Cybernetics and Human Knowing. Vol. 17, no. 4, pp. 7-50

Life After Ashby:
Ultrastability and the Autopoietic Foundations of Biological Autonomy

Tom Froese[1] and John Stewart[2]

The concept of autopoiesis was conceived by Maturana and Varela as providing the necessary and sufficient conditions for distinguishing the living from the non-living (and, by extension, the cognitive from the non-cognitive). More recently however, there has been a growing consensus that their original conception of autopoiesis is necessary but insufficient for this task as it fails to meet a number of constructive, interactive, normative, and historical requirements. We argue that it also fails to satisfy crucial phenomenological requirements that are motivated by the ongoing appropriation of autopoiesis as a key concept in enactive cognitive science. The root of these problems can be traced to the abstract general systems framework in which the ideas were first formulated, as epitomized by Ashby's cybernetics. While this abstract generality has helped the concept's popularity in some circles, we insist that a restriction of autopoiesis to a radical embodiment in chemical self-production under far-from-equilibrium conditions is necessary if the concept is to live up to its original intentions.

Introduction

What is life? What is mind? These questions are more likely to be asked by a philosopher than a biologist or cognitive scientist, even despite the fact that these natural phenomena supposedly constitute the scientists' respective fields of inquiry. But if we scientists do not try to address these definitional questions in some way, then how do we decide what to investigate? To be sure, if they were pressed for an answer most scientists would likely respond by appealing to a conceptual mixture featuring the genetic code, information processing, and survival of the fittest. We will not enter here into the persisting debates about the merits and detriments of this standard perspective. We simply wish to point out that something crucial is missing in this response if we are to make sense of the very attempt itself: our own lived perspective as human subjects.

Indeed, what is currently missing from science is any consideration of the fact that we experience ourselves as living agents who are engaged in a meaningful and goal-directed activity, and that this authenticity of purpose is necessary for reasoned debate and the establishment of a scientific community. But what precisely is such lived and living agency? In order to address this question we will consider an alternative response to the question of life, one based on the notion of autopoiesis.

1. Centre for Computational Neuroscience and Robotics (CCNR), University of Sussex, Brighton, UK. Email: t.froese@gmail.com
2. Knowledge, Organization and Technical Systems (COSTECH), University of Technology of Compiegne, Compiegne, France. Email: johnmacgregorstewart@gmail.com

1.1. Introducing Maturana and Varela's Autopoiesis

The term *autopoiesis* was first coined by the Chilean biologists Humberto Maturana and Francisco Varela (1973) in their attempt to define the essential organization of living systems, namely as a network of processes that produces its own system identity (cf. section 2 for a standard definition). In contrast to the dominating position of the neo-Darwinian synthesis, which prefers to define life in terms of genes, information, and natural selection, Maturana and Varela proposed that autopoiesis is both necessary and sufficient for life, and that life is both necessary and sufficient for cognition.

However, even though the primary literature written by Maturana and Varela can give the impression that the final word on the topic has been said, it is evident that even their own definitions of autopoiesis have more or less subtly changed over the years (cf. Bourgine & Stewart, 2004). The formulations which one encounters in their publications do not vary substantially, but there are noticeable shifts of emphasis depending on which aspects of life happen to be of interest at the time (e.g., organizational closure, semi-permeable membrane, molecular production, etc.). In any case, so compelling were their arguments and so forceful (or imposing?) was their style of writing that the notion of autopoiesis = life = cognition continues to be an influential idea today, though still largely outside of mainstream biology (cf. Stewart, 1992, 1996).

Can their proposal be turned into a workable scientific research program for biology and the cognitive sciences? There have been some valiant attempts in the past, but the state of the art of today's autopoietic framework is largely the same as when it was conceived 40 years ago. Part of the problem has been that Maturana and Varela assumed the role of final arbiters of how the concept of autopoiesis should be used and interpreted, while at the same time being unable to clarify the concept sufficiently to disambiguate its intended meaning. For instance, even though they often claimed that autopoiesis was restricted to the chemical domain, their insistence on abstracting away from concrete material and energetic requirements appeared to make the concept applicable more generally (e.g., to social systems; cf. Luhmann, 1984, 2002). Similarly, it is to Maturana's credit that he managed to devise an alternative worldview based on the notion of structure determinism (i.e., the idea that an organism's response to a stimulus depends on its current structure). But in its abstract formulation this notion appears to be nothing but state determinism, which is a general mathematical concept that applies to all dynamical systems except for idealized reactive systems. Again, it is not clear how this general notion helps us to better understand what is special about the phenomenon of life as such.

