Book Review

F. Varela, E. Thompson and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*[1]

Anand Rangarajan

Dept. of Computer Science, Yale University, New Haven, CT 06520-2158, USA

Dept. of Diagnostic Radiology, Yale University, New Haven, CT 06510-3333, USA

1 What the book is about

The Embodied Mind classifies all research in cognitive science into three major categories; cognitivist, emergent and enactive. Cognitivism is described as "no computation without representation" and is criticized for holding to the view of a world with pregiven features that are subsequently represented by cognitive systems. The criticism is not directed at the failure to answer questions of meaning. Instead, cognitivism is criticized for failing to take into account the feedback from action to perception via the actions of a situated, embodied cognitive agent in the real world. The emergence school (connectionism) is described as "learning representations via an optimal fit between system and environmental features" and is criticized for being silent on the issue of representation. Despite the emergent school's focus on self-organization, it's lumped in the same boat as cognitivism due to its failure to take into account the perception/action loop alluded to earlier. The enactive approach is preceded by an extensive survey of color and color vision in birds, animals and humans. External objectivism (the view that colors are surface reflectances and are to be found in the external

[1] (The MIT Press, Cambridge, MA, 1991); xx + 308 pages, \$25.00 (hardcover), \$13.95 (paper).

world) is criticized for failing to account for the opponent nature of hues, colors as perceived attributes of the sky, afterimages and dreams and for neglecting the role of color vision in surface segmentation. Neurophysiological subjectivism (the view that colors are in the head) is criticized for failing to recognize that colors belong to a shared biological and cultural world. A circular definition of color is adopted. It is argued that this circularity does not prevent color from exhibiting universals that can be studied by cognitive science. The discussion of color serves as a springboard for the introduction of the enactive approach. The enactive approach is presented as a middle way between the extremes of objectivism and (neurophysiological) subjectivism. Self-organization as viable, structural coupling between the cognitive system and the world (with appropriate perception/action loops) forms the kernel of the enactive program.

The second major theme in the book is the nature of the self. A big departure is taken from all forms of materialism. A fundamental circularity is introduced between cognitive science and external behavior plus personal experience. While the role of cognitive science in performing an information processing characterization of behavior and experience is recognized, the entire scientific endeavor called cognitive science is itself brought into circulation as collective behavior and experience. This circularity is introduced to forestall all attempts at identifying the information processing level with "all that's really going on." After introducing the circularity, current research into the nature of the self are examined. Materialist approaches that deconstruct the self by breaking down the single agent into a society of interacting agents are discussed. It is asserted that a naive, introspectionist view of the self prevails among most researchers in cognitive science (and in society at large) and that this prevents the realization of the lack of a unitary self at the heart of experience. After arguing that a schism between cognitive science and experience is unavoidable if the theoretical discovery (via cognitive science) of a decentered self is not taken seriously, a bridge to the Buddhist tradition of no-self is constructed. An introduction to the Buddhist theories of no-self is presented. The important aspect of Buddhism, namely, a lack of an inviolate level of "all that's really going on", is highlighted. The way in which Buddhism prevents an inevitable slide from groundlessness into nihilism is discussed. Finally, it is stressed that the self-consistent path (taken from the fundamental circularity to groundlessness) was laid down by exploring cognitive science from within and not by grafting Buddhism from without.

We comment on the authors' criticism of cognitivism and emergence in Section 2.1. The enactive approach is presented in Section 2.2 followed by comments. After issuing a caveat that the Buddhist section of the book may well be impossible to summarize without doing injustice to the book and the authors (and maybe to all sentient beings), we present the materialist and Buddhist deconstruction of the self along with comparisons in Section 2.3. After taking the message of the enactive approach to heart, different approaches to cognitive science are presented as different ways of reviewing the book in Section 3. Overall comments in Section 4 wrap up the review.

2 Commentary

2.1 Cognitivism and Connectionism

The authors begin with a criticism of the AI theme of representation. Their attack on representation is simplistic. However, they argue that their intent is to highlight the philosophical underpinnings of AI.

> Our point is not to caricature a sophisticated research program but simply to render explicit some tacit epistemological assumptions in as clear a fashion as possible. Thus although everyone agrees that representation is a complex process, it

> is nonetheless conceived to be one of recovering or reconstructing extrinsic, independent environmental features. Thus in vision research, for example, one speaks of "recovering shape from shading" or "color from brightness." Here the latter features are considered to be extrinsic properties of the environment that provide information needed to recover "higher-order" properties of the visual scene, such as shape and color. The basic idea of a world with pregiven features remains. [page 136]

The authors are critical of the notion of recovery of world features especially when the world features are taken to be pre-given, external and independent of the cognitive system. Here, they seem to be critical of external objectivism (as opposed to, say, neuroscience which is a form of internal objectivism). However, they do not frame their criticisms in terms of external objectivism's lack of response to *why* questions (instead of *what* and *how* questions). External objectivism is attacked not because of a failure to address questions of meaning but due to its lack of concern with the cognitive systems' embodiment and situatedness in the real world. We'll have more to say later on these issues.

Their summary of cognitivism is as follows:

> *Question 1*: What is cognition?
>
> *Answer*: Information processing as symbolic computation—rule-based manipulation of symbols.
>
> *Question 2*: How does it work?
>
> *Answer*: Through any device that can support and manipulate discrete functional elements—the symbols. The system interacts only with the form of the symbols (their physical attributes), not their meaning.
>
> *Question 3*: How do I know when a cognitive system is functioning adequately?

> *Answer*: When the symbols appropriately represent some aspect of the real world, and the information processing leads to a successful solution of the problem given to the system. [page 42–43]

When the focus is shifted to emergence (connectionism), their main criticism regarding representation remains intact. They argue that connectionism attacks the notion of cognition as symbol manipulation and offers an alternative view of cognition as emergent activity of a network. According to the authors, connectionism has nothing to say about representation per se. Instead, they claim that knowledge representation is replaced by rules for manipulating network connections (i.e. supervised learning) with the goal of optimally encoding environmental regularity.

