homologue of tectum) has mainly focused on "approach behavior", namely eye movements to fixate a target, whereas frog studies have long addressed both "approach to prey" and "avoidance of enemies". It is only relatively recently that studies in rat (Dean, Redgrave and Westby [18]) have brought the balanced study of approach and avoidance into the mammalian literature.

(iv) The models of these frog behaviors have inspired work in robotics (e.g., Arkin [12, 13]).

In this spirit, mechanisms of frog visuomotor coordination are again presented in the "Functional Overview" in AESz, with less important results from TMB2 omitted, and new results and evolutionary insights added. More generally, comparative neuroethology, the study of the neural mechanisms underlying diverse forms of behavior in different species, provides vital lessons both for the study of the human brain, and for the design of novel

robots. Unfortunately, a lack of understanding of the historical and continuing importance of studies of frog vision and behavior has made it very hard to get funding for related research, and the database for frog neuroethology has come close to a standstill.

4. From action to language

This section of the response provides a brief snapshot of recent work on action-oriented perception in mammals which builds atop my earlier studies of the frog. Arbib [7] explores the hypothesis that various subregions (but by no means all) of posterior parietal cortex are specialized to process visual information to extract a variety of *affordances* for behavior. This discussion of affordances is informed by the attempt of Section 7.2 of TMB2 to place the ecological perception of J.J. Gibson within a computational framework. We introduce two biologically-based models of regions of monkey posterior parietal cortex. The model of the lateral intraparietal area (LIP) emphasizes its roles in *dynamic remapping* of the representation of targets during a double saccade task, and in combining stored, updated input with current visual input (Dominey and Arbib [19]). The model of the anterior intraparietal area (AIP) addresses parietal-premotor interactions involved in *grasping*, and analyzes the interaction between AIP and premotor area F5 (Fagg and Arbib [22]). The model represents the role of other intraparietal areas working in concert with inferotemporal cortex as well as corollary discharge from F5 to provide and augment the affordance information in AIP, and suggests how various constraints may resolve the action opportunities provided by multiple affordances. Finally, we develop a systems-level model of hippocampo-parietal interactions underlying rat *navigation* motivated by the monkey data used in developing the above two models as well as data on neurons in monkey posterior parietal cortex sensitive to visual motion. We note the formal similarity between dynamic remapping (primate saccades) and path integration (rat navigation), and explain certain available data on rat posterior parietal cortex in terms of affordances for locomotion. We also suggest the utility of further modeling linking the World Graph model of cognitive maps for motivated behavior with hippocampal-parietal interactions involved in navigation (Guazzelli et al. [23]). These models demonstrate that posterior parietal cortex is not only itself a network of interacting subsystems, but functions through *cooperative computation* with many other brain regions, thus demonstrating anew one of the basic claims of TMB2.

Section 5.4 of TMB2 presents three models of language which are described in much greater detail by Arbib, Conklin and Hill [8]. My main foray into language since then has built on the work on brain mechanisms of grasping mentioned above. A key finding for that work was that neurons located in the rostral part of inferior area 6 (area F5) discharge during active hand and/or mouth movements and that discharge in most F5 neurons correlates with an action rather than with the individual movements that form it (Rizzolatti et al. [35]). One may thus classify F5 neurons into various categories corresponding to the action associated with their discharge, the most common being "grasping-with-the-hand" neurons, "grasping-with-the-hand-and-the-mouth" neurons, "holding" neurons, "manipulating" neurons, and "tearing" neurons—a "vocabulary" of motor schemas.

Further study revealed something unexpected—a class of F5 neurons that discharge not only when the monkey grasped or manipulated objects, but also when the *monkey*

observed the experimenter make a gesture similar to the one that, when actively performed by the monkey, involved activity of the neuron (Rizzolatti et al. [36]). Movements yielding mirror neuron activity when made by the experimenter include placing objects on or taking objects from a table, grasping food, or manipulating objects. Mirror neurons, in order to be visually triggered, require an interaction between the agent of the action and the object of it. The simple presentation of objects, even when held by hand, does not evoke the neuron discharge. These neurons (mirror neurons) appear to represent a system that matches observed events to similar internally generated actions, forming in this way a link between the actor and observer. Transcranial magnetic stimulation and PET experiments showed that a mirror system for gesture recognition exists also in humans and includes Broca's area. Rizzolatti and I (Arbib and Rizzolatti [10]; Rizzolatti and Arbib [34]) have argued that it is the addition of such an observation/execution matching system that provides the necessary bridge from "doing" to "communicating about doing" in the evolution of the neural mechanisms for language. Many authors have argued for a gestural basis for language (e.g., Hewes [25]; Kimura [28]; Armstrong et al. [14]). We bolster the argument by asserting the crucial role of an observation/execution matching system for grasping in monkeys which provides a representation of "expectations" which can both guide the monkey's own actions and enable it to comprehend the actions of other monkeys. We posit that a crucial process in language evolution was that the link between actors and observers became a link between the senders and receivers of messages.

