

From Cybernetics to Second-Order Cybernetics: A Comparative Analysis of Their Central Ideas

Tom Froese • University of Sussex, UK • t.froese@gmail.com

> Context • The enactive paradigm in the cognitive sciences is establishing itself as a strong and comprehensive alternative to the computationalist mainstream. However, its own particular historical roots have so far been largely ignored in the historical analyses of the cognitive sciences. **> Problem** • In order to properly assess the enactive paradigm's theoretical foundations in terms of their validity, novelty and potential future directions of development, it is essential for us to know more about the history of ideas that has led to the current state of affairs. **> Method** • The meaning of the disappearance of the field of cybernetics and the rise of second-order cybernetics is analyzed by taking a closer look at the work of representative figures for each of the phases – Rosenblueth, Wiener and Bigelow for the early wave of cybernetics, Ashby for its culmination, and von Foerster for the development of the second-order approach. **> Results** • It is argued that the disintegration of cybernetics eventually resulted in two distinct scientific traditions, one going from symbolic AI to modern cognitive science on the one hand, and the other leading from second-order cybernetics to the current enactive paradigm. **> Implications** • We can now understand that the extent to which the cognitive sciences have neglected their cybernetic parent is precisely the extent to which cybernetics had already carried the tendencies that would later find fuller expression in second-order cybernetics. **> Key Words** • W. Ross Ashby, Heinz von Foerster, constructivism, enactive cognitive science.

1. Introduction

This paper makes a specific contribution to our historical understanding of the origins of the enactive paradigm in the cognitive sciences. Its aim is to investigate the defining changes in thought that led from the original tradition of cybernetics in the 1940s and '50s to the establishment of second-order cybernetics in the '60s and '70s. The main focus will be less on providing a catalogue of objective dates and events, and more on finding a way of how best to make sense of this history from the perspective of more recent developments. Accordingly, this is not meant to be merely a scholarly exercise. It is intended to increase our understanding of the past, so that we may act with more awareness in the present toward shaping a desirable future.

So far, the bulk of historical research into the origins of the enactive paradigm has been limited to the goal of introducing its central ideas to a wider cognitive science audience, especially by relating them to the historical developments that took place *within* the cognitive sciences, and that

eventually led up to the publication of the initial manifesto of the enactive approach by Varela, Thompson and Rosch (1991). A very brief summary of this history of ideas is presented in the subsection below.

Here, we instead focus on the important historical developments that took place *outside* the cognitive sciences, and that eventually helped to give rise to the enactive paradigm. More specifically, we will analyze the culmination of the first generation of cybernetics, as represented by the work of W. Ross Ashby (Section 2), and the ensuing development of second-order cybernetics, as envisioned by Heinz von Foerster (Section 3). This transition will also enable us to make better sense of why the other offspring of cybernetics, computationalist cognitive science, has largely preferred to ignore its cybernetic parent (Section 4).

The implicit assumption that guides this particular interpretation is that it is more accurate to think of the enactive paradigm as deriving from a distinct tradition of science of cognition, which similarly originated in the first wave of cybernetics, but which, at the end of that era, formed a largely inde-

pendent tradition of thought that continued in the shadow of the mainstream cognitive sciences (cf. Varela 1996; Varela, Thompson & Rosch 1991: 37–40). We will highlight some of the potential directions for future research that are entailed by this assumption at the end of this paper (Section 5).

On the cognitive science origins of the enactive paradigm

It is possible to interpret the recent appearance of the enactive paradigm in the cognitive sciences as resulting from an inner development that was driven by the need to address shortcomings of the previous paradigms (Froese 2007). From this view, the historical sequence of the rise of computationalism (e.g., Fodor 1975), connectionism (e.g., McClelland, Rumelhart et al. 1986), embodied-embedded cognitive science (e.g., Clark 1997), and the enactive paradigm (Stewart, Gapenne & Di Paolo, in press) can be seen as governed by a specific explanatory need. In order to develop a more complete explanation of cognition, researchers had to appeal to an ever-expanding context of factors in order to explain the underlying

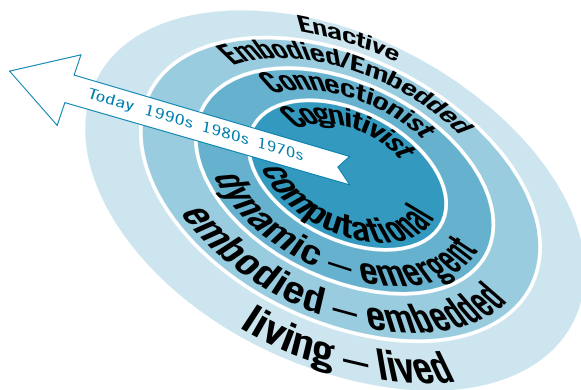


Figure 1: Illustration of the possibility of interpreting the history of the enactive paradigm in terms of events that happened *within* the cognitive sciences. The history of ideas appears to be shaped by a sequence of emerging paradigms. These changes were driven by the explanatory need to address previous shortcomings in explaining cognition by appealing to a more inclusive set of determining factors.

mechanisms: symbolic, dynamical, bodily, and living (cf. Thompson 2007: 3–16). For instance, it is now becoming more widely accepted that it is not sufficient to appeal to logical structures to explain cognition, but that biological realization and physical constraints play a constitutive role as well.

At the same time, this expansion of context has the effect of restricting the domain of validity of the previous paradigms by helping to delimit the restricted conditions under which their principles apply (cf. Figure 1).

Nevertheless, the possibility of interpreting the history of the cognitive sciences in terms of this more or less orderly sequence of paradigms should not blind us to the fact that the arrival of each subsequent stage also brought to fruition another part of the rich heritage that cybernetics had already bequeathed to the cognitive sciences, but that had lain mostly dormant until the conditions for its re-appropriation arose.

For instance, the central idea of understanding and building cognitive systems in terms of a variety of technological platforms was already an essential part of the cybernetics tradition. In addition, each paradigm’s particular scientific tool of choice, i.e., information processing for computationalism (e.g., Newell & Simon 1976), artificial neural networks for connectionism (e.g., McClelland, Rumelhart et al. 1986), and mobile

robots for embodied-embedded cognitive science (e.g., Brooks 1991), has its own cybernetic precursors. One only has to think of McCulloch & Pitts’ (1943) seminal paper, which became a manifesto for both symbolic and connectionist AI, and the robotic tortoises of Grey Walter (1950) to see the continuing influence of the early cybernetics tradition. What is striking, however, is how rarely the field of cybernetics in general has been mentioned in accounts of the cognitive sciences. We will explore some of the reasons for this neglect later on in this paper.