We are therefore left with two choices: Either (i) we accept that these particular concepts specify the phenomenon of life while at the same time being more widely applicable, thus in effect admitting that there is nothing special about life as such, or (ii) we argue that life does have unique characteristics that distinguish it from non-living phenomena, and we thereby acknowledge that the traditional formulation of these concepts, whatever their value may be in other contexts, is inadequate. The first option is tantamount to giving up the notion of life altogether, while the second requires us to move beyond Maturana and Varela's initial account of life—and to properly distinguish between the phenomenon and its explanation, as Maturana and Varela so often remind us.

1.2. Beyond Maturana and Varela's Autopoiesis

Given this situation, it should come as no surprise that in the last two decades a growing number of scientists in the fields of artificial life, robotics, theoretical biology, cognitive science, and bio-semiotics have voiced a number of fundamental criticisms of the original concept of autopoiesis, while still accepting the idea that chemical self-production plays an essential role for life, cognition, and agency. More precisely, this return from the lofty rhetoric of abstract systems to the concrete phenomenon of life itself has helped to uncover a number of essential requirements which are not satisfactorily addressed by Maturana and Varela's original definition of autopoiesis (we will refer to this definition as Maturana and Varela's autopoiesis, or MV-A). These missing requirements and the specific effects that their absence has on our understanding of the phenomena of life and mind can be summarized as follows:

(i) **Constructive requirements:** MV-A is too static and abstract to distinguish the living from the non-living (e.g., Sheets-Johnstone, 2000; Fleischaker, 1988); it is lacking all consideration of material and energetic conditions for self-organization, emergence, and individuation (e.g., Collier, 2004, 2008; Bickhard, 2008), as well as being generally silent on the implications and requirements of self-production under far-from-equilibrium conditions (e.g., Ruiz-Mirazo & Moreno, 2004; Moreno & Etxeberria, 2005; Christensen & Hooker, 2000).

(ii) **Interactive requirements:** MV-A is too isolated; it thus ignores a possible incorporation of environment (e.g., Virgo, Egbert, & Froese, in press) and a dependence on other living systems (e.g., Ruiz-Mirazo, Peretó, & Moreno, 2004; Hoffmeyer, 1998); it also lacks consideration of functional interaction cycles (e.g., Arnellos, Spyrou, & Darzentas, 2010; Di Paolo, 2009; Barandiaran & Moreno, 2008; Moreno & Etxeberria, 2005).

(iii) **Normative requirements:** MV-A cannot account for normative activity such as adaptivity and goal-directed action (e.g., Barandiaran, Di Paolo, & Rohde, 2009; Di Paolo, 2005; Barandiaran & Moreno, 2008; Bickhard, 2009), as well as sensorimotor action and cognition (e.g., Barandiaran & Moreno, 2006; Bourgine & Stewart, 2004; Christensen & Hooker, 2000).

(iv) **Historical requirements:** MV-A is lacking a capacity for memory, learning, and development (e.g., Hoffmeyer, 2000), as well as for genetic material (e.g., Hoffmeyer,1998; Emmeche, 1998; Brier, 1995), and is thus unsuitable for open-ended evolution (e.g., Ruiz-Mirazo et al., 2004).

Since the satisfaction of the latter requirements constitutively and historically depends on the satisfaction of the former, it is widely recognized that improving on the concept of autopoiesis first and foremost depends on paying closer attention to the physical and chemical realization of living systems in far-from-equilibrium conditions (constructive requirements). In other words, similar to the growing appreciation of the recent embodied turn in the cognitive sciences (e.g., Varela, Thompson, & Rosch, 1991), we have to start taking the *embodiment* of life more seriously if we want a naturalized account of living and lived agency. Fortunately, as indicated by the selection of references above, there are already several research groups around the world which are currently working on fleshing out our understanding of life so that it satisfies these missing requirements.

There is one more essential aspect of the phenomenon of life, however, which has not yet been specifically included in this list of requirements, namely the lived presence from the perspective of the living being itself. As von Uexküll (1934/1957) would say, all living beings are embodied subjects which are embedded in an intricate web of intrinsically meaningful relationships, their *Umwelt*. Interestingly, in the case of human beings, the presence of a world has sometimes been taken as a definition of consciousness. However, since we are referring to the presence of mind in life more generally, we prefer to use the less contentious term *lived experience* which is taken from the Husserlian tradition of phenomenology (cf. Thompson, 2007). In other words, we are not simply talking about meaning in terms of informational or normative constraints. We want to go further and suggest that, using Nagel's (1974) famous turn of phrase, there is always something "it is like to be" a living being. As will become clearer in the next section, Maturana and Varela's original formulation of autopoiesis is insufficient to account for this lived dimension of living being. We can therefore add another category of requirements to our existing list:

(v) **Phenomenological requirements:** MV-A is insufficient for grounding a lived perspective of concern (e.g., Di Paolo, 2005), desire (e.g., Barbaras, 2002), and thus of lived experience more generally.