> Thus even when the very ideas of representation and information processing change considerably, as they do in the study of connectionist networks, self-organization, and emergent properties, some form of the realist assumption remains. In cognitivism, the realism is at least explicit and defended; in the emergence approach, however, it often becomes simply tacit and unquestioned. [page 133]

When supervised learning is employed, the network is usually unable to find an unbiased representation of the incoming patterns, without increasing the error variance. Based on this observation, Stuart Geman *et al.* [1] argue that fundamental challenges in neural modeling are still about finding the right representation rather than learning. In [1], in the context of a discussion on the limitations of supervised learning:

> These limitations are well known, and well understood in terms of what we will call the bias/variance dilemma.
>
> The essence of the dilemma lies in the fact that estimation error can be decomposed into two components, known as bias and variance; wheras incorrect models

> lead to high bias, truly model-free inference suffers from high variance. Thus model-free (*tabula rasa*) approaches to complex inference tasks are slow to "converge," in the sense that large training samples are required to achieve acceptable performance. This is the effect of high variance, and is a consequence of the large number of parameters, indeed infinite number in truly model-free inference, that need to be estimated. . . . The only way to control the variance in complex inference problems is to use model-based estimation. However, and this is the other face of the dilemma, model-based inference is bias prone: proper models are hard to identify for these more complex (and interesting) inference problems, and any model-based scheme is likely to be incorrect for the task at hand, that is, highly biased.

After arguing that connectionism is silent on the issue of representation, the authors allow that a combination of cognitivism and connectionism would be useful and go so far as to state that a "mixed-mode endeavor" might dominate future cognitive science.

> The emergence view is, however, both in its early phase of the study of self-organizing systems and in its present connectionist form, is open to encompassing a greater variety of cognitive domains. An inclusive or mixed mode seems, therefore, a natural strategy to pursue. A fruitful link between a less orthodox cognitivism and the emergence view, where symbolic regularities emerge from parallel distributed processes, is a concrete possibility, especially in AI with its predominantly engineering, pragmatic orientation. This complementary endeavor will undoubtedly produce visible results and might become the dominant trend for many years to come in cognitive science. [page 103]

This "fruitful link" has been pointed out by many connectionists [2, 3, 4]. The main theme in [3] is the effective combination of powerful symbolic representations and effective learning procedures.

The authors summarize the connectionist approach as follows:

> *Question 1*: What is cognition?
>
> *Answer*: The emergence of global states in a network of simple components.
>
> *Question 2*: How does it work?
>
> *Answer*: Through local rules for individual operation and rules for changes in the connectivity among the elements.
>
> *Question 3*: How do I know when a cognitive system is functioning adequately?
>
> *Answer*: When the emergent properties (and resulting structure) can be seen to correspond to a specific cognitive activity—a successful solution to a required task. [page 99]

Unfortunately, in their criticism of connectionism, they do not clearly distinguish between supervised and unsupervised learning. In supervised learning, the criterion of an optimal fit between system and environment holds – in unsupervised learning, it does not (this distinction is merely stated on page 92). Supervised learning has more in common with well established principles of nonlinear regression and less with self-organization. They do point out that self-organization is common to connectionism *and* the enactive approach but a clearer distinction between the different learning paradigms would have served them well. If the authors had made a clear three-way distinction between representation, supervised learning and self-organization, the road to enactive cognitive science might have been less bumpy. As Christine Skarda notes [5]:

> It is misleading to identify, as Varela does, connectionism with self-organizing, *emergentist* systems, and to say that *all* connectionist systems are wedded to the representations of traditional cognitivism. Some connectionist models are self-organizing, but others are not. All connectionist systems use distributed, highly parallel processing, but that is not the same thing as being self-organizing. PDP systems are susceptible to Varela's attack on representations, self-organized systems are not. I believe that Varela's distinction between emergent and enactive systems is ultimately intended to capture the same fundamental distinction, but it is a mistake to equate emergent systems with connectionism as a whole and set all connectionist systems against the enactive approach. This dichotomy is a false one.

Self-organization as unsupervised learning is an important theme in connectionism but self-organization as emergent, viable, structured coupling between organism and environment has a stronger resonance. A useful tool in connectionism becomes the heart and soul of enactive cognitive science.

2.2 The Enactive approach

Before the introduction of their enactive approach, the authors discuss color vision in detail. Color vision, they argue, offers a domain that is illustrative of the enactive approach. In [6], two of the authors have presented their case in greater detail than in the book. Their position, briefly stated is as follows:

> Different explanations of color vision favor different philosophical positions. Computational vision is more compatible with objectivism (the color is in the object), psychophysics and neurophysiology with subjectivism (the color is in the head).

> Comparative research suggests that an explanation of color must be both experientialist (unlike objectivism) and ecological (unlike subjectivism).

By experientialist, the authors do not mean a retreat into ineffable qualia. Instead, "experientialist" should be taken to mean that color emerges from the interaction of a color vision processing system and an environment. Also, by ecological, the authors wish to take into account the active way in which animals shape their environments.

First, we examine the authors' criticism of external objectivism and then their criticism of neurophysiological subjectivism. They attack the notion that a completely objective account can be given for color. This is done by pointing out that the opponent nature of certain hues (red/green and yellow/blue) are not to be found in surface reflectances.

> Surface reflectances can be classified according to whether they reflect more or less light in the short-, middle-, and/or long-wave regions of the spectrum, but they cannot be classified as standing in opponent relations to other reflectances. Nor can these properties of uniqueness, binariness, and opponency be found in the structure of light. For these reasons, the properties that specify what colors are simply have no nonexperiential, physical counterparts.

They go on to point out that color is also a perceived attribute of the sky, dreams and afterimages. But the most important argument is the fact that color vision is also used in segmentation.

> Finally, there is a hidden, but much deeper problem with the objectivist view of color vision: the objectivist simply assumes that surface reflectances are to be found in some pregiven world that is independent of our perceptual and cognitive capacities. But how are we to specify what counts as a surface? How are we to

> specify its edges, boundaries, texture, and orientation, if not in relation to some perceiver for whom these distinctions are relevant? [pages 166–167]

After an attack on objectivism, they turn their criticism to neurophysiological subjectivism – the theory that colors are in the head. According to the authors, subjectivism neglects the role animals play in shaping their environment. For example, in discussing bee color vision:

> This coevolution implies not only that bee color vision is sensitive to ultraviolet because it is advantageous for bees to detect flowers that have ultraviolet reflectances, but also that flowers have ultraviolet reflectances because it is advantageous for them to be seen by bees. Thus the evolution of bee color vision did not simply provide the bee with a practical knowledge of its environment; it also contributed to the very determination of that environment.