In addition, as we will see, it is no coincidence that some of the dynamical systems work that is going on within the enactive paradigm (e.g., Di Paolo 2003) is heavily indebted to the general systems theory of W. Ross Ashby. In fact, we will argue that Ashby can be seen as the starting point for a series of historical developments that took place *outside* the mainstream cognitive sciences, and that eventually contributed to the conception of the enactive paradigm in the cognitive sciences. It is the beginning of this “external” sequence of events that will be the focus of our analysis.

It should also be briefly mentioned that this history of ideas omits some significant precursors to this tradition of thinking, in particular David Hume’s (1739–1740) philosophy of life and mind, Immanuel Kant’s (1790) philosophy of the organism and

also Jakob von Uexküll’s theoretical biology (1928). The meaning of Hume’s philosophy from the perspective of the enactive paradigm is analyzed by Froese (2009). The relation of Kant’s approach to cybernetics and the autopoietic tradition is discussed by Froese & Ziemke (2009), and the relevance of von Uexküll’s approach to contemporary research is assessed by Ziemke & Sharkey (2001). Finally, for an extensive “first-person” history of the events leading up to the development of second-order cybernetics, the reader is referred to the dialogical autobiography of von Foerster & Bröcker (2002).

A philosophical primer

As a final point before we begin our analysis, some general philosophical orientation is in order. It is helpful to remember that all of the scientific traditions discussed in this paper responded to the same deep problem in modern thought in relation to which they emerged as systematic alternatives, though each in their own distinctive way. Thus, whatever their essential individual differences may be, they are at least *externally* unified in their shared opposition to the dualist Cartesian worldview, which had been the dominant paradigm of thought ever since it paved the way for the scientific revolution beginning in the 18th century. Descartes’ absolute rupture between mental and physical phenomena provided the necessary metaphysical protection for scientific activity to make progress in its quest for objectivity, while conveniently leaving the problem of subjectivity aside.

However, as is well known, the same philosophical move that freed the physical world for scientific exploration and technological domination has simultaneously made the very foundation of knowledge, the relationship between the observer and the observed, simply mysterious. Even today we are still faced by this “explanatory gap” (Levine 1983). As will become evident in what follows, each of the scientific traditions can be seen as a specific way of dealing with this fundamental mind-body problem.

The significance of trying to resolve this problem should not be underestimated. The “gap” is not a mere philosophical conundrum; it has subtle but important implications for how people believe they can, or cannot, act in the world.

2. Cybernetics and the Ashbyan crisis

After a long period of scholarly silence on the topic, it is now becoming increasingly recognized that the current information age in many respects owes its existence to one of the most influential yet enigmatic periods of thought in the first half of the 20th century: *cybernetics*. Beginning in the 1940s, the age-old idea of “mind as mechanism” was turned into a progressive scientific and technological research program, eventually culminating in the inception of the cognitive sciences in the '70s. By this point, the field of cybernetics had already begun to disappear, the general notion of “mechanism” had been replaced by the more specific concept of “computation,” and the computational theory of mind had established itself as the only game in town (cf. Chapters 6 & 7 in Boden 2006).

The history of the early American cybernetics tradition has recently been eloquently analyzed by Jean-Pierre Dupuy (2009), a study to which this paper owes much. Dupuy pays particular attention to the role of the Macy Conferences and to the contributions of some of the greatest minds of the twentieth century, including John von Neumann, Norbert Wiener, Warren McCulloch and Walter Pitts. While his interpretation of this tradition is informed from the perspective of more recent developments in the cognitive sciences, his historical analysis actually ends in the 1950s. At that point, according to Dupuy, the confrontation between the revolutionary work of the British cyberneticist W. Ross Ashby and the old guard of the Macy Conferences heralded the end of an era and the beginning of another: cognitive science.

This paper aims to shed some further light on this momentous event, in particular on the rise of the new generation of cybernetics that developed in the long shadow of the prestigious cognitive sciences: *second-order cybernetics*. In so doing we will provide a tentative answer to a central question posed by Dupuy, which, as Di Paolo (2001) points out, has not yet received a satisfactory answer: why was the field of cybernetics largely forgotten by the cognitive science mainstream to which it gave birth? The argument put forward in this paper is that cognitive science's cybernetic origins were

ignored precisely to the extent to which cybernetics already harbored the tendencies that only found their fuller expression within the second-order approach.

A science of the mindless mind

The main finding of recent historical investigations into the early cybernetic tradition is that the defining paradigm of thought essentially comes down to the *mechanization of the mind* (cf. Dupuy 2009; Boden 2006; Husbands, Holland & Wheeler 2008). What does this mean? It is clear that we are dealing with a specific attempt at naturalizing the mind by making it amenable for scientific study. More specifically, we can say that the cybernetic paradigm is a modern form of scientific materialism that posits feedback mechanisms, algorithms and non-linear dynamic systems as the foundational building blocks of cognitive systems, and hence of how our experiential reality shows up for us. Though this is not necessarily true of the work of all cyberneticists, its general philosophical framework consists in a monist metaphysical position in which mental activity is reduced to the operation of purely physical mechanisms¹.

For example, in an influential paper by Rosenblueth, Wiener & Bigelow (1943), we find that the necessary and sufficient conditions for what *appears* to be “purposeful” activity are defined entirely without reference to the conditions of an active subject. The notion of purpose and goal-directed behavior is *reduced* to the presence of feedback systems such that there is no longer any essential difference between the behavior of a guided missile system and a human agent. In this sense, cybernetics can be seen as a form of *eliminative materialism* in which the reality of the subjective aspect of existence is simply denied as nothing but mere appearance.

1| Note that since physics describes the world in terms of mathematics, it seems that this kind of scientific materialism is actually a form of Platonic idealism: the essence of the physical world appears to consist in observer-independent ideas! However, since materialism remains blind to the subjective source of this reification, it is not a form of idealism in the standard philosophical sense of the term.