This particular category of requirements is mainly being developed in the context of the *enactive* approach in the cognitive sciences, which is currently positioning itself as a comprehensive alternative to the computationalist mainstream (e.g., Stewart, Gapenne, & Di Paolo, in press; Di Paolo, Rohde, & De Jaegher, in press; Thompson 2007; Varela, et al. 1991). Very briefly, one of its main contributions on a practical level has been to raise awareness of the need for a rigorous examination of lived experience if we are to better understand the biological foundations of the mind (e.g., Varela & Shear, 1999). And on a conceptual level it is advancing a resolution of the mind-body problem based on the idea that life is a unified phenomenon which integrates both lived and living being (cf. Hanna & Thompson, 2003; Weber & Varela, 2002). We will provide a more detailed treatment of the enactive approach in the next section.

1.3. Overview of the Paper

The upshot of our analysis will be that the concept of autopoiesis only makes sense when conceived as a very specific form of chemical self-production under far-from equilibrium conditions. Of course, not everyone will be happy about this strict concretization of the concept, especially if their work has benefited from the generality of Maturana and Varela's original abstract definition. And while arguments can certainly be made for conceiving autopoiesis as a general systems concept, we prefer to side with Maturana and Varela's original intention of trying to provide an operational account of the living. As a consequence this means that the precise definition of autopoiesis cannot be static, as would be the case for an abstract and universal systemic principle, but always depends on our current best understanding of the biological phenomena.

For instance, at the moment there is a growing consensus that an essential aspect of the phenomenon of life is its precarious existence in far-from-equilibrium conditions, and that this must therefore be taken into consideration when talking about autopoiesis (e.g., Ruiz-Mirazo et al., 2004). At the same time we must acknowledge that, even though the terms in which Maturana and Varela originally framed their ideas are no longer adequate to capture what we currently know about the phenomenon of life, their original systemic expression may still be valid for other, non-living phenomena. We therefore propose to trace the historical roots of the autopoietic tradition to the general cybernetic framework in which they were first expressed. This will enable us to make their original definition available in general terms in the form of Ashby's notion of ultrastability.

At the same time this historical analysis will cast doubt on purely theoretical approaches to biology and the cognitive sciences. It turns out that two highly influential cybernetic discoveries, namely that the organization of organisms can be described in systemic terms and that the organization of the brain can be described in terms of binary logic, have seduced generations of our best scientists into a Platonic realm of idealities. Thus, the latter discovery led to the still prevailing movement that placed computer science at the heart of cognitive science, while the former placed the mathematics of dynamical systems at the core of organismic biology. However, from the perspective of biological embodiment, both traditions have gotten carried away by the theoretical implications of their initial assumptions, and in both cases merely formal disputation has overshadowed the concrete phenomena given in our experience. The task faced by this generation of scientists is therefore to bring these magnificent castles in the sky back down to the living where they can be turned into fruitful and progressive scientific research programs.

In brief, the next sections will unfold as follows: In section 2 we will summarize the key insights of the enactive approach to the cognitive sciences as they pertain to the category of phenomenological requirements on autopoiesis. This will provide us with a reference point for what we need from that concept if it is to provide the biological foundations of a phenomenologically informed notion of life and mind. Then, in section 3, we return to the first era of cybernetics, which can be considered the modern starting point of both the autopoietic tradition and the cognitive sciences (cf. Froese, 2010), in order to have a closer look at the conceptual tools that were available to biologists during that time. More specifically, the work of Ashby will serve as a paradigmatic example for the cybernetic roots of systems biology. In section 4 we will examine how this framework of thinking influenced the expression of Maturana's biology of cognition, including his notion of the circular organization of the living which was a direct precursor to the concept of autopoiesis. The final step of our historical analysis, in section 5, presents a reassessment of Maturana and Varela's original conception of autopoiesis. The historical perspective we developed in the preceding sections enables us to uncover a tension between the inherited mode of cybernetic thought and a deeper appreciation of the phenomenon of life. In section 6 we discuss ways in which the concept of autopoiesis has been subsequently refined in the individual work of Maturana and Varela so as to satisfy more adequately the various requirements that are made upon it. We note how this has resulted in two competing viewpoints, which we call the Ashbyan (e.g., Maturana) or the Kantian (e.g., Varela) conception of autopoiesis. Some conclusions are presented in section 7.