5. Schema theory

Pribram notes that the concept of a "[schema], used extensively by Henry Head and by Sir Frederick Bartlett, is meant to give unity to the exposition. A schema, Arbib holds, is much like a program in that it has many different instantiations. On p. 207 the attributes of schemas are listed. I interpret these as defining distributed map-like structures which can be cascaded in a variety of combinations during processing. In a sense they are a more sophisticated version of cell assemblies (such as those proposed by Donald Hebb (1949)) which partake of the property that any particular neuron or schema can participate in a variety of processes by being coupled to a variety of other assemblies or schemas."

Indeed, schema theory attempts to bridge between structure and function at the highest level. This theme is further developed in Chapter 3, "A Functional Overview", of AESz.

Barnden is concerned that "the detailed discussion of the dynamics and feedback control of muscles makes little direct contact with schema theory (though clarifying what motor schemas need to do in order to control muscles), and the same is true of the overviews of artificial neural net frameworks." But this is to miss the point that the essence of a multi-level approach to a brain, or an artificial cognitive system, is that no one level is sufficient. Schemas (functional) and neural networks (structural) are just two of the levels extending down into neurons, dendrites, synapses, and the molecular processes beneath them. As is clear from my response to Pribram's concerns about "The Hologram Metaphor", there is no single level of analysis of the nervous system (structurally stated) or cognition (functionally stated). Methodologically, detailed modeling at one level may serve to calibrate an approximate model at a higher level—it is possible to model the differential

effect of thousands of synapses on a single neuron; but if one must model hundreds of thousands of neurons, then one must determine a reasonably accurate way of aggregating the effect of those synapses on each neuron. Another form of aggregation is the use of statistical mechanics to analyze neural networks, though here the constraint of biological subtleties is still scant (Zippelius and Engel [39]). Barnden is concerned that "some readers may feel that [Arbib's] approach leads to too discrete a division of behavior into schemas. For instance, in discussing prey acquisition by toads, Arbib proposes (pp. 219–222) one perceptual schema for detecting a barrier, one for detecting a chasm, and one for detecting the free prey condition (no barrier or chasm). But there is no argument that it is indeed appropriate to postulate separate schemas here." However, whether or not there are "separate schemas here", I stress that schema-level models can be invalidated, and provide the basis for yet better models, without descending to lower levels, on the basis of lesion and behavioral data (see, e.g., Fig. 3.3 and Section 3.4 of AESz).

Barnden continues "[Arbib] points out that a single region of the brain might subserve many schemas, and a schema (or instance) might involve several regions. (I would have preferred less talk of regions as opposed to just subsets of the total set of neurons.)" This last point is well taken. A given brain region contains many subsets of neurons which may be distinguished anatomically or, even within an anatomically defined class, physiologically by their differential involvement in different functions. For example, Dominey and Arbib [19] analyze several brain regions involved in the rapid eye movements known as saccades, and in each region separate those neurons involved in the Current Saccade Motor Schema from the Target Memory Schema that holds data about the next saccade.

Barnden asserts that "analogical processing as a replacement for more rigid deductive reasoning is an important and fruitful idea, one that could usefully have received extensive and detailed discussion in the book." Mary Hesse and I have discussed such ideas at some length in *The Construction of Reality* (Arbib and Hesse (1986)) within an epistemological context, but it is true that much further work must be done to link studies of analogy with the schema theory of TMB2. Hesse and I also study, informally, "social schemas". These are patterns of behavior exhibited by many members of a society (cf. the "collective representations of Durkheim [21, p. xii]), and which we see as providing the statistical regularities for a newcomer to a society (e.g., a child acquiring language, TMB2, Section 5.4) to interiorize these regularities via the formation of "schemas in the head".