However, the intelligibility of the cybernetic tradition is therefore always threatened from within its own framework because of the blatant absurdity of a radical nihilism, i.e., by the self-refuting position that there are *nothing but* physical processes that create the mere “as-if” *appearance* of a subject acting for authentic reasons. In other words, what reason do I have to accept this view if its proposal denies the very efficacy of rationality? The philosopher, Hans Jonas, whose phenomenological work on the biological foundations of individuality has been extremely influential for the enactive paradigm (cf. Weber & Varela 2002), wrote in his critique of cybernetics:

“As part of the history of life's quest to know itself, materialistic biology, its arsenal newly strengthened by cybernetics, is an attempt to apprehend life by eliminating that which affords the possibility of the attempt itself – the authentic nature of awareness and purpose.” (Jonas 1966: 133–134)

At least in the beginning of the cybernetic tradition, it still seemed possible to retain some measure of self-consistency in the face of this paradox, namely by supposing that the physical mechanisms responsible for the mere illusion of subjective agency are *themselves* sufficiently intelligent. The appropriation of agent-level terminology for component-level descriptions, especially noticeable in relation to the success of such fundamental concepts as “information” and “communication” (e.g., Shannon & Weaver 1949), should therefore come as no surprise.

However, with the culmination of the British cybernetics movement in the work of W. Ross Ashby, the field underwent a radical transformation that forcefully showed the futility of a crude reductionism that hides behind a confusion of discourse. Ashby showed, for example, that behavior that appears to be intelligent to an external observer can potentially be accounted for by means of a simple mathematical system driven by *random* step functions (e.g., Ashby 1947, 1960). In other words, to explain the operation of this adaptive system there was no need to make use of any kind of agent-level concept at all! And to give weight to his arguments he not only demonstrated them

BOX 1: Representationalist computationalism

In response to the Ashbyan crisis, the information-based approach to cybernetics morphed into the field of symbolic artificial intelligence, as exemplified by the famous work of Newell & Simon (1976). Indeed, the effects of this fundamental shift in focus are still discernable today. Mainstream computationalist cognitive science, the modern inheritor of the information processing paradigm, has continued the same strategy of keeping the paradox of eliminative materialism at bay, namely by indiscriminately mixing distinct levels of discourse. Notions centred on computation, communication and information processing that normally only have validity on the personal or even social level (Hutchins 1995) are viewed as necessary in explanations of mechanisms that are operational in the sub-personal level.

Indeed, for mainstream cognitive science there is simply no longer any systematic distinction between person-level discourse and brain-level explanation. The idea of a little homunculus that perceives and directs the so-called “Cartesian theatre” in our brains remains a powerful illusion that allows the worst excesses of materialism to proceed relatively unchecked. Unfortunately, the existential need for this kind of discourse, especially when faced with the onslaught of an overbearing and reductionist science, is so great that it can even obscure the very possibility of meaningful alternatives.

Andy Clark, for example, a well-known philosopher of cognitive science who is in many respects a progressive and more moderate computationalist, also reacts with a sense of scepticism and bewilderment, somewhat akin to that of Ashby’s Macy Conference audience, when faced by a thoroughly dynamical approach to explaining cognition (cf. Clark 2001: 128–138). Of course, even the computationalists have always worked within a mechanistic framework as well, but the crux of the problem is that a radical dynamical approach exposes the concept of “inner mental representation” as a label borne out of convenience and convention rather than of scientific necessity. Then, as now, the radical dynamicists called the bluff of computationalism by attempting to frame the scientific investigation of cognition in a purely mathematical formalism that is hardly amenable to the comfortable confusion of standard folk psychological terminology.

with formal mathematics, but even built an actual device, the famous homeostat, as further proof of concept.

Of course, the idea that the teleology of purposeful, goal-directed behavior could be turned into the teleonomy, or “as-if” teleology, of feedback mechanisms had a similar style of approach. Nevertheless, there still appears to be an explicit *representation* of the target goal state that the system tries to reach. And this is the essential point: no such representation could be found in Ashby’s homeostat. No wonder that the first generation of cyberneticists present at the Macy Conferences reacted to Ashby’s work with such outrage and disbelief. Indeed, his talks turned out to be a dramatic encounter that can be interpreted as marking the end of that particular era (cf. Dupuy 2009: 148–155).

As is usually the case, those who question the validity of the preferred narrative and point out that the emperor is without clothes are rarely looked upon kindly by anyone. Thus, in a sense, Ashby’s role for the field of cybernetics, though he certainly did not see himself like this, can be interpreted akin to the role that is often attributed to Hume: a great thinker who radicalizes a prevalent framework to such an extreme that it unwittingly refutes itself because of its own patent absurdity.² To be sure, the

2| This is not meant to suggest that Hume’s work can be reduced to this role. On the contrary, I actually believe that Hume’s philosophy of life and mind has in many respects been misjudged and underappreciated. I argue elsewhere that he can be seen as a forerunner of the enactive paradigm (cf. Froese 2009).

scope of Ashby’s contributions goes well beyond this role, but it is nevertheless the case that the homeostat presented an important philosophical challenge to the cybernetic tradition. Researchers were now faced with two choices if they wanted to avoid undermining the rationality of their scientific project. They could either (i) introduce a personal level of explanation that is irreducibly distinct to its mechanistic underpinnings, or (ii) insist on the necessity of using personal-level concepts within the level of mechanistic explanation.

Accordingly, the once more-or-less unified field of cybernetics became irreparably divided after the work of Ashby. In Box 1 we briefly analyze what happened on one side of this historical split, namely the emergence of symbolic AI and the attendant computational theory of mind. This development, which I will call *representationalist computationalism*, became officially inaugurated as the field of cognitive science in the 1970s (Harnish 2002). It thereby provided to its proponents a fresh start from which to conveniently forget about the Ashbyan trauma inherent in the field’s tumultuous past.

However, the disintegration of the field of cybernetics that is traditionally associated with the Macy Conferences also provided the opportunity for a second generation of researchers to take the principles of cybernetics into uncharted territory. Most importantly, Ashby’s own work on adaptation and ultrastability (e.g., Ashby 1956; 1960) set the stage for the emergence of Maturana and Varela’s Santiago school of cognitive biology (e.g., Varela, Maturana & Uribe 1974). Moreover, Ashby’s presence at von Foerster’s “Biological Computer Laboratory” was surely influential for the later development of second-order cybernetics (cf. Foerster 1979), which we will analyze in the next section.

Both of these traditions can be seen as direct responses to the challenges deriving from Ashby’s general systems framework by systematically extending its foundational principles into new directions. In other words, they accepted Ashby’s homeostat as a proof of concept and adopted his dynamical framework, but then had to find ways to address some of its fundamental limitations. Two challenges deriving from Ashby’s work are particularly noteworthy in this regard.

Challenge 1: Can there be self-organization?