2. The Enactive Paradigm: Autopoiesis and Sense-making

The origin of the autopoiesis concept, which tries to capture the circular organization of the living as the root of biological autonomy, can be traced back to the Santiago school in systems biology that originated with the seminal research of Humberto Maturana and Francisco Varela in the late 1960s.[3] The term *autopoiesis* was coined in 1971, though it wasn't until 1974 that it was first introduced to the English-speaking community (Varela, Maturana, & Uribe, 1974). In this section we are interested in the role of this concept for the enactive paradigm, so we will begin our discussion by considering a definition that has been used by Varela in a series of biological publications in the 1990s (e.g., Varela, 1997), and which has since carried over into the literature of enactive cognitive science:

> An autopoietic system—the minimal living organization—is one that continuously produces the components that specify it, while at the same time realizing it (the system) as a concrete unity in space and time, which makes the network of production of components possible. More precisely defined: an autopoietic system is organized (defined as a unity) as a network of processes of production (synthesis and destruction) of components such that these components:
>
> 1. continuously regenerate and realize the network that produces them, and
> 2. constitute the system as a distinguishable unity in the domain in which they exist.
>
> (Varela, 1997, p. 75)

In order to avoid any confusion we should immediately clarify that the self-constitution of an identity also entails the constitution of a relational domain between the system and its environment. The shape of this relational domain is not a pre-given

3. See Varela (1996) and Maturana (2002) for more detailed accounts of the historical circumstances under which the notion of autopoiesis was first conceived and developed.

a priori, but is co-determined by the organization of the system and its current environment. Accordingly, any system which fulfils the criteria for autopoiesis also generates its own domain of possible interactions in the same movement in which it gives rise to its emergent identity (Thompson, 2007, p. 44). It is important to realize that such a self-producing system, which is generating its own identity by separating itself precisely from that which it is not, simultaneously generates the very conditions by which it can relate to this "other," that is, its environment (Weber, 2001). It is this fundamental *asymmetry* of the organism-environment relationship, which places the locus of activity on the side of the system, which essentially constitutes the organism's *perspective* on that environment:

> Now, in this dialogic coupling between the living unity and the physicochemical environment, there is a key difference on the side of the living since it has the active role in this reciprocal coupling. In defining what it is as unity, in the very same movement it defines what remains exterior to it, that is to say, its surrounding environment. … the autopoietic unity creates a perspective from which the exterior is one, which cannot be confused with the physical surroundings as they appear to us as observers, the land of physical and chemical laws *simpliciter*, devoid of such perspectivism. (Varela, 1997, p. 78)

Of course, this is not to say that we cannot provide a description of the organism and its environment in physicochemical terms.[4] The point is merely that this type of description does not exhaust the domain of phenomena with which an organismic biology should be concerned. In other words: "One could envisage the circularity metabolism-membrane entirely from the outside (this is what most biochemists do). But this is not to deny that there is, at the same time, the instauration of a *point of view* provided by the self-construction" (Weber & Varela, 2002, p. 116). Moreover, this point of view is not only an abstract reference point for the biologist who is trying to understand the behavior of the living system, thereby moving us from biochemistry to ethology, it provides the passage from mere material happenings to a lived existence:

> In other words by putting at the center the autonomy of even the minimal cellular organism we inescapably find an intrinsic teleology in two complementary modes. First, a *basic* purpose in the maintenance of its own identity, an affirmation of life. Second, directly emerging from the aspect of concern to affirm life, a *sense-creation* purpose whence meaning comes to its surrounding, introducing a difference between environment (the physical impacts it receives), and world (how that environment is evaluated from the point of view established by maintaining an identity). (Weber & Varela, 2002, p. 117)

Let us briefly consider these two aspects in turn. First, there is the notion of *intrinsic teleology* in terms of the organism's relation to its own identity, that is, its capacity to constitute its own purposeful and goal-directed existence. Weber and Varela derive this notion by combining Kant's conception of a *natural purpose*, namely the idea that

4. Nor should Varela be misunderstood as implying that the physical surroundings that are described by scientists are devoid of all perspectivism in that they reflect an absolute reality (cf. Varela et al., 1991, pp. 9-12). We can better understand his point by realizing that the scientific description of the surroundings is generated precisely by stripping the world of its significance (cf. Dreyfus, 1991, pp. 112-121).

a self-organizing system that is both cause and effect of itself is also its own means and purpose (cf. Kant, 1987, §64-65), with our modern understanding of autopoiesis. Then, by appealing to the philosophical biology of Hans Jonas (1966/2001), they move beyond Kant's conception of teleology as a useful regulative idea for the observer, and posit teleology as intrinsic to the phenomenon of life itself.