The authors then point out the significance of their investigation of color:

> We can now appreciate, then, how color provides a paradigm of a cognitive domain that is neither pregiven nor represented but rather experiential and enacted. It is very important to note that just because color is not pregiven does not mean that it does not exhibit universals or that it cannot yield to rigorous analysis by the various branches of science. [page 171]

And later, they attack the notion that the varieties of color vision can be explained by natural selection. Their position is worth quoting in full:

> When we last left this cognitive domain [color], we had seen that there are different, incommensurable "color spaces": some require only two dimensions for their description (dichromacy), some require three (trichromacy), others require

> four (tetrachromacy), perhaps even five (pentachromacy). Each of these different kinds of color space is enacted or brought forth through a specific history of structural coupling.
>
> One of our motives in this chapter has been to show how such unique histories of coupling can be understood from the vantage point of evolution. To this end, we have provided a critique of the adaptationist view of evolution as a process of (more-or-less) progressive fitness, and we have articulated an alternative view of evolution as natural drift. We claim, then, that these unique histories of coupling, which enact incommensurable kinds of color space, should not be explained as optimal adaptations to different regularities in the world. Instead, they should be explained as the result of different histories of natural drift. [page 201]

Their position is that color is enacted via structural coupling between species and environment. They claim that the objectivist ignores experiential factors and that the subjectivist ignores ecological factors. Finally, they insist that the evolution of color vision systems cannot be explained through a process of optimization between a species and its environment. The final claim forestalls a "retreat into natural selection." Daniel Dennett offers a good assessment of their position in [7]:

> But what, then, are *colors*? Doesn't this imply that only a circular definition of color is possible? Yes, but as the authors say, "one should not be put off by the circularity" (sect. 3, para 16). The real bogey is the fear that if we cannot give a foundational, objective definition – either in the form of an external objective definition (e.g., Hilbert 1987)[2] or in the form of an equally objective internal or neurophysiological definition – we will be stuck with "intrinsic, ineffable, unknow-

[2]This reference corresponds to [8]

> able" qualia. The true value of the enactive view is that it explains why neither of these sorts of objective account is necessary to avoid the extreme subjectivism of Nagel (1974)[3]. Both these varieties of reduction are embarrassed by the specter of having to deal with all the counterexamples by brute force enumeration of exceptions – a telltale sign, presumably, that they have failed to capture the "essence" of color. An enactive account, however, can explain, objectively and scientifically, everything that needs explaining – including the fact that no compact, noncircular definition of color is possible – with no leftovers conceded to mystery.

The main points can be summarized. Color offers a case study in cognitive science wherein an objectivist or subjectivist reduction is not possible. Yet, it is amenable to scientific investigation. Also, color shows that we must eschew the notion of optimal fitness of a species to its environment since the plethora of different color systems (trichromatic, tetrachromatic etc.) cannot be explained by natural selection. Instead, viable structural coupling between a species and its environment is emphasized.

> If this coupling were to be optimal, the interactions of the system would have to be (more or less) prescribed. For coupling to be viable, however, the perceptually guided action of the system must simply facilitate the continuing integrity of the system (ontogeny) and/or its lineage (phylogeny). Thus once again we have a logic that is proscriptive rather than prescriptive: any action undertaken by the system is permitted as long as it[4] does not violate the constraint of having to maintain the integrity of the system and/or its lineage. [page 205]

---

[3]This reference corresponds to [9]

[4]I have taken the liberty to make a minor correction to the text as it appears in the book. The line in the book reads ". . . as long as it is does not violate . . ." I hope I have not violated the constraint of having to maintain the integrity of the book and/or the previous works by the authors.

The enactive approach can now be summarized:

> *Question 1*: What is cognition?
>
> *Answer*: Enaction: A history of structural coupling that brings forth a world.
>
> *Question 2*: How does it work?
>
> *Answer*: Through a network consisting of multiple levels of interconnected, sensorimotor subnetworks.
>
> *Question 3*: How do I know when a cognitive system is functioning adequately?
>
> *Answer*: When it becomes part of an ongoing existing world (as the young of every species do) or shapes a new one (as happens in evolutionary history). [pages 206–207]

Figure 1 (in [10]) and Table 1 (in [11]) lay out the enactivist manifesto. The circular relationship between representation and action is very important to the enactive approach and is illustrated in the figure. Several distinctions between cognitivism and enactivism are shown in the table. After our discussion of the overall nonobjectivist/nonsubjectivist theme of the enactive approach, the distinctions listed in the table should be self-explanatory.

The polar map in Figure 2 shows the authors' classification of researchers in cognitive science. A major flaw in the book is the seemingly arbitrary classification of certain researchers as enactive researchers. Two examples are John Holland and Stephen Grossberg. Other than the mention of Holland in the polar map, there's no discussion of his work. Why does Holland qualify? Holland's genetic algorithms do not perform category formation. Instead, as pointed out by Stephen Smoliar [12], they work with existing categories and find a good representation. Why is Grossberg's approach an enactive one? There's no sensorimotor coupling – Grossberg is interested in adaptive pattern recognition and not in adaptive behavior in the real world.

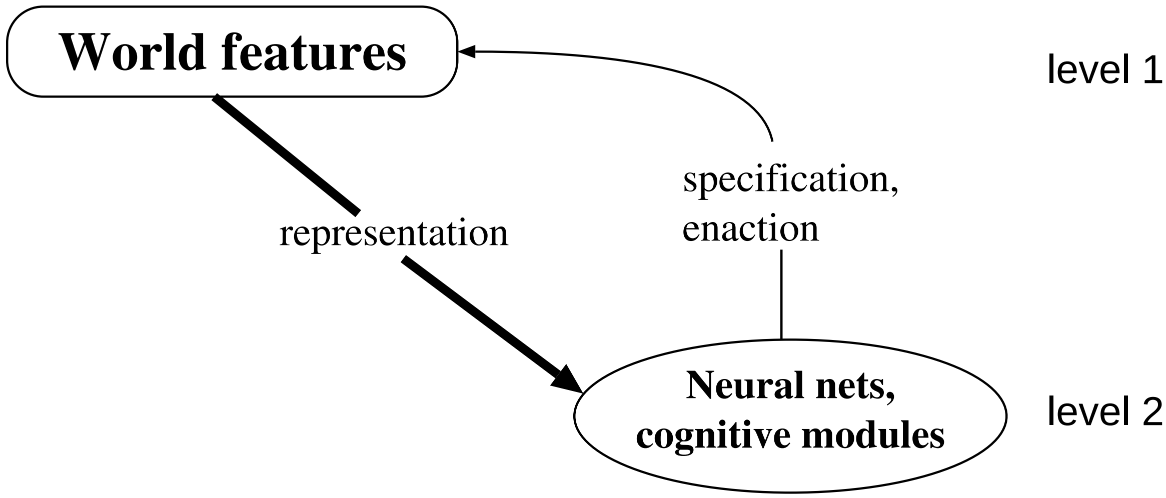

Figure 1: Representation and enaction of world features and cognitive modules

Table 1: From cognitivism towards enactivism

From:	Towards:
task-specific	creative
problem solving	problem definition
abstract, symbolic	history, body bound
universal	context sensitive
centralized	distributed
sequential, hierarchical	parallel
world pre-given	world brought forth
representation	effective action
implementation by design	implementation by evolutionary strategies
abstract	embodied