First, it is important to recall that amidst all the clamor and excitement about the supposed wonders of self-organization, Ashby stood up and dryly demonstrated the notion's principal impossibility (cf. Ashby 1962). In brief, if one accepts general systems theory's two major premises, namely that issues of materiality are simply irrelevant and that self-reference is illogical and must thus be excluded, then one is left in a position where the concept of self-organization simply makes no sense. It can still be used loosely to denote a situation in which some aspect of a system's organization dynamically changes some other parts of that organization, but in this case there is still merely a static organization at the level of the whole system.

Ashby's own solution to this problem is to explicitly incorporate the possibility of external noise into the "self-organizing" system in the form of the mathematically defined concept of a machine "breaking" (Ashby 1947). The demonstration of this concept can be seen, for example, in the principle of ultrastability and is represented by the random step functions of the homeostat. This kind of approach is also pursued by von Foerster who develops it into an "order from noise" principle by grounding it in the perspective of physics (e.g., Foerster 1960), and it continues to inspire research related to the enactive paradigm in robotics (e.g., Di Paolo 2003) and artificial life (Ikegami & Suzuki 2008).

Maturana, on the other hand, can be seen to respond to this particular challenge by grounding the possibility of self-reference in the material self-production inherent in the metabolic activity of biological systems, i.e., the "circular organization" of the living (e.g., Maturana 1970a). This approach also still retains the essence of the noise principle in the notion of "perturbation," which denotes the triggering influence of environmental conditions. Unfortunately, it is beyond the scope of this paper to take a closer look at Maturana's conception of the circular organization of the living. But for our purposes it is only important to realize that the work of both von Foerster and Maturana begins to loosen the restriction of the first premise of Ashby's general systems theory by taking

on board material (physical and biological) considerations. Remarkably, this shift of interest toward aspects of the *material object* that is described by the systemic framework is complemented by a growing awareness of the constitutive role of the observer, i.e., the *cognitive subject* who is actually doing the describing.

Challenge 2: What is the role of the observer?

This epistemological shift can be seen as a response to a second challenge posed by general systems theory, namely the problem of how best to understand the *epistemic status* of its insights. In fact, this epistemic uncertainty appears precisely at the point when one attempts to apply general systems theory's otherwise mathematically pure framework to explain concrete phenomena that are observed in the physical and biological domains. As Ashby puts it when tracing the meaning of the notion of a "systemic constraint" to the uncertainty of the observer:

“The ‘constraint’ is thus a relation between observer and thing; the properties of any particular constraint will depend on both the real thing and on the observer. It follows that a substantial part of the theory of organization will be concerned with *properties that are not intrinsic to the thing but are relational properties between observer and thing.*” (Ashby 1962: 258)

Interestingly, the fact that this admission is a first indication of the fundamentally relational nature of *all* knowledge, in so far as it depends on the perspective of our observations, is not acknowledged by Ashby. On the contrary, he laments that the systemic approach that he is developing has to deal with a "peculiarity not found in the more objective sciences of physics and chemistry" (Ashby 1962: 257). It is possible that this rather naïve conception of the work of the "hard sciences" represents for him the ideal scientific situation, which he tries to imitate in his attempts to devise concepts of biological phenomena that are "purely objective" (e.g., Ashby 1940: 483).

Against the assumption of a representationalist epistemology of the first generation of cyberneticists, which Ashby has retained, the decisive step that launches *second-order*

cybernetics as a distinct approach can be defined by an epistemological shift. It accepted and incorporated what previous research in cybernetics had already started to show: that scientific knowledge is a relational phenomenon in the domain of explanations that does not and cannot represent an observer-independent reality (Foerster 1973). In other words, the relativity of systemic insights to the perspective of the observer, previously only known to the first generation of cybernetics in terms of the arbitrariness of choices involved in distinguishing a particular system of interest, had by now been made explicit and further radicalized.

3. Transitions: Heinz von Foerster's second-order cybernetics

It is important to realize that the development of this second-order cybernetics was a historic moment for Western thought as a whole. It was only then, I submit, that Warren McCulloch's original vision for cybernetics as an *experimental epistemology* was for the first time fully implemented since the field's inception. In effect, the first cybernetics movement unreflectively accepted the standard representationalist epistemology that was dominating mainstream philosophy of science at the time, and thus proceeded to put it to a scientific test – perhaps more implicitly than explicitly, but tested nevertheless – by devising systems that embodied its theoretical principles. However, when the results of this cybernetic "experiment" started to call the prevailing epistemology into question, the majority of subsequent researchers simply abandoned the experiment rather than the epistemology (cf. Box 2).

It therefore seems fair to say that the creative circle that was first started by McCulloch and others during the first generation of cyberneticists was finally completed only after his death in 1969, when several leading researchers following in his footsteps decided to adjust their epistemological outlook according to what their experimental insights appeared to show (e.g., Maturana 1970a; Foerster 1973; Varela 1979, 1986). This new group of researchers shunned the prospects of a career in mainstream sci-

BOX 2: Computer science as cognitive science?

As we have seen, many cyberneticists sought refuge from the Ashbyan crisis in the idealized world of computer science, a strange utopia of trivial systems where you get out what you put in, and so there was less danger that one's cherished epistemology would be confronted and challenged by its own results. With this unfortunate move into the domain of informatics, the field of computer science became equivalent with cognitive science and the progressive potential of McCulloch's experimental epistemology was simply turned on its head. Instead, it developed into a conservative metaphysics: the domain of research was chosen so as to support the theory of knowledge, rather than vice versa.

If this retreat into symbolic AI and the computer metaphor of the mind had been the end of the story for cybernetics, the daring project of an experimental epistemology might have ended in failure. Indeed, even McCulloch (1960) had sensed that his grand vision might come to an end with his death, though he was fortunate enough to have witnessed the tentative beginnings of a new generation of inspired cybernetics (including Maturana's research on frog vision, which was a decisive influence on his later development of the concept of the circular organization of the living).

To be sure, even without this renewal of experimental epistemology, the checks and balances of the "reality principle" are never far away. Thus, when the field of symbolic AI eventually attempted to leave the sanctuary of its digital ivory tower and used robotic platforms to make tentative forays into the "real world" (still consisting of severely constrained and simplified artificial environments), the old problems reappeared (cf. Dreyfus & Dreyfus 1988).

It should also be noted that it is possible to make use explicitly of the peculiar property of the artificial medium to mirror our premises, namely as a tool to expose and undermine naive theoretical assumptions (cf. Di Paolo, Noble & Bullock 2000).

ence by taking it upon themselves to enter more deeply into that unfamiliar world of complex systems and self-involvement that had opened up before them. More specifically, they responded to the epistemological challenges that cybernetics had laid down before them by according the utmost significance to the active role of the observer in the constitution of objects for scientific study (cf. McGee 2005). This was the birth of a reflexive or "second-order" cybernetics that also addressed the activity of *observing systems* and not just the activity of the *observed systems*.