In addition, Jonas argues that the precarious situation of the living furnishes the organism with more than just an inherently purposeful existence: "The organism has to keep going, because to be going is its very existence—which is revocable—and, threatened with extinction, it is *concerned* in existing" (Jonas, 1966/2001, p. 126, emphasis added). He thus argues that it is the generation of a precarious identity through material self-production that simultaneously enables the generation of a lived existence concerned with existential values. More poetically expressed:

> The basic clue is that life says yes to itself. By clinging to itself it declares that it values itself. … Are we then, perhaps, allowed to say that mortality is the narrow gate through which alone *value*—the addressee of a yes—could enter the otherwise indifferent universe? (Jonas, 1992, p. 36)

This brings us to the second aspect of the organism's perspective, namely its capacity for sense-creation or *sense-making*. This notion highlights that the generation of values always happens in the context of a particular organism-environment relationship. The internal and external encounters that perturb the process of identity generation take on a value in relation to this precarious identity: "The perspective of a challenged and self-affirming organism lays a new grid over the world: a ubiquitous scale of value. To have a world for an organism thus first and foremost means to have value which it brings forth by the very process of its identity" (Weber & Varela, 2002, p. 118). In other words, an organism's world is first and foremost a meaningful *presence* (a world of significance) that is related to its particular manner of physically realizing its identity. To quote a famous illustrative example from Varela: "There is no food significance in sucrose except when a bacteria swims upgradient and its metabolism uses the molecule in a way that allows its identity to continue" (Varela, 1997, p. 79). In this way we can define living as a process of sense-making (Thompson, 2004).

But is Maturana and Varela's original concept of autopoiesis sufficient to account for these basic phenomenological requirements, that is, of the phenomenon of life as living and lived sense-making? For instance, according to Weber and Varela's (2002, p. 117) assessment, the constitution of values requires an evaluation of significance in relation to the metabolic needs of the organism. And the realization of this capacity appears to depend on a separate dedicated mechanism, for example on *adaptivity* (e.g., Di Paolo, 2005; Barandiaran & Moreno, 2008). Furthermore, the original definition says nothing about the kind of material and energetic conditions under which the autopoietic system's operations are taking place. It therefore remains unclear whether the system's existence is actually the necessary result of its own ongoing constructive/interactive activity unless we specifically qualify it, for instance, as existing in precarious far-from-equilibrium conditions (Di Paolo, 2009). And even if these additional requirements have been met we are still left with an essentially conservative notion that is insufficient to account for why living is a dynamic process of becoming (e.g., change, transformation, and development). From a lived perspective we can observe that the mere conservation of identity cannot ground the proactive presence of drives, instincts and desires (Barbaras, 2002). This is likely to require some internally generated thermodynamic instability.

But why are these extra qualifications needed if autopoiesis was supposed to be necessary and sufficient to distinguish the class of living beings? One possibility is that we are simply helping to refine and make explicit what was already implicit in the original definition. At the same time, however, the number and profundity of these issues shows that the problem goes a little deeper than that. Accordingly, we will need to determine the way of thinking in which the concept was originally expressed in order to address the problem at its roots. It will turn out that what made the concept inadequate for the phenomenon of life is precisely what enabled the facility with which the concept has subsequently been appropriated to serve other contexts, that is, as an all-or-nothing set theoretic concept which is applicable to general systems.

The task of the following sections will be to determine the explicit and implicit premises that led to the formulation of the concept of autopoiesis, and to consider which of these premises is causing difficulties in its current appropriation in organismic biology and enactive cognitive science.

3. Ashby's Cybernetics

We begin with an extensive critical analysis of the hugely influential work by the British cyberneticist W. Ross Ashby (1903-1972; cf. Ashby, 2009), who played a pivotal role in the development and establishment of general systems theory as a rigorous discipline in its own right (Klir, 2009), and in particular in its application to biological and psychological phenomena of adaptation (cf. Umpleby, 2009; Pickering, 2009). Ashby was a leading member of a mechanistic school of psychiatry. His practical experiments with patients motivated him to construct a better theoretical understanding of the operations of the human brain, which finally led him to devise a general mathematical theory of the machine as a dynamical system (including living machines).