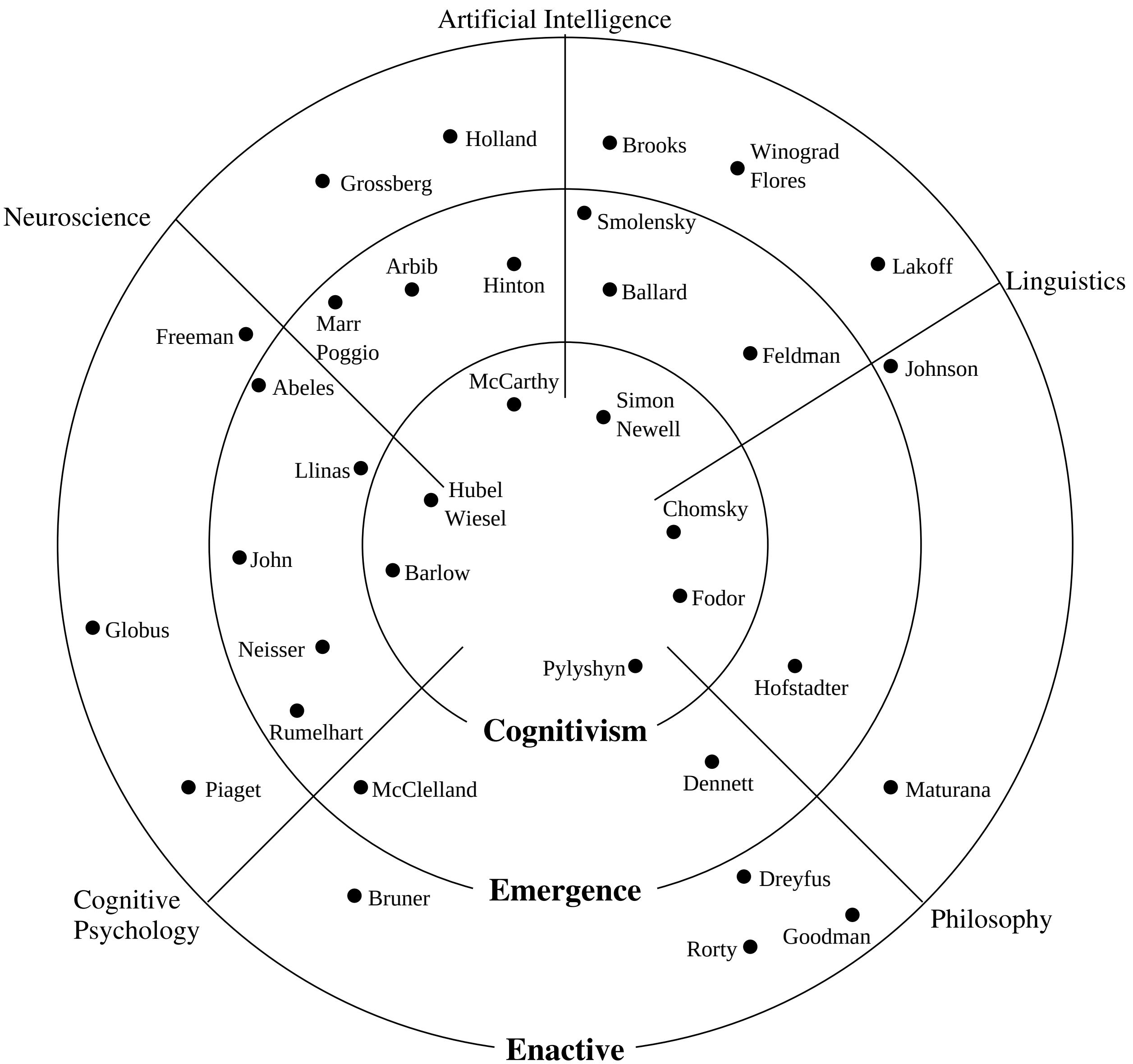

Figure 2: A conceptual chart of the cognitive sciences today in the form of a polar map, with the contributing disciplines in the angular dimensions and different approaches in the radial axis. [page 7]

Also, Marvin Minsky is missing. The authors draw on Minsky's work (especially in the context of consciousness) but either Minsky couldn't be pigeonholed (into the polar map) or the omission was an oversight.

In particular, the authors stress that Rodney Brooks' program [13] is fully compatible with the concerns of the enactive approach. Brooks and the authors share several common themes. The key aspects of Brooks' work are as follows:

> **Situatedness**: The world is its own best model. The idea here is to let a mobile robot keep referring to its sensors rather than to an explicit world model.
>
> **Embodiment**: The world grounds regress. Regress refers to the possibility of self-delusion when there is no feedback from the world. When an embodied robot validates its results in the world, there is less scope for self-delusion.
>
> **Intelligence**: Intelligence is determined by the dynamics of interaction with the world. Intelligence has less to do with abstractions and more to do with adaptability in a dynamic environment. The theme here is "Elephants don't play chess."
>
> **Emergence**: Intelligence is in the eye of the observer. Intelligence emerges from the interaction of the components of the system. There is no homunculus.[5]

All four themes have correlates in the enactive program. Since the enactive approach would have the cognitive systems bring forth a world via structural coupling, it can be said to go further than Brooks' program. The authors (deliberately?) do not clearly distinguish between the different time scales of evolution and development, and as we have seen, the active feedback from the species to its environment is an important theme.

It's obvious from the authors' classification of researchers that they were unaware of the emerging, enactive orientation in computational vision. For example, Dana Ballard in [14]:

[5] We'll have more to say about this topic in the next section

> However, a paradigm that we term *animate vision* argues that vision is more readily understood in the context of the visual behaviors that the system is engaged in, and that these behaviors may not require elaborate categorical representations of the 3-D world.

and Ruzena Bajcsy [15]:

> Hence the problem of Active Sensing can be stated as a problem of controlling strategies applied to the data acquisition process which will depend on the current state of the data interpretation and the goal or the task of the process.