The second-order approach is perhaps best exemplified by the work of von Foerster, whose lifelong quest for more adequate ways of knowing exemplifies one of the most fascinating success stories of experimental epistemology in action (cf. Foerster & Bröcker 2002). We will here only focus on some noteworthy aspects of his complex and wide-ranging work that are specifically related to the discussion in this paper.

Linguistic hygiene

For example, one remarkable element that ties in nicely with the preceding discussion is von Foerster's aversion to confused discourse. When he joined the cybernetics community of the Macy Conferences in the 1950s and '60s, his position was actually not that different from what we have called representationalist computationalism, but by the '70s he was an ardent radical constructivist, and along the way his initial commitment to a moderate computationalism was gradually replaced by a profound concern for ethics (e.g., Foerster 1966, 1973, 1991).

Thus, even von Foerster, who in his later life worked tirelessly to eliminate the threat of what he called "pathogenic linguistic pollutants" (Foerster 1980), especially those resulting from the trivialization of such concepts as memory, information and communication, was unable to avoid the lure of the computer metaphor in the domain of neurophysiology in much of his early work. But at the same time, his intellectual and

personal development showed that there is no shame in being wrong – as long as one becomes aware of the mistake and recognizes it as an opportunity for change and growth.

The problem of trivialization

A few more words on von Foerster's notion of *trivialization* are in order because it informs a decisive part of his later period of work. In a broad sense, this notion is his diagnosis of the root cause of what he sees as the degenerate state of modern society as expressed, for example, in the widespread symptoms of lacking self-knowledge, limited capacity for insightful perception, and little desire for change by creative action (Foerster 1972). However, for second-order cybernetics, which understands the observer as being intimately related to the observed, a mere diagnosis is not enough. The fundamental self-involvement of the observer in her world is the basis for a non-moralistic ethics, and so the field has to practice what it preaches: it has to accept a part of the responsibility and engage in corrective action. A potential remedy, von Foerster (1984) suggests, lies in transforming the effective carrier of the pollutants, namely confused and restrictive discourse, by introducing distinctions that enable us to deal more effectively with the complexities of the human condition.

We have already mentioned the conceptual framework of the information-based approach as one important target for linguistic clarification in this regard. However, the most fundamental distinction for empowering the "silent majority" is the differentiation between trivial and non-trivial machines (Foerster 1972). The former are essentially reactive systems, whereby there is a one-to-one relationship between "input" (stimulus, cause) and "output" (response, effect), and the latter are what we could call "state-determined systems," whereby the "output" is not directly determined by the current "input" but is a combined result of the system's history of interactions and recursive operations on its own internal state.

Note that from this perspective, it becomes clear where von Foerster's original approval of the computer metaphor of mind came from. Since digital computers are also a type of non-trivial system, he could iden-

tify with the computationalist paradigm's staunch opposition against the trivial (reactive) systems that are exemplary of behaviorist psychology. Indeed, in this context it is important to recognize that the computationalist paradigm played an essential role in bringing the reign of behaviorism to an end, and even attempted to create a systematic methodology in which the existence of the subjective perspective was once more explicitly acknowledged within a scientific discourse (cf. Ericsson & Simon 1980).

Thus, it was only after the end of behaviorism, when there was time to take a closer look at the position of your "friends and neighbors," that von Foerster rejected the metaphor of the general purpose computer for political reasons. He began to realize that the computationalist paradigm was the right step in the wrong direction. For even though the digital computer was strictly speaking a non-trivial (state-determined) system, as a model of cognition it was suitable only for those who unquestioningly obey the authority of external commands, but not for those who decide to live according to autonomous choice and who thereby embrace their own inherent responsibility (Foerster 1980). Accordingly, his focus shifted more to the constitutive role of operational closure (recursion) for personal autonomy and the need for active movement in bringing forth the perceptual world.

In sum, we can describe the general framework of second-order cybernetics as a form of *constructivist mechanismism*: the representationalist epistemology has been replaced with a concern for the active role of the observer in "constructing" what is observed, and the notion of symbolic computation has been marginalized in favor of a dynamical approach grounded in the mechanisms of physics and biology.

Mind as a constructivist mechanism

The essential contribution of this framework, we argued above, was its resolute insistence on bringing the first round of experimental epistemology to completion – even when it meant risking personal careers to explore previously unknown terrain outside the accepted limits of mainstream science. In the end, the field's epistemology was adjusted to fit the insights generated by its own cybernetic experiment.

In fact, just as McCulloch was instrumental in getting the ball rolling, it is fair to say that von Foerster was decisive in setting the stage for the second round. In particular, he attempted to ground the newfound active role of the observer in a quintessential cybernetic manner, namely in the notion of the non-trivial machine. However, while taking another look at the science of cognitive systems from the new perspective of the adjusted epistemology was certainly a step in the right direction, von Foerster's own attempt ultimately fell short of its goal. The concept of the non-trivial machine, defined as a state-determined system, did not sufficiently safeguard the *autonomy* of the observer, which could ground their perspective. A rock, a computer, a living being and a person can all be described as different kinds of state-determined systems. Accordingly, the concept of the non-trivial machine is fundamentally insufficient to account for what makes some of the observed systems also be *observing* systems. At this point two possible responses become available: we can (i) refine the systemic concept, and/or (ii) further refine the constructivist epistemology. How do we address this challenge?

Before we highlight two traditions that systematically took up this challenge, it should be noted that this criticism of the insufficiency of non-triviality is not to be understood as the failure of second-order cybernetics. On the contrary, the emphasis of this particular shortcoming of second-order cybernetics is only supposed to mark the point at which a second round of experimental epistemology can begin. Indeed, that von Foerster himself was sympathetic to both of the possible responses noted above can be seen by his final efforts (i) to restrict the notion of non-triviality to denote only those machines for which the problem of identifying their structure is in principle unsolvable (and hence retain a space for the expression of their autonomy), and (ii) to develop an appropriate epistemology that would better allow for this kind of non-triviality by centering on "in principle undecidable questions" (e.g., Foerster 1991). It is also important to mention his idea of the "double closure" of the nervous system in this regard, which was an attempt to account for the possibility of self-regulation

(e.g., Foerster 1973). However, he never developed these tantalizing ideas into a more detailed and systematic research tradition.