3.1 Living Is Surviving: The Concept of Stability

It is possible to view Ashby's oeuvre as a sustained and systematic attempt to elucidate the origin of the capacity to survive by means of adaptive behavior (cf. Cariani, 2009). He insisted that what it means to survive, and therefore to be living, can be captured by the general systems concept of stability. Before we go into the details of his proposal, it is important to remember the revolutionary ambition that this turn toward a mechanistic approach entailed (cf. Husbands et al., 2008). The advances in cybernetics appeared to put within reach the historic opportunity of becoming a Newton of the sciences of life and mind. It was the time of an unprecedented generalization of the scope of scientific explanation by appealing to the framework of general systems theory, which enabled one to abstract away from the intractable mess of bio-chemical complexity:

> It is, I think, one of the substantial advances of the last decade that we have at last identified the *essentials* of the "machine in general." Before the essentials could be seen, we had to realize that two factors must be *excluded as irrelevant*. The first is "materiality"—the idea that a machine must be made of actual matter …. Also to be excluded as irrelevant is any reference to energy, for any calculating machine shows that what matters is the *regularity* of the behavior—whether energy is gained or lost, or even created, is simply irrelevant. (Ashby, 1962, pp. 260-261)

This founding assumption of the irrelevance of material and energetic considerations for the *machine* in general and, consequently, for the *living machine* in particular, sets the stage for most of cybernetic research, including Ashby's. Indeed, there must have been a sense of elation when the invention of feedback mechanisms, telecommunication systems and calculating machines demonstrated conclusively that life-like behavior could be understood rigorously and formally, even without making any reference to the messy details of its physical realization.[5] To see the potential power of this approach when it is applied to biology, it is worthwhile quoting one of Ashby's illustrative examples at length. The case is of a mouse being chased by a cat:

> The concept of "survival" can thus be translated into perfectly rigorous terms .... The various states *(M* for Mouse) that the mouse may be in initially and that it may pass into after the affair with the cat is a set $M_1, M_2, ..., M_k, ..., M_n$. We decide that, for various reasons of what is practical and convenient, we shall restrict the words *"living* mouse" to mean the mouse in one of the states in some subset of these possibilities, in M_1 to M_k say. If now some operation C (for cat) acts on the mouse in state M_i, and C(M_i) gives, say, M_2, then we may say that *M* has "survived" the operation of C, for *M*2 is in the set $M_1, ... M_k$.
> If now a particular mouse is very skilled and always survives the operation C, then all the states $C(M_1), C(M_2), ..., C(M_k)$, are contained in the set $M_1, ..., M_k$. We now see that this representation of survival is *identical* with that of the "stability" of a set (S. 5/5). Thus the concepts of "survival" and "stability" can be brought into an exact relationship; and facts and theorems about either can be used with the other, provided the exactness is sustained.
> The states M are often defined in terms of variables. The states $M_1, …, M_k$, that correspond to the living organism are then those states in which certain **essential variables** are kept within assigned ("physiological") limits. (Ashby, 1956, p. 197)

Accordingly, if the survival of a living being is identical to the stability of a dynamical system, the central problem is how a system could maintain its essential variables (i.e., to keep on living) when confronted with external perturbations. The problem of life thus becomes the problem of adaptation. Ashby's first important move in this respect

---

5. It is important to note that this has turned out to be an extremely costly shortcut for science. In biology it has diverted attention from the metabolic existence of the organism in favor of the pure genetic code, and in cognitive science immense resources have been invested in creating its foundations in symbolic artificial intelligence. In business terms these trends have been productive (e.g., genetic engineering and information technology), but science has yet to recover from their overbearing influence. As we will see in the next sections, even the autopoietic tradition was confused by the prevalent drive toward abstraction.

was to suggest that the biological concept of adaptive behavior is equivalent to the systemic concept of stable equilibrium:

> It is suggested here that adaptive behaviour may be identical with the behaviour of a system in stable equilibrium, and that this latter concept may, with advantage, be substituted for the former. The advantages of this latter concept are that (1) it is purely objective, (2) it avoids all metaphysical complications of "purpose," (3) it is precise in its definition, and (4) it lends itself immediately to quantitative studies. (Ashby, 1940, p. 483)[6]