Bajcsy stresses active control of data gathering in solving object recognition. In her formulation and to a large extent in Ballard as well, the emphasis is on foveated vision with active gaze, vergence and focusing mechanisms. It would have been interesting if the authors had related the works of these researchers with Brooks, especially since the former are interested in solving problems in visual perception and not just in building robots. When one builds a robot in the real world, the robot has to be viable – when one solves problems, optimal solutions are relevant if efficiently computable. By not stressing the middle ground between these approaches, the authors may alienate researchers of a more problem solving orientation. Also, I am puzzled by their treatment of Paul Smolensky's harmony theory:

> Consider as an example Paul Smolensky's harmony theory. Smolensky's paradigm of subsymbolic computation is generally compatible with the concerns of the enactive program. The remaining point of difference consists in Smolensky's evaluation of his models by reference to an unviolated level of environmental reality. . . . The enactive program . . . would require that we eschew any form of optimal fitness by taking this kind of cognitive system into a situation where en-

dogenous and exogenous features are mutually definitory over a prolonged history that requires only a viable coupling. [page 212]

This is like saying that backpropogation is similar to self-organization except for the minor point that backprop descends on a pre-specified error surface. Essentially, they've neglected to tell us how optimal fitness is to be replaced with viable coupling. Given the similarity of harmony theory to the Boltzmann machine, it's difficult to see how viable coupling can be substituted for optimal fitness without losing the essence of harmony theory. This is because the learning rules (in the Boltzmann machine) are derived by minimizing the Kullback distance [16] between the probability distributions of the fixed (inputs and outputs fixed) and free phases (inputs alone fixed). If descent on the error surface is removed, what is the substitute?

In sum, there are two principal problems with the authors' transition from connectionism to enactivism. First, they do not properly distinguish between supervised and unsupervised learning and second, there is too sharp a distinction between problem solving and viable coupling. The first problem leads them to lump all emergentist approaches together. A proper distinction would have allowed them to find elements in connectionism that supported the enactive perspective. Furthermore, the principal difference between self-organization as unsupervised learning, and self-organization as enaction, is the feedback from the actuators to the sensors via action in the real world. If they had expanded their treatment of self-organization to include a spectrum of approaches with no feedback at one end of the spectrum (exemplified by clustering) and enaction at the other, this would have enabled them to have a spectrum of enactive cognitive science with problem solving at one end (with optimal solutions) and viability at the other. This mistake is even more egregious when the inclusion of Grossberg in the polar map (lauded with full enactive honors) is taken into consideration.

2.3 The nature of the self

In this section, the second major theme in the book – the nature of the self – is examined. As mentioned in Section 1, the fabric of materialism is merely stretched by the enactive approach but the discussion on the self rips it apart. Being of good material, however, it doesn't get torn to shreds – it may be possible to form a patchwork quilt out of the torn fragments. Colorful metaphors aside, we must insist on a suspension of disbelief, in order to follow the argument made by the authors on the nature of the self. The presentation of the authors' position may seem somewhat schizophrenic in places. This is because I don't make a clear distinction between their argument and my own opinion. What follows then is a presentation of their position with little or no commentary.

Briefly, the authors examine materialist and Buddhist views of self but it's the similarities between the two positions that are highlighted. There are inherent difficulties in presenting the views of a completely different culture and by and large the authors accomplish their goal of presenting the Buddhist case. Unfortunately, presenting a précis of their position is a tall order.

In our discussion of color in the previous section, we saw that the authors were insistent on the lack of an objective and (neurophysiologically) subjective ground for color. Consequently, their enactive approach embraced a circular definition of color. The authors take this several steps further and claim that a fundamental circularity exists between the scientific study of cognition and our own cognitive structure. This circularity manifests itself when we try and associate behavior and experience with brain structure.

> The basic assumption, then, is that to every form of behavior and experience we can ascribe specific brain structures (however roughly). And, conversely, changes in brain structure manifest themselves in behavioral and experiential alterations. . . .

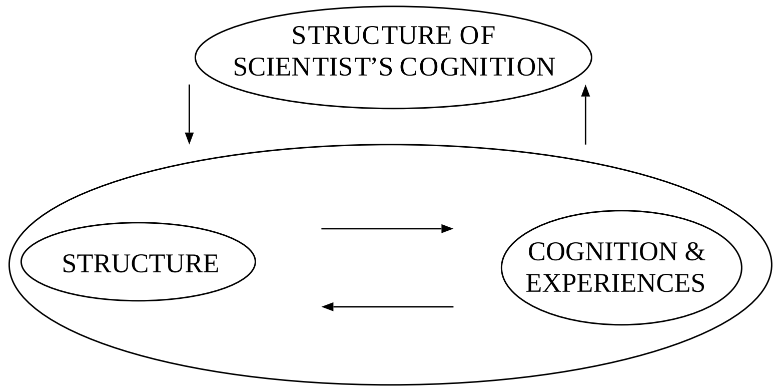

Figure 3: Interdependency of scientific description and our own cognitive structure. [page 10]

> Yet, upon reflection, we cannot avoid as a matter of consistency the logical implication that by this same view any scientific description, either of biological or mental phenomena, must itself be a product of the structure of our own cognitive system. [page 11]

When behavior and experience are associated with brain structures, the former are usually seen as phenomenological and derivative while the latter are seen as empiricial and causal. The authors, by introducing the fundamental circularity, are trying to forestall the characterization of behavior and experience as merely derivative. They claim that without the fundamental circularity, we would identify the information processing level of the brain with all that is really going on and everything else as useful fiction. With the fundamental circularity, the collective behavior of a scientific community could also be seen as "useful fiction." This is not a handwaving argument – the authors cannot create a bridge to Buddhism without insisting on the fundamental circularity. The status of ordinary experience gets strengthened by bringing our scientific endeavor into circulation as behavior and experience. Also, by bringing the scientific endeavor into circulation, the authors do not allow mechanistic explanations of behavior from becoming an inviolate level of "all that's really going on."