Fortunately, we can find a comprehensive treatment of these possible responses in two other traditions for which von Foerster himself also had great affinity, namely (i) the *cognitive biology* initiated by Maturana & Varela's (1973) Santiago school of biology, and (ii) the *cognitive psychology* of Ernst von Glasersfeld's (1974, 1995) radical constructivism. Von Foerster was involved in the development of both of these traditions, even providing decisive help in getting the Santiago school scientifically established on the international stage (cf. Varela 1996). In terms of the published literature, however, he participated more actively in the development of radical constructivism (cf. Foerster & Glasersfeld 1999). While this constructivist approach retained many operational concepts of the original cybernetics, it also entailed a fundamental epistemological switch from one extreme perspective to the other: the essential locus of cognitive activity, previously reduced to feedback operations in the external world, was now located purely in the mental operations of the observer. As such, radical constructivism can be defined as a form of "constructivist intellectualism." In future work I plan to look more closely at how biology of cognition and radical constructivism are related to second-order cybernetics.

4. Discussion

Our journey started with the end of the cybernetics era, so it is fitting to close the circle by linking back to that momentous event. We argued that cybernetics died while giving birth to its two children, symbolic AI and second-order cybernetics: the one was the golden boy that became the foundation of the prestigious cognitive sciences, while the other was shunned as the ugly duckling and is still struggling for recognition. However, it is possible that with the ongoing development of the enactive paradigm in the cognitive sciences the transformation of this neglected offspring might finally have begun.

We have argued that two factors seem to have been decisive in the culmination in

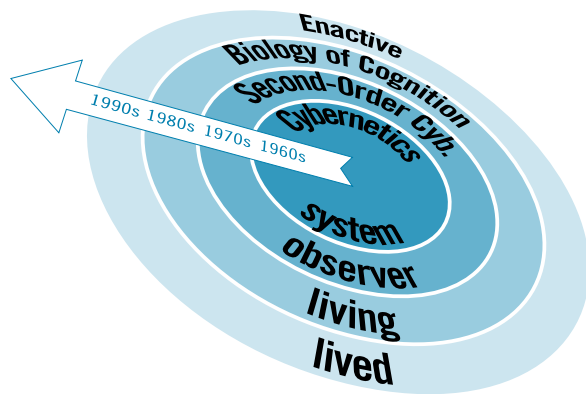


Figure 2: Illustration of the possibility of interpreting the history of the enactive paradigm in terms of events that happened *outside* the cognitive sciences. The history of ideas appears to be shaped by a sequence of emerging paradigms. These changes were driven by the explanatory need to address previous shortcomings in explaining cognition by appealing to a more inclusive set of determining factors.

cybernetics: (i) the constitutive interdependence of the observer with the observed was becoming more evident, and (ii) the dynamical systems perspective developed by Ashby and others left no room for a rational subject inside the physical world. In response, the computationalist movement within the original cybernetics movement was first solidified into symbolic AI and then became the experimental and theoretical foundation of the cognitive sciences. This paradigm of representationalist computationalism insisted that: (i) the observer is independent of the observed, and (ii) the rational subject is realized in the physical brain as a symbolic computer.

We further argued that this countermove to the Ashbyan crisis of cybernetics, which is still being scientifically promoted today as the mainstream approach to perception and cognition, entailed the double effect that: (i) the observer is safely distanced from the world, since the world is observer-independent, and at the same time (ii) safely distanced from their own existence, since the observer is actually nothing but a brain-based digital computer.

It is the second of these consequences that adds fuel to the computationalists'

charge of nihilism against a pure dynamical systems approach. The latter, if it were to completely replace the representationalist discourse, would seriously undermine the only possible sense of subjectivity that is conceivable within a computationalist framework: a brain-based computer in the world. Thus, the idea seems to be that it is better to be physically realized as a *computer* rather than as a *dynamical system* because at least the computer, as a system to which we can ascribe representations, has some resemblance to how we reason about ourselves and the world. Could the existential threat posed by a dynamical systems approach possibly explain why the cognitive sciences have generally failed to acknowledge their cybernetic heritage?

According to an idealized view of scientific progress, the development of novel paradigms should progress according to objective factors, and so we might be inclined to discount this line of argument. However, even if we do not adopt a social constructivist attitude to the history of science, it is difficult to deny that personal choice must play a significant role in this particular history. This is especially so considering that the computationalist paradigm provides no

serious scientific challenge to the possibility of a full-blown dynamical systems approach. Computationalist discourse is fundamentally confused (i) by using an irrational epistemology that attributes an absolute God's eye perspective the observer, and (ii) by an irresponsible use of personal-level agential notions in the domain of sub-personal mechanisms. The first confusion puts it on a par with the state of physics before the role of the observer became operationalized in general relativity and quantum mechanics, and the second confusion is akin to a pre-scientific animist view of the world where natural processes are explained in terms of agent-like forces acting within them.

These combined difficulties do not arise in the traditions of second-order cybernetics, radical constructivism, and biology of cognition. Nevertheless, it is clear that these alternative traditions are also challenged by their own set of significant problems, and we have looked more closely at those faced by second-order cybernetics in particular. The enactive paradigm has retained and further developed many of the central insights of these traditions, but it has also complemented them with a practical insistence on taking the concreteness of our own *lived experience* into consideration. A proper defense of this move is beyond the scope of this current paper, but I would like to indicate at least how it deals with the two difficulties identified above.

In effect, such phenomenological and existential analysis can show us that (i) we always already participate in the world before we isolate ourselves through abstract reflection, and (ii) we do not need to find a representation of ourselves as a cognitive subject in our "computer-like brains" because we already coincide with our own lived experience. In the end, the irony might therefore be that if the cognitive sciences are to become more objective, we researchers have to become more aware of our own lived subjectivity. It appears that only in this way can we make some existential room for the kind of dispassionate gaze that is necessary in order to study the material underpinnings of our minds without inadvertently projecting subjective content onto what is observed.

The general suggestion put forward in this article is that the origins of the enac-

tive paradigm are better understood if we trace the history of its central ideas to developments that took place outside of the cognitive sciences. In general, these developments appear to have expanded the domain of inquiry to wider regions of interest. A tentative summary of these trends is illustrated in Figure 2.

In this paper we have solely focused on the transition from the end of the early generation of cybernetics, as epitomized by the work of Ashby in the 1950s and '60s, to its later second-order manifestation, as represented by the work of von Foerster in 1970s and '80s. In essence, the mathematics of general systems theory was retained in this shift, though it introduced an increased awareness of the importance of reflexivity and the active role of the observer. As such, we can say that the core of this first shift consisted of a revised, *constructivist epistemology*, though it remains to be seen how this shift related to the tradition of radical constructivism.