Ashby's illustrative example of a system in stable equilibrium is a cube resting on one face. If I perturb it slightly, it will tilt and then return to its original position. Accordingly, we can conceive of the cube's automatic return to its initial state in reaction to external perturbation as being a form of 'adaptive behavior', whereby the stability of the system is maintained. To be sure, Ashby is aware of the limitations of this analogy and is therefore quick to emphasize that "'stable equilibrium' does *not* entail immobility" and that the concept equilibrium is essentially a dynamic one (Ashby, 1940, p. 479). For instance, if I happen to apply vibration to the table on which the cube is resting, then it will be forced to cope with disturbances in a continuous manner, and throughout this ongoing flux of motion it will nevertheless remain in stable equilibrium. The cube can therefore be stable and mobile at the same time. Moreover, by emphasizing the dynamic nature of the concept Ashby wants to highlight the fact that we need to interact with the cube and provoke a response before we can determine its specific kind of equilibrium. According to Ashby the cube's response to perturbation is therefore only different in degree, and not in kind, from the adaptive behavior of a living being.

However, even despite these qualifications it is still the case that, due to the abstractions from material and energetic processes unfolding within the organism, living is portrayed as being essentially passive and static in its natural state. A living being is roused from this inherent stability-as-stasis only contingently, namely when it has been perturbed by external events: living is thus nothing but a latent capacity for conservative reaction. Of course, Ashby is quick to point out that in most practical cases, when we are "talking of a stable system we are implicitly assuming that we are dealing with deviations from this neutral point" (Ashby, 1940, p. 480). Nevertheless, this does not negate the fact that the intrinsic tendency of a stable system is to be at absolute rest, and that the locus of activity, if any should occur, is therefore placed entirely on the side of the environment.

This brief analysis has already shown that there is a significant disagreement between the autonomous behavior displayed by living beings, which happens on the basis of continual metabolic needs, and the concept of stable equilibrium which

6. Note that Ashby provides a systemic definition of adaptive behavior, i.e., of the survival mechanism by which a living system maintains its identity, which is explicitly devoid of the notion of purpose. Might this be an unreferenced inspiration for Maturana and Varela's (1973, p. 86) subsequent insistence that "living systems, as physical autopoietic machines, are purposeless systems"? Of course, this rejection of teleology was also very much a part of the cybernetics project as a whole (e.g., Rosenblueth et al., 1943).

reduces behavior to mere passive reactivity to external events. Ashby's account of biological self-maintenance has committed him to conceiving living movement according to an outdated conception of object motion. In other words, moving inanimate objects remain motile (unless impeded by external forces), but living movement essentially tends toward stasis (unless moved by external forces). Thus, while the mechanistic philosophy of Galileo and Newton had freed the movements of matter from the mysterious hands of the "unmoved mover," and thereby placed it closer the dynamic activity of life, Ashby in turn binds the movements of life to the whims of an external environment, and thereby places it closer to a pre-scientific view of material motion. His dynamical framework of adaptive behavior has implicitly turned Aristotle's philosophy of motion on its head: Matter is intrinsically dynamic and life is intrinsically static.

That there is little possibility in Ashby's conception of life to account for an inherently dynamic form of existence is further underlined by his disregard for the potential role of temporality. In particular, he seemed to have been unable to contextualize his viewpoint in relation to the role of different timescales. The possibility that a limited lifespan of an unstable system is not necessarily a problem as long as there is indeed some span of time to be lived, does not appear to have crossed his mind:

> Finally, there is one point of fundamental importance which must be grasped. It is that stable equilibrium is necessary for *existence*, and that systems in unstable equilibrium inevitably destroy themselves. Consequently, if we find that a system *persists*, in spite of the usual small disturbances which affect every physical body, then we may draw the conclusion with absolute certainty that the system must be in stable equilibrium. This may sound dogmatic, but I can see no escape from this deduction. (Ashby, 1940, p. 482)

Ashby neglects the transient temporal distance between life and death: On his view it does not matter if an unstable organism manages to cope with its instability for seconds, days or even years—strictly speaking, it can only be said to exist if it is formally a system in stable equilibrium. Accordingly, there is simply no place in his cybernetic approach for a precarious type of existence that is inherently unstable, but which can effectively delay the occurrence of its final disintegration for the duration of a normal lifespan.

The mechanistic cybernetics in which Ashby specifies biological phenomena leads to a conception of (i) living as stasis and, more specifically, (ii) living motility as a degenerate form of physical movement, and (iii) adaptive behavior as a mere passive reaction. This cybernetic attempt to model biological phenomena in a purely formal framework has bypassed concerns whether there might not be a more appropriate way of distinguishing the phenomena as material processes (for instance, one in which activity and decay are intrinsically tied to physical realization as a special type of dissipative structure).