Barnden finds me "sympathetic to high-level units of behavior emerging in some subtle, distributed way from neural networks. This resonates with standard connectionist claims that such things as rules can be merely emergent from subsymbolic computation.... Arbib also says that high-level units such as rules can enable the system to avoid a great deal of expensive, highly parallel, low-level processing (p. 247) ... [but] one might have conjectured that highly parallel, low-level processing is not a drain on the overall system in the first place." I can address this by expanding upon Seidenberg's [38] claim that a connectionist model of past-tense learning in English does more justice to psycholinguistic data than a rule-based model which sees the user of English learning the rule "add -ed" plus a list of exceptions "is-was, go-went, etc." because, for example, subjects will provide an irregular past form for a nonsense word if it is similar to a set of irregular verbs with parallel past tenses. However, Seidenberg does not address the fact that, when learning a

second language, we do memorize rules for the past tense and invoke them in the early stages of learning. My claim would be that we have (to simplify) two interacting neural nets, one which can learn rules, and one which can process data in a "connectionist way". When learning our first language, we first learn "connectionistically" (i.e., statistically through experience) and may later come to learn rules by reflecting upon this mechanism. Conversely, when we learn a second language in the usual manner, the "rule network" can both serve to control language use initially, and to train the "connectionist net". Fluency comes when the connectionist net can provide correct behavior more quickly than can invocation of the rule net. Much connectionism focuses on single networks being trained by a single learning rule. By contrast, TMB2 stressed the interaction of multiple networks with different architectures and functions. This idea has at last begin to enter the connectionist literature (Jordan and Jacobs [27]). The deeper point is that the ability of the mind to reflect on its own operations is a crucial aspect of human rationality (whatever its imperfections) and must eventually find its proper place within brain theory—just as the notion of "knowledge about knowledge" presents an enduring challenge for AI in general. This is related to the view of consciousness offered in TMB2, Section 8.3.

Barnden notes that "The dynamic formation of... higher-level schemas is one type of learning that Arbib hints at, although he does not provide a detailed account of how it might work;" while Hanson suggests that "Schema acquisition might be an appropriate place to start understanding the structure of schemas and the nature of their organization for brain function. After all, the way schemas are learned and then subsequently guide learning might be diagnostic of the way they are used and organized by the brain. Surely, the subtle interplay of perception and memory would be a key element of schema acquisition and consequent usage." The work of Jane Hill (briefly reviewed in Section 5.4 of TMB2) is one excursion in that area. Recently, Fernando Corbacho has laid the groundwork for a general theory of Schema-Based Learning (SBL). A very preliminary version appeared in (Corbacho and Arbib [17]); more general papers are now in preparation.

However, most work on schema learning by my group has not involved a general theory of SBL, but has rather involved the following sequence: (i) Given some behavior of interest, show how it may be achieved through the interaction of multiple schema instances, then (ii) show how, by careful attention to the data of neuroscience, these schemas may be implemented through the cooperative interaction of multiple biological neural networks of the brain, and, finally (iii) show how learning (modeled in terms of synaptic plasticity in neural networks) can improve the efficacy of these schemas. Studies of this kind include:

- the role of cortico-striatal plasticity—modeled using reinforcement learning—in visuo-motor conditional learning and sequence learning (AESz, Sections 10.4.2 and 10.5.1), and
- the role of cerebellum—modeled using error-based learning—in improving motor control in both saccades and throwing (Schweighofer, Arbib and Dominey [37]; Arbib, Schweighofer and Thach [11]; AESz, Section 9.4.4).

Barnden found it unclear how my "schemas resemble or differ from the schemas of Piaget, or the similar constructs of other authors whose work he reviews." I would say that Piaget tends to concentrate on large undifferentiated schemas, and tends to see them as passing through a pre-ordained series of stages, whereas I see schemas as being of a far finer grain so that an assemblage of schema instances is active at any one time, with

stages then emergent from schema interactions and accommodation, rather than being the determinants of schema formation. Arbib [6] is a further attempt to place my work within a far broader context of what may be seen as contributions to schema theory from Kant to the present day.