After insisting on this circular relationship between the phenomenological and empirical levels, we are in a position to examine the Buddhist view of phenomena. The following quote by the Dalai Lama [17] illustrates the view that Buddhists have of ordinary phenomena:

> So we do not say that phenomena exist by the force of conceptual designation just because we are enamored of conceptual designation, or simply like the whole idea. Rather, when we look at a phenomenon, a pen, for example, on one hand we investigate whether it exists from its own side, that is, independently of being cognized. Upon such analysis, we find that no such entity exists from its own side. If you investigate the pen, you investigate its shape, its color, its components, and you take each of these apart, you don't come across something that is the very pen itself. It doesn't satisfy that mode of analysis. But having come to the conclusion that there is no pen-in-itself to be found, then you can't simply say there is no pen. When you pick up the pen and write something, it performs functions that can harm you or benefit you. Any phenomenon that can harm or benefit you cannot simply be written off as nonexistent. But now you are in a quandary. You can't say it is nonexistent, yet at the same time you have investigated it and not found the thing-in-itself. So the question becomes, How does it exist? It exists by the force of conceptual designation. It is not that you like the idea so much, but rather that you have no alternative.

The fundamental circularity was introduced to keep at bay a materialist reduction of phenomena to an information processing level. The most important point to keep in mind when examining Buddhist views is that Buddhism does not have an inviolate level of "all that's really going on." For example, in the previous quote, it's clear that the the pen is just a collective abstraction for the Dalai Lama. He doesn't insist that there's something else that gives rise to the pen and its properties. All he wants to do is identify the pen as an abstraction.

The sense in which we should take the Dalai Lama's point is that there is no independently existing pen. Rather, it's existence is context dependent. This is a precise point and is the basis of everything else that follows. We might be willing to grant this point but because it seems superfluous. However, the most interesting aspect of Buddhism is the extrapolation of this sort of context-dependent existence all the way: this extrapolation includes all objects, phenomena and finally the self. Once again, the Dalai Lama [17]:

> There are many of these kinds of phenomena that are mental constructs, that is, they are abstract or merely designated. Generality falls into this same category and applies to the kind of innate notion of self that you have, not designated on the basis of particular instances but rather on the basis of the continuities. . . .

The ease with which the Dalai Lama switches from external phenomena to the self might seem disorienting at first. However, the internal logic is straightforward. We are asked to find anything that can exist all by itself. Not finding anything externally, we are then asked to find anything that corresponds to a fixed and independently existing self. All we seem to find is pure process and flux. For the Buddhists, the self is an abstraction based on a perceived continuity of experience and behavior. However, as opposed to a scientific reduction of the self to a material level, there is no attempt made to find an absolute level from which self and phenomena can be said to spring forth. The authors reiterate the Buddhist position on the self:

> We found dispositions, volitions, motivations—in short, all those things that make up our personality and emotional sense of self. We also found all the various forms in which we can be aware—awareness of seeing and hearing, smelling, tasting, touching, even awareness of our own thought processes. So the only thing we didn't find was a truly existing self or ego. But notice that we did find experience.

> Indeed, we entered the very eye of the storm of experience, we just simply could discern there no self, no "I." [page 79]

Again, there's a specific sense in the way things are said to be found. According to the authors, we can discern several ongoing processes – streams of thought, experience etc., but we cannot discern anything corresponding to a fixed entity. In the storm of activity, no concrete, absolute, independently existing self can be found.

Now, most of us who are not dualists would probably agree that the self does not exist independently and is some kind of emergent. Recently, Dennett [18] has developed such a model of the self – the Multiple Drafts model – wherein, the self is carefully deconstructed:

> It all depends on how far away you are. The closer you get, the more the disunity, multiplicity, and competitiveness stand out as important. The chief source of the myth of the Cartesian Theater[6], after all, is the lazy extrapolation of the intentional stance[7] *all the way in*. Treating a complex, moving entity as a single-minded agent is a magnificient way of seeing pattern in all the activity; the tactic comes naturally to us, and is probably even genetically favored as a way of perceiving and thinking. But when we aspire to a science of the mind, we must learn to restrain and redirect those habits of thought, breaking the single-minded agent down into miniagents and microagents (with no single Boss). Then we can see that many of the *apparent* phenomena of conscious experience are misdescribed by the traditional, unitary tactic.

[6]The Cartesian Theater refers to models of consciousness that postulate a spatially located (presumably in the brain) "seat of the soul."

[7]The intentional stance refers to a specific attitude adopted by a cognitive scientist; an attitude where the subject of investigation (whether a zombie or a human) is to treated as a rational agent "who harbors beliefs and desires."

Now, as pointed out earlier, the biggest difference between a Buddhist and materialist deconstruction is the sense in which one level is taken as absolute. Once again, Dennett:

> In some regards, you could say that my theory identifies conscious experiences with information-bearing events in the brain – since *that's all that's going on*, . . . [emphasis mine]

We can now appreciate an overarching difference between a materialist and a Buddhist description of the self. The materialist description treats the self as it would any observable phenomenon and reduces it to "real" information processing activities of the brain. The Buddhists on the other hand stop at a deconstruction of the self and do not try and reduce the self to a more fundamental level.

What happens when we take something like the Multiple Drafts model seriously? The following comments by Drew McDermott in [19] are a good example of some of the dilemmas that arise. After approving wholeheartedly, he writes:

> The dethroning of the self brings us to the third objection, ethics, which is the most crucial. Perhaps the consequences of believing D & K's[8] theory are so awful that even if it's true we should suppress it. It has always seemed axiomatic that if I care about nothing else in the world, I can care[9] about "the dear self." If the self is just a construct, then who cares whether it cares about itself?

According to the authors, this feeling of gloom generated by a materialist reduction of the self points to our inability to transform the scientific finding via experience. It's because we habitually go through life with a naive, introspectionist view that we cannot see that there is

[8] D & K refers to Dennett and Kinsbourne.

[9] In the original *Behavioral and Brain Sciences* commentary, the word "care" is missing. I've taken the liberty of adding the word after checking with the author.

no unitary self at the heart of experience. They quote the Tibetan teacher Tsultrim Gyatso to lend support to their claim:

> To have any meaning such a self has to be lasting, for if it perished every moment one would not be so concerned about what was going to happen to it the next moment; it would not be one's "self" anymore. Again it has to be single. If one had no separate identity why should one worry about what happened to one's "self" any more than one worried about anyone else's? It has to be independent or there would be no sense in saying "I did this" or "I have that." If one had no independent existence there would be no-one to claim the actions and experiences as its own . . . We all act as if we had lasting, separate, and independent selves that it is our constant preoccupation to protect and foster. . . . The meditator does not speculate about this "self." He does not have theories about whether it does or does not exist. Instead he trains himself to watch . . . how his mind clings to the idea of self and "mine" and how all his sufferings arise from this attachment. [pages 62–63]