3.2 The Origin of Adaptive Behavior in Ultrastable Systems

It might be argued that this critical analysis has been unfair to Ashby because we have not yet addressed what many see as his greatest achievement, namely the specification of the concept of ultrastability and its physical implementation in a machine called the *homeostat* (cf. Cariani, 2009). Indeed, since this concept is still influential in enactive cognitive science (e.g., Di Paolo, 2003), and is the only aspect of Ashby's work that is explicitly cited in Maturana's subsequent biology of cognition (cf. Maturana, 1970, p. 25), it is important to be clear about its meaning.

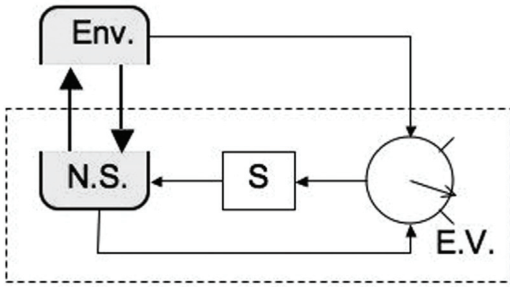
Ashby's work in this area is guided by the illustrative question: Why does the burnt kitten avoid the fire? In order to avoid potential confusion it is important to point out again that Ashby is not interested in the evolutionary origins of reflexes, but in the manner in which it is possible for a system to learn adaptive behavior during its lifetime. His well-known answer is: because the kitten, as an organism with a nervous system, is an *ultrastable* system (Ashby, 1952). Let us unpack this answer in more detail.

In order to better explain the cybernetic mechanism underlying adaptive behavior Ashby makes what could be considered one of his most important conceptual innovations: He introduces a mathematical notion that specifies the conditions under which a system is said to *break*. Simply put: "a break is a change of organization" (Ashby, 1947, p. 50). The occurrence of such a break can, for example, be formalized as a change in the activation of a step-function in a more encompassing system, since it thereby determines how one sub-system is transformed into another. With the addition of this novel concept Ashby now has the means for specifying the origins of adaptation as a cybernetic process:

> Thus, suppose we start with a great number of haphazardly assembled machines which are given random configurations and then started. Those which are tending towards equilibria will arrive at them and will then *stop there*. But what of the others, some of whose variables are increasing indefinitely? In practice the result is invariable—something breaks. After a break the organisation is changed, and therefore so are the equilibria. This gives the machine fresh chances of moving to the new equilibrium or, if not, of breaking again. … This may be stated in the approximate but arresting form: dynamic systems stop breaking when, and only when, they reach a state of equilibrium. And since a break is a change of organisation the principle may be restated in equivalent form: if we allow breaks to occur, then all dynamics systems change their organisations until they arrive at a state of equilibrium. (Ashby, 1947, p. 55)

In the context of an assumed equivalence between living and stability, this theory is an elegant solution for the mechanistic origin of adaptive behavior. It powerfully illustrates that, in contrast to the founding assumption of symbolic AI, the appearance of what looks like intelligent behavior to an observer does not necessarily require equally intelligent internal mechanisms (Di Paolo, 2003). The precise technical details of its implementation do not need to concern us here, but we can note four essential aspects. We need (1) a *physical system* which is capable of interaction with (2) its *environment* such that this perturbs the value of (3) an *essential variable* with a specified boundary of viability which, when this boundary is exceeded, triggers (4) a *selector* that specifies a break in the system such that it randomly changes its organization. This general situation, as applied to the nervous system, is illustrated in fig 1.

Figure 1: Schematic of Ashby's Notion of Ultrastability



The dashed line delimits the boundary of the organism which includes at least one essential variable (E.V.) which is linked to a selector of random step-functions (S). The behavior of the organism arises as the nervous system (N.S.) interacts with a part of the environment (Env.) by means of its effector and receptor surfaces. The state of the essential variable can be perturbed by external and internal activity. If its limit of viability is exceeded then the selector triggers a random change in the organization of the nervous system. This in turn leads to a change in behavior of the sensory-motor loop, which will potentially expose the state of the essential variable to a different set of perturbations. This loop of re-organization continues until a new stable configuration has been found.

Note that our criticisms cannot be satisfactorily addressed by the notion of ultrastability, although it does introduce the possibility of transient dynamic instability in a subpart of the homeostatic system. Once again it is only when the external environment pushes an essential variable out of bounds that the system does anything at all. Accordingly, even the famed homeostat is an essentially passive system whose natural state is to indefinitely remain in stability-as-stasis. To be sure, Ashby's theory ultrastability does not conceive of