Hanson accepts the need for schemas, but laments that "there are promissory notes throughout the early part of the text not paid off later. Instead we get a scholarly discussion of the topics and concepts, but typically without much resolution. For example, one might posit different types of schemas, ones perhaps that vary in abstractness or perhaps have some domain or topic dependency and then this begins to look like a vocabulary, dictionary or even simple taxonomy. Unfortunately, nothing like this is spelled out in the book.... what seems to be avoided here are serious issues of Generalization... given a schema how do I calculate its similarity to other schemas and other input data? What determines the level of abstractness of a given schema? When does a schema have precedence over sensory data?" As the earlier paragraphs make clear, some of these promises have been kept, but others are overdue and I can only hope that others will join the small group who are already contributing to the development of the theory. However, I think it may be useful to note that the term "schema theory" is being used in two different senses (Arbib [4]):

> There are two facets to schema theory, as a language and as a theory of intelligent behavior. In, for example, RS (Lyons and Arbib [29]) and the VISIONS schema system (Draper et al. [20]), we see the beginnings of a language for distributed systems at a level abstract enough to convey some real understanding of complex problem solving behaviors and yet precise enough that we can refine the specification to some concrete implementation. Yet when we discuss brain models and advocate a class of designs for AI systems that is inspired by them, we see the other facet of schema theory, a model of intelligence, which uses schema theory *qua* language for *expressing* such models. In the language sense, schema theory is more like group theory than relativity theory. Relativity theory is a model of the physical world—it can be falsified or revised on the basis of physical experiments. However, group theory stands or falls for the scientist seeking to explain the world (as distinct from the mathematician proving theorems) not by any criterion of whether it is true or false, but rather on whether its terminology and theorems aid the expression of successful models. Schema theory as an abstract model of computation does not yet have the rigor or stock of theorems of group theory, but the success of models using the language of instantiation, modulation, activity levels, etc., strengthens its claim to be a valuable tool in the development of Artificial Intelligence as well as Brain Theory. However, the language of schema theory has developed in tandem with a schema-based theory of human and artificial intelligence, such as that given in the sections [of [4], and related sections of TMB2] on Schema Modulation and Evolution, Memory, Perception, and Action, Schemas for *Rana computatrix*, Schemas and High-Level Vision, and Schemas and Learning. It is in the latter sense that one can make such statements as that in schema-theoretic models of language and other cognitive functions "there is a tendency (though not a necessity) to root such models in action and perception." In either sense, the paradigm of schema theory is indeed evolving in a fashion well suited to contribute to distributed artificial intelligence as well as to bridge between cognitive science and brain theory, too.

Appendix A. Structuring the book

TMB2 has an unevenness of tone because, as Hanson says, "At times TMB seems to be an attempt at more of a textbook genre, as there is considerable introductory material... [but Arbib] migrates beyond simple textbook structures to a more personal view of the field involving computation, brains and mathematics." For example, computer scientists and psychologists know so little "in-discrete" mathematics that the author has a difficult choice: to avoid mathematics entirely, to supply the necessary exposition, or to assume that the reader has the necessary background. Churchland and Sejnowski [16] take the first approach (there are a few skippable equations), TMB2 takes the second, while AESz takes the third. The middle way is probably less elegant but, for a certain class of readers, more helpful than the other two. The reader who wishes to "hum the mathematics" (in the memorable phrase of Peter Medawar) rather than read the details will still be able to reap most of the benefit of TMB2.

Pribram finds the organization of the themes that unify the book unsatisfactory: "Beginning with the section on Schema Theory (p. 204) we are shuttled from frogs to Shepard figures and machine vision and shortly thereafter to human language. Arbib uses these various topics as examples of the importance of a certain type of internal representation which he calls schemas.... My problem is not with the concept of schema per se but with separating this section from discussions in Part III of vision, action and memory. In Part III one loses sight of schemas and thus of the putative coordinating theme of the volume."

In TMB2, my strategy was to offer an introductory overview before introducing neural networks with examples from vision, action and memory, then introduce schemas with examples from vision, action and memory, and finally to offer separate chapters on vision, action and memory in turn. The reader who wishes a new ordering of the sections when reading TMB2 might use the following course syllabus. The material was supplemented by lectures from HBTNN:

> *Introduction*: Chapter 1. Brain, Behavior and Metaphor; 2.1. Action-Oriented Perception and Intelligence; 2.3. The Brain as a Network of Neurons; 2.4. A Functional Perspective on Neuroanatomy; 3.3. Visual Preprocessing; the portion of 4.3 on the Didday Model of Prey-Selection (Winner-Take-All); NSL Neural Simulation Language (HBTNN).
> *Systems Theory*: 3.1. Systems Concepts, 3.2. Feedback and the Spinal Cord.
> *An Introduction to Schema Theory*: 2.2. Perceptual and Motor Schemas; 4.1. McCulloch's Principles; 4.2. Constraint Satisfaction & Relaxation; 5.1. Schemas for Frog and Toad.
> *Learning*: 3.4. Adaptive Networks; 8.2. Connectionist Models of Adaptive Networks; Reinforcement Learning (HBTNN), Reinforcement Learning and Motor Control (HBTNN); 4.3. Competition and Cooperation in Neural Nets; Self-Organizing Feature Maps—Kohonen maps (HBTNN). Statistical Mechanics of Neural Networks (from Hertz et al. [24]); Artificial Intelligence and Neural Networks (HBTNN), Cognitive Modeling: Psychology and Connectionism (HBTNN); 8.3. Memory and Consciousness.

Vision: 7.1. Depth Perception; 7.2. Optic Flow; 5.2. Vision in Perspective; 7.4. The Many Visual Systems.

Action: Reactive Robotic Systems (HBTNN); 6.1. Rhythm Generators and Locomotion, 6.2. Control of Eye Movements.

TMB2 material not covered this time were: 4.4. A Mathematical Model of Neural Competition; 5.4. From Schema Theory to Language; 7.3. Modeling the Frog Tectum; 8.1. Mechanisms of Neuronal Plasticity; 9.1. Where we are; and 9.2. From Brain Theory to the Sixth-Generation.

Hanson speculates that "The revolution [in the study of artificial neural networks] in recent years, may have begat interest in the updating of the book." In fact, this is not true— the growth in understanding of the themes introduced in TMB was the main motivation for TMB2. Hanson is correct however that "very little about the coherence of neural networks receives much confirmation from [TMB2.... The] book seems to be unaffected by the revolution. Indeed, in a late chapter on Memory and Learning we find references and discussions to PDP, Kohonen etc. almost like an afterthought." However, this is in part the "Pribram Problem" again—*something* had to come last, and I chose the order Vision, Action, Memory for my concluding round. What Hanson fails to note is that key ideas (as distinct from formal theory) are elaborated early, in Section 4.3, in highlighting the theme of Competition and Cooperation in Neural Nets. However, in 1988 there were few results that strongly linked PDP and Kohonen to my main concern—the integration of action and perception. As both HBTNN and AESz (with Péter Érdi's strong emphasis on dynamics at various levels of neural modeling) demonstrate, much has changed in the last decade, and the above course outline reflects those changes.

References

[1] M.A. Arbib, Brains, Machines and Mathematics, McGraw-Hill, New York, 1964.
[2] M.A. Arbib, The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory, Wiley-Interscience, New York, 1972.
[3] M.A. Arbib, The Metaphorical Brain 2: Neural Networks and Beyond, Wiley-Interscience, New York, 1989.
[4] M.A. Arbib, Schema theory, in: S. Shapiro (Ed.), The Encyclopedia of Artificial Intelligence, 2nd ed., Wiley-Interscience, 1992, pp. 1427–1443.
[5] M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, A Bradford Book/The MIT Press, Cambridge, MA, 1995.
[6] M.A. Arbib, Schema theory: from Kant to McCulloch and beyond, in: R. Moreno-Diaz and J. Mira-Mira (Eds.), Brain Processes, Theories and Models. An International Conference in Honor of W.S. McCulloch 25 Years after His Death, The MIT Press, Cambridge, MA, 1995, pp. 11–23.
[7] M.A. Arbib, From visual affordances in monkey parietal cortex to hippocampo-parietal interactions underlying rat navigation, Phil. Trans. Roy. Soc. London B 352 (1997) 1429–1436.
[8] M.A. Arbib, E.J. Conklin, J.C. Hill, From Schema Theory to Language, Oxford University Press, 1987.
[9] M.A. Arbib, P. Érdi, J. Szentágothai, Neural Organization: Structure, Function, and Dynamics, A Bradford Book/The MIT Press, Cambridge, MA, 1997.
[10] M.A. Arbib, G. Rizzolatti, Neural expectations: a possible evolutionary path from manual skills to language, Communication and Cognition 29 (1997) 393–424.
[11] M.A. Arbib, N. Schweighofer, W.T. Thach, Modeling the cerebellum: from adaptation to coordination, in: D.J. Glencross, J.P. Piek (Eds.), Motor Control and Sensory-Motor Integration: Issues and Directions, North-Holland Elsevier Science, Amsterdam, 1995, pp. 11–36.