Future work in this area could focus on the other major transitions that are indicated in Figure 2. Even though we did not specifically support these claims in this paper, it is possible to argue that the subsequent shift from second-order cybernetics to Maturana & Varela's (1987) biology of cognition essentially grounded this revised epistemology in a specific form of *systems biology* that is centered on the autonomy of the living. The enactive paradigm then complemented this biology with an *existential phenomenology* centered on our lived experience (Varela, Thompson & Rosch 1991). Though these developments largely happened in parallel to the history of cognitive sciences, which was illustrated in Figure 1, they are now once more becoming intertwined in the form of enactive cognitive science. A simple schematic of this historical process is shown in Figure 3.

The fact that the central ideas of the enactive paradigm have been shaped by a tradition that has developed almost entirely outside of the scope of the cognitive sciences is something that is worth exploring in more detail. It might, for instance, explain why there is sometimes an ambiguity about whether the enactive paradigm entails another minor reformation or a major revolution of the cognitive sciences.

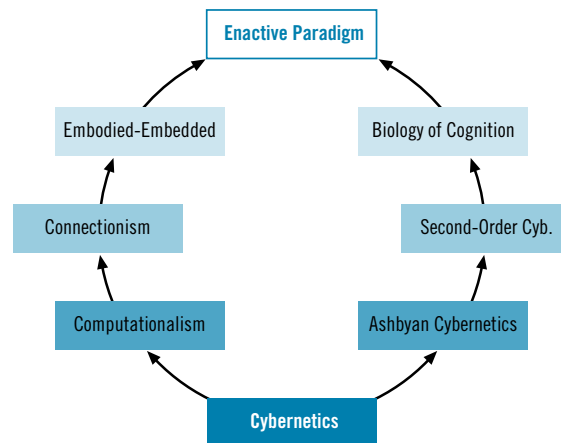


Figure 3: The enactive paradigm unifies two distinct scientific traditions focused on cognition. Both of these traditions first arose out of the collapse of cybernetics, but have followed largely separate historical developments since then.

5. Conclusion

Finally, it is also worth emphasizing what is not shown in Figure 3, namely that the enactive paradigm is busily incorporating additional new influences and traditions into its expanding and refining framework. For instance, while cybernetics unfortunately missed several encounters with other intellectual traditions of its time, including existentialism and phenomenology (cf. Dupuy 2009: 102–107), the continued development of these philosophical traditions has now become a foundational part of the enactive research program. Moreover, despite having a specific theoretical framework, the paradigm is still flexible enough to include impulses from a variety of different directions (e.g., Stewart, Gapenne & Di Paolo in press). For example, a renewed appreciation for cultural and social anthropology and analytic psychotherapy are also on the horizon.

We are witnessing the emergence of a constellation of theoretical, empirical and practical disciplines that is unique in the history of ideas. My personal feeling is that we might be on the verge of another momentous period of stimulating work akin to the emergence of the original cybernetics tradition.

References

- Ashby W. R. (1940) Adaptiveness and equilibrium. *Journal of Mental Science* 86: 478–484.
- Ashby W. R. (1947) The nervous system as physical machine: With special reference to the origin of adaptive behavior. *Mind* 56(221): 44–59.
- Ashby W. R. (1956) An introduction to cybernetics. Chapman and Hall, London.
- Ashby W. R. (1960) Design for a brain: The origin of adaptive behavior. Second edition. Chapman and Hall, London.
- Ashby W. R. (1962) Principles of the self-organizing system. In: Foerster H. von & Zopf G. W. (eds.) *Principles of self-organization: Transactions of the University of Illinois symposium*. Pergamon Press, London: 255–278.
- Boden M. A. (2006) *Mind as machine: A history of cognitive science*. 2 Volumes. Oxford University Press, Oxford.
- Brooks R. A. (1991) Intelligence without representation. *Artificial Intelligence* 47(1–3): 139–160.
- Clark A. (1997) *Being there: Putting brain, body, and world together again*. MIT Press, Cambridge MA.
- Clark A. (2001) *Mindware: An introduction to the philosophy of cognitive science*. Oxford University Press, Oxford.



THE AUTHOR

Tom Froese received his D.Phil. in Cognitive Science from the University of Sussex in 2009. The focus of his doctoral research was the relationship between life, mind and sociality, which he explored using a variety of methods, including evolutionary robotics and philosophical phenomenology. More recently he has been interested in the cross-section of phenomenological and systemic approaches in the study of consciousness. He is currently a visiting post-doctoral researcher at the Sackler Centre for Consciousness Science.