The authors expand on this relationship between cognitive science and human experience. They argue that unless the discovery of the decentered self is taken seriously, we would be operating in a schizophrenic state; denying the existence of a self in cognitive science while affirming it in experience. Unless we examine experience with a method, our everyday conviction regarding the self is not likely to change. According to the authors, the discovery of the lack of a self in cognitive science offers a perfect bridge to the Buddhist tradition of no-self. They point out that the relative cultural unfamiliarity with the Buddhist nonfoundationist doctrines is no longer a serious obstacle due to the presence of a living Buddhist tradition in the West. They discuss the Buddhist mindfulness/awareness meditation tradition and claim that it allows us to experientially discover the lack of a unitary self. They first take pains to say what mindfulness/awareness is not:

> (1) a state of concentration in which consciousness is focused on only one object; (2) a state of relaxation that is psychologically and medically beneficial; (3) a dissociated state in which trance phenomena can occur; and (4) a mystical state in which higher realities or religious objects are experienced. . . .
>
> Buddhist mindfulness/awareness practice is intended to be just the opposite of these. Its purpose is to become mindful, to experience what one's mind is doing as it does it, to be present with one's mind. [page 23]

In [20], a precise description of mindfulness/awareness is given. The mindfulness method is nondualistic since the objective is complete identification with the ongoing perception/action of the moment so that there is no watcher of experience. The term "awareness" is more mysterious.

> *Mindfulness* is that state in which mind is fully present with whatever action we are executing: placing a flower, wiping a teacup, washing the car, programming a computer. It is attention to detail, careful and almost deliberate. It is identifying fully with one's body, thoughts and actions so that there is nothing left over, no self-consciousness, no watcher, no split mind. It is not watching what we are doing but simply *being* fully what we are doing, thinking, and feeling in its smallest, most insignificant detail. It is therefore altogether different from the early 'introspectionist' methods of Wundt and others. One is not looking *for* something, one is not particularly interested in the content of what arises to consciousness as such; one is healing the split between mental contents and knowing those contents rather than deliberately promoting that split as introspection tends to do.

> *Awareness* is acknowledging the quality of sudden openness that comes in when we are fully present. It is a sudden glimpse, a sudden flash of freshness and wider perspective. We cannot discover where it comes from, we cannot hold onto it, and we cannot artificially recreate it. Because it comes from beyond our perceptual process of conceptual projection, this quality of openness brings with it a sense of inquisitiveness, of interest in the environment within which our actions and thoughts take place. From this glimpse, awareness expands to see the associations and causal connections to even the smallest thought or perception, and to unravel and reveal the details and causal processes of thought and perception itself.

We can now appreciate the degree of convergence between the materialist and Buddhist views of the self. Both negate the unitary self that is posited by most religious traditions and affirmed by most of us in our habitual patterns of perception and action. However, the main difference between the two views is that Buddhism emphatically does not rely on a fundamental level of "all that's really going on." This lack of an ultimate level is quite basic to all forms of Buddhism. The most interesting aspect of Buddhism is the ability to insist on groundlessness without degenerating into nihilism.

The path taken from the fundamental circularity to groundlessness can be summarized. i) The fundamental circularity is introduced in order to be able to see scientific activity as behavior and experience. (ii) The inner workings of present day scientific activity has resulted in a mechanistic deconstruction of behavior, experience and the self. (iii) Rather than exploit the circularity via paradox[10] or negation, the authors choose to explore a self-consistent path wherein the discovery of groundlessness of the self in cognitive science is accepted but only as a

[10] The circularity can be succinctly summarized as follows: (i) Cognitive science asserts that all experience and behavior are fiction and that the information processing level is real. (ii) Cognitive science is also a form of behavior and experience. From (i) and (ii) it follows that the veracity of cognitive science is open to question. Different resolutions can be found. (i) Paradox: Cognitive science is real. Behavior and experience

purely theoretical discovery. (iv) A bridge is created to a tradition that allows its practitioners to experientially discover the lack of a self.

> We have therefore chosen to follow Nishitani's[11] lead by building a bridge between cognitive science and mindfulness/awareness as a specific practice that embodies an open-ended approach to experience. Furthermore, since we cannot embody groundlessness in a scientific culture without reconceptualizing science itself as beyond the need of foundations, we have followed through the inner logic of research in cognitive science to develop the enactive approach. This approach should serve to demonstrate that a commitment to science need not include as a premise a commitment to objectivism or to subjectivism. [page 244]

Initially, I had decided to write a ponderous commentary on the authors' position along with detailed criticisms, points of agreement etc. Upon reflection, this seemed to be a rather dense way of going about doing it. Instead, I felt that a more productive conclusion would be to sketch under what circumstances different readers would profit from reading the Buddhist section of the book. To this end, I briefly sketch a few possible scenarios.

First, let me state categorically that if you are troubled by cognitive science's assertion

are ultimately fiction. Cognitive science is behavior and experience. (ii) Negation: Cognitive science, being behavior and experience is fiction. Therefore, cognitve science has no ontological status and has only pragmatic and utilitarian dimensions. (iii) Fuzziness: My head hurts thinking about this circularity. All this is just philosophical mumbo-jumbo. (iv) Self-consistency: Cognitive science asserts that behavior and experience are fiction. Cognitive science being behavior and experience is fiction. All is fiction, Buddha nature alone is real.

I am probably in danger of mis-representing the authors' position in describing a path from the circularity to the Buddhist tradition. However, I decided, in the interest of clarity, to proceed and sin boldly as St. Paul advises.

[11] Nishitani Keiji is a Japanese philosopher, trained in the Zen tradition of mindfulness/awareness and was also one of Heidegger's students.

of a lack of a unitary self and feel that this violates your (seemingly) unchanging center of awareness, then the Buddhist section of the book will be equally infuriating and challenging. On the other hand, if you take the self to be an emergent and are interested in models of the self and awareness etc., you will be almost entirely unpersuaded. Next, if you are worried about a possible schism between cognitive science and experience due to our inability to take a deconstruction of the self seriously, then this is the book for you. In a more philosophical vein, if you've been obsessed with the paradoxes of eternal flux and constancy, of process and archetype, this book is completely orthogonal to your interests. Finally, if you're merely interested in a good summary of Buddhism and its intersections with cognitive science, this is where to begin, since the authors are perfectly situated to do a good treatment of both topics.