[12] R.C. Arkin, Neuroscience in motion: the application of schema theory to mobile robotics, in: J.-P. Ewert, M.A. Arbib (Eds.), Visuomotor Coordination: Amphibians, Comparisons, Models, and Robots, Plenum Press, 1989, pp. 649–671.

[13] R. Arkin, Behavior-Based Robots, The MIT Press, Cambridge, MA, 1997.

[14] D.F. Armstrong, W.C. Stokoe, S.E. Wilcox, Gesture and the Nature of Language, Cambridge University Press, Cambridge, 1995.

[15] H.B. Barlow, Single units and sensation: a neuron doctrine for perceptual psychology?, Perception 1 (1972) 374–394.

[16] P.S. Churchland, T.J. Sejnowski, The Computational Brain, A Bradford Book/The MIT Press, Cambridge, MA, 1992.

[17] F.J. Corbacho, M.A. Arbib, Learning to detour, Adaptive Behavior 4 (1995) 419–468.

[18] P. Dean, P., Redgrave, G.W.M. Westby, Event or emergency? Two response systems in the mammalian superior colliculus, Trends in Neurosci. 12 (1989) 137–147.

[19] P.F. Dominey, M.A. Arbib, A cortico-subcortical model for generation of spatially accurate sequential saccades, Cerebral Cortex 2 (1992) 153–175.

[20] B.A. Draper, R.T. Collins, J. Brolio, A.R. Hanson, E.M. Riseman, The schema system, Internat. J. Computer Vision 2 (1989) 209–250.

[21] E. Durkheim, The Rules of Sociological Method, Translated by S.A. Solovay, J.H. Mueller, Collier-Macmillan, London, 1938.

[22] A.H. Fagg, M.A. Arbib, Modeling parietal-premotor interactions in primate control of grasping, Neural Networks (1998) to appear.

[23] A. Guazzelli, F.J. Corbacho, M. Bota, M.A. Arbib, Affordances, motivation, and the world graph theory, Adaptive Behavior (1998) in press.

[24] J. Hertz, A. Krogh, R.G. Palmer, Introduction to the Theory of Neural Computation, Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA, 1991.

[25] G. Hewes, Primate communication and the gestural origin of language, Current Anthropology 14 (1973) 5–24.

[26] M. Jeannerod, The Cognitive Neuroscience of Action, Blackwell Publishers, Oxford, 1997.

[27] M.I. Jordan, R.A. Jacobs, Modular and hierarchical learning systems, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, A Bradford Book/The MIT Press, Cambridge, MA, 1995, pp. 579–582.

[28] D. Kimura, Neuromotor Mechanisms in Human Communication, Oxford University Press, New York, 1993.

[29] D.M. Lyons, M.A. Arbib, A formal model of computation for sensory-based robotics, IEEE Trans. Robotics and Automation 5 (1989) 280–293.

[30] A.D. Milner, M.A. Goodale, The Visual Brain in Action, Oxford University Press, Oxford, 1995.

[31] K.H. Pribram, A further experimental analysis of the behavioral deficit that follows injury to the primate frontal cortex, J. Exper. Neurol. 3 (1961) 432–466.

[32] K.H. Pribram, Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[33] K.H. Pribram, Brain and Perception: Holonomy and Structure in Figural Processing, Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

[34] G. Rizzolatti, M.A. Arbib, Language within our grasp, Trends in Neuroscience 21 (1997) 188–194.

[35] G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, M. Matelli, Functional organization of inferior area 6 in the macaque monkey: II. Area F5 and the control of distal movements, Exp. Brain Res. 71 (1988) 491–507.

[36] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi, Premotor cortex and the recognition of motor actions, Cogn. Brain Res. 3 (1996) 131–141.

[37] S. Schweighofer, M.A. Arbib, P. F. Dominey, A model of adaptive control of saccades: I. The model and its biological substrate, Biological Cybernetics 75 (1996) 19–28.

[38] M. Seidenberg, Linguistic morphology, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, A Bradford Book/The MIT Press, Cambridge, MA, 1995, pp. 546–549.

[39] A. Zippelius and A. Engel, Statistical mechanics of neural networks, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, A Bradford Book/The MIT Press, Cambridge, MA, 1995, pp. 930–934.