- Di Paolo E. A. (2001)** Book review of "The mechanization of the mind: On the origins of cognitive science" by Jean-Pierre Dupuy. *Journal of Cognitive Systems Research* 2: 291–295.
- Di Paolo E. A. (2003)** Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop. In: Murase K. & Asakura T. (eds.). *Dynamical systems approach to embodiment and sociality*. Advanced Knowledge International, Adelaide, Australia: 19–42.
- Di Paolo E. A., Noble J. & Bullock S. (2000)** Simulation models as opaque thought experiments. In: Bedau M. A., McCaskill J. S., Packard N. H., Rasmussen S. (eds.), *Artificial life VII: Proceedings of the 7th international conference on artificial life*. MIT Press, Cambridge MA: 497–506.
- Dreyfus H. L. & Dreyfus S. E. (1988)** Making a mind versus modelling the brain: Artificial intelligence back at a branch-point. *Daedalus* 117(1): 15–44.
- Dupuy J.-P. (2009)** On the origins of cognitive science: The mechanization of mind. MIT Press, Cambridge MA.
- Ericsson K. A. & Simon H. A. (1980)** Verbal reports as data. *Psychological Review* 87(3): 215–251.
- Fodor J. A. (1975)** *The language of thought*. Harvard University Press, Cambridge MA.
- Foerster H. von (1960)** On self-organizing systems and their environments. In: Yovits M. C. & Cameron S. (eds.) *Self-organizing systems*. Pergamon Press, London: 31–50.
- Foerster H. von (1966)** From stimulus to symbol: The economy of biological computation. In: Kepes G. (ed.) *Sign, image and symbol*. George Braziller, New York: 42–61.
- Foerster H. von (1972)** Perception of the future and the future of perception. *Instructional Science* 1(1): 31–43.
- Foerster H. von (1973)** On constructing a reality. In: Preiser W. F. E. (ed.) *Environmental research design*. Volume 2. Hutchinson and Ross, Stroudsburg, Dowden: 35–46.
- Foerster H. von (1979)** Cybernetics of cybernetics. In: Krippendorff K. (ed.) *Communication and control*. Gordon and Breach, New York: 5–8.
- Foerster H. von (1980)** Epistemology of communication. In: Woodward K. (ed.) *The myths of information: Technology and postindustrial culture*. Coda Press, Madison: 18–27.
- Foerster H. von (1984)** Disorder/Order: Discovery or invention? In: Livingston P. (ed.) *Disorder and order: Proceedings of the Stanford international symposium (14–16 September 1981)*. Anma Libri, Saratoga CA: 177–189.
- Foerster H. von (1991)** Through the eyes of the other. In: Steier F. (ed.) *Research and reflexivity*, Sage, London: 63–75.
- Foerster H. von & Bröcker M. (2002)** *Teil der Welt. Fraktale einer Ethik – oder Heinz von Foerstertanz mit der Welt*. Carl-Auer Verlag, Heidelberg.
- Foerster H. von & Glasersfeld E. von (1999)**. *Wie wir uns erfinden. Eine Autobiographie des radikalen Konstruktivismus*. Carl-Auer Verlag, Heidelberg.
- Froese T. (2007)** On the role of AI in the ongoing paradigm shift within the cognitive sciences. In: Lungarella M., Iida F., Bongard J. & Pfeifer R. (eds.) *50 years of artificial intelligence: Essays dedicated to the 50th anniversary of artificial intelligence*. Springer, Heidelberg: 63–75.
- Froese T. (2009)** Hume and the enactive approach to mind. *Phenomenology and the Cognitive Sciences* 8(1): 95–133.
- Froese T. & Ziemke T. (2009)** Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence* 173(3–4): 366–500.
- Glasersfeld E. von (1974)** Piaget and the radical constructivist epistemology. In: Smock C. D. & Glasersfeld E. von (eds.) *Epistemology and education*. Follow Through Publications, Athens GA.
- Glasersfeld E. von (1995)** *Radical constructivism: A way of knowing and learning*. Falmer, London.
- Grey Walter W. (1950)** An imitation of life. *Scientific American* 182(5): 42–45.
- Harnish R. M. (2002)** *Minds, brains, computers: An historical introduction to the foundations of cognitive science*. Blackwell, Malden MA.
- Hume D. (1739–1740)** *A treatise of human nature*. English translation in: Norton D. F. & Norton M. J. (eds.) (2000) *The Oxford philosophical texts edition of David Hume, A treatise of human nature*. Oxford University Press, Oxford: 1–396.
- Husbands P., Holland O. & Wheeler M. (eds.) (2008)** *The mechanical mind in history*. MIT Press, Cambridge MA.
- Hutchins E. (1995)** *Cognition in the wild*. MIT Press, Cambridge MA.
- Ikegami T. & Suzuki K. (2008)** From homeostatic to homeodynamic self. *BioSystems* 91(2): 388–400.
- Jonas H. (1966)** *The phenomenon of life: Toward a philosophical biology*. Harper & Row, New York. Republished in: (2001) *Northwestern University Press, Evanston IL*.
- Kant I. (1790)** *Kritik der Urteilskraft*. English translation: Kant I. (1987) *Critique of Judgment* (Translated by W. S. Pluhar). Hackett, Indianapolis IN.
- Levine J. (1983)** Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64: 354–361.
- Maturana H. R. (1970a)** Neurophysiology of cognition. In: Gavin P. L. (ed.) *Cognition: A multiple view*. Spartan Books, New York, NY: 3–23.

- Maturana H. R. (1970b)** Biology of cognition. In: Maturana H. R. & Varela F. J. (1980), *Autopoiesis and cognition: The realization of the living*. Kluwer, Dordrecht: 1–58.
- Maturana H. R. & Varela F. J. (1973)** Autopoiesis: The organization of the living. In: Maturana H. R. & Varela F. J. (1980) *Autopoiesis and cognition: The realization of the living*. Kluwer Academic, Dordrecht: 59–140.
- Maturana H. R. & Varela F. J. (1987)** The tree of knowledge: The biological roots of human understanding. Shambhala Publications, Boston MA.
- McClelland J. L., Rumelhart D. E. & the PDP Research Group (1986)** Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models. MIT Press, Cambridge MA.
- McCulloch W. S. (1960)** What is a number, that man may know it, and a man, that he may know a number? *General Semantics Bulletin* 26/27: 7–18.
- McCulloch W. S. & Pitts W. H. (1943)** A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–133.
- McGee K. (2005)** Enactive cognitive science. Part 1: Background and research themes. *Constructivist Foundations* 1(1): 19–34.
- Newell A. & Simon H. A. (1976)** Computer science as empirical enquiry: Symbols and search. *Communications of the Association for Computing Machinery* 19(3): 113–126.
- Rosenblueth A. N., Wiener N. & Bigelow J. (1943)** Behavior, purpose and teleology. *Philosophy of Science* 10: 18–24.
- Shannon C. E. & Weaver W. (1949)** The mathematical theory of communication. University of Illinois Press, Urbana IL.
- Stewart J., Gapenne O. & Di Paolo E. A. (eds.) (in press)** *Enaction: Towards a new paradigm for cognitive science*. MIT Press, Cambridge MA.
- Thompson E. (2007)** *Mind in life: Biology, phenomenology, and the sciences of mind*. MIT Press, Cambridge MA.
- Varela F. J. (1979)** *Principles of biological autonomy*. Elsevier North Holland, New York.
- Varela F. J. (1986)** *Experimental epistemology: Background and future*. *Revue Internationale De Systemique* 1(2): 209–223.
- Varela F. J. (1996)** The early days of autopoiesis: Heinz and Chile. *Systems Research* 13(3): 407–416.
- Varela F. J., Thompson E. & Rosch E. (1991)** *The embodied mind: Cognitive science and human experience*. MIT Press, Cambridge MA.
- Varela F. J., Maturana H. R. & Uribe R. (1974)** Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems* 5: 187–196.
- Uexküll J. von (1928)** *Theoretische Biologie*. 2nd revised edition. Julius Springer, Berlin. English translation of the 1st edition: Uexküll J. von (1926) *Theoretical Biology*. Harcourt Brace, New York.
- Weber A. & Varela F. J. (2002)** Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences* 1: 97–125.
- Ziemke T. & Sharkey N. (2001)** A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica* 134(1–4): 701–746.

RECEIVED: 9 JANUARY 2010

ACCEPTED: 11 MARCH 2010