3 An example of what the book is about

An example of the mindfulness/awareness approach to the examination of experience can be seen by a three-way review of the book: cognitivist, emergent and enactive. In what is to follow, the three positions have been caricatured. The cognitivist review is disembodied. The book is treated as an object in the "real" world with fixed ideas and properties that can be objectively described. The simplistic three-way distinction of approaches to cognitive science follow from this standpoint. The emergent review is describing the book as a deep object which has resulted from a historical process. Likewise, the reviewer (also the result of a long process) is reviewing the book with unavoidable bias. The enactive review is not a review of the book but a view of the book in the process of embodied reflection.

Cognitivist:

"*The Embodied Mind* classifies all research in cognitive science into three major categories; cognitivist, emergent and enactive. Cognitivism is described as "no computation without representation" and is criticized for holding to the view of a world with pregiven features that are subsequently represented by cognitive systems. . . ."

Emergent:

". . . In sum, there are two principal problems with the authors' transition from connectionism to enactivism. First, they do not properly distinguish between supervised and unsupervised learning and second, there is too sharp a distinction between problem solving and viable coupling. . . ."

Enactive:

. . . These guys don't seem to understand that all the power of cognitive science comes from assuming that the information processing level is where we can do science. "I" don't understand how they can accept all the findings of cognitive science and then turn around and claim that actually there was no ground to anything that we did. Won't this always result in not being able to take those findings seriously? Seems like the Buddhist stuff is an add on and doesn't really do anything. You have to accept a materialist reduction first and then appeal to Buddhism to ward off all the ethical problems and also to stop the fury of introspection that we indulge in. . . .

Note that the enactive review is a view of the thoughts of the reviewer in the process of reviewing the book. The cognitivist and emergent reviews are quotes taken from the Section 1 and the commentary in Section 2 respectively. However, the enactive review is "hot off the press" and does not occur in the previous sections. The decision was taken to review the book and then the thoughts were written down as they occurred. Consequently, the enactive review does not break the review down into first person or third person. Hence, it is possible to extract an objectivist or subjectivist review from the enactive review but not vice-versa.

4 Conclusion

In conclusion, there are two books here. The first book takes a sober look at the varieties of cognitive science and tries to point out tacit, unquestioned epistomological commitments lurking in the various frameworks. In particular, cognitivism is criticized for the lack of emphasis on the role played by cognitive agents in shaping a world and connectionism is criticized for not being clear regarding the role of representations and for not taking self-organization all the way into viable, structural coupling. The enactive approach is presented as a framework that allows for precisely this sort of self-organized interaction and mutual specification between sensorimotor networks and an environment. The second book is a murmur of protest regarding the materialist deconstruction of the self in cognitive science and an encouraging nod toward Buddhist deconstructions of the self. This is followed by a warning of an approaching schism between cognitive science and ordinary experience due to our inability to take these findings (of a decentered self) seriously, resulting in us becoming materialists on weekdays and naive, introspectionists on weekends.

References

[1] S. Geman, E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, 4, pp. 1–58, 1992.

[2] C. von der Malsburg and E. Bienenstock, "Statistical coding and short-term synaptic plasticity: A scheme for knowledge representation in the brain", In E. Bienenstock, F. Fogelman-Soulie, and G. Weisbuch, editors, *Disordered Systems and Biological Organization*, Springer-Verlag, Berlin, 1986.

[3] G. Hinton, editor, *Connectionist Symbol Processing*, The MIT Press, Cambridge, MA, 1990.

[4] E. Mjolsness, "Bayesian Inference on Visual Grammars by Neural Nets that Optimize", Technical Report YALEU/DCS/TR-854, Yale Computer Science Department, May 1991.

[5] Christine Skarda, "Perception, Connectionism and Cognitive Science", In Francisco J. Varela and Jean-Pierre Dupuy, editors, *Understanding Origins: Contemporary Views on the Origin of Life, Mind and Society*, Kluwer Academic Publishers, Boston, MA, 1992.

[6] E. Thompson, A. Palacios, and F. J. Varela, "Ways of coloring: Comparative color vision as a case study for cognitive science (includes commentaries)", *Behavioral and Brain Sciences*, 15(1), 1992.

[7] D. C. Dennett, "Hitting the nail on the head", *Behavioral and Brain Sciences*, 15(1), pp. 35, 1992, comment on E. Thompson, A. Palacios and F. J. Varela.

[8] D. R. Hilbert, "Color and color perception: A study in anthropocentric realism", 1987, Center for the Study of Language and Information, Stanford University.

[9] T. Nagel, "What is it like to be a bat?", In Ned Block, editor, *Readings in the philosophy of psychology*, volume 1, Harvard University Press, 1974/1980.

[10] Jean-Pierre Dupuy and Francisco J. Varela, "Understanding Origins: An Introduction", In Francisco J. Varela and Jean-Pierre Dupuy, editors, *Understanding Origins: Contemporary Views on the Origin of Life, Mind and Society*, Kluwer Academic Publishers, Boston, MA, 1992.

[11] F. J. Varela, "Whence Perceptual Meaning? A Cartography of Current Ideas", In Francisco J. Varela and Jean-Pierre Dupuy, editors, *Understanding Origins: Contemporary*

Views on the Origin of Life, Mind and Society, Kluwer Academic Publishers, Boston, MA, 1992.

[12] S. W. Smoliar, "Book Review of The Remembered Present: A Biological Theory of Consciousness by Gerald M. Edelman", *Artificial Intelligence*, 52(3), pp. 295–318, December 1991.

[13] R. Brooks, "Intelligence without Reason", *Artificial Intelligence*, 47(1-3), pp. 139–159, January 1991.

[14] D. H. Ballard, "Animate Vision", *Artificial Intelligence*, 48, pp. 57–86, 1991.

[15] R. Bajcsy, "Active Perception", *Proc. IEEE*, vol. 76(8), pp. 996–1005, August 1988.

[16] S. Kullback, *Information Theory and Statistics*, Dover Publications, New York, NY, 1968.

[17] Jeremy W. Hayward and Francisco J. Varela, editors, *Gentle Bridges: Conversations with the Dalai Lama on the Sciences of the Mind*, Shambala, Boston, MA, 1993.

[18] Daniel C. Dennett, *Consciousness Explained*, Little, Brown and Company, Boston, MA, 1991.

[19] D. McDermott, "Little me", *Behavioral and Brain Sciences*, 15(2), pp. 183, 1992, comment on D. C. Dennett and M. Kinsbourne.

[20] J. W. Hayward, *Shifting Worlds, Changing Minds: Where the Sciences and Buddhism Meet*, New Science Library, Shambala, Boston and London, 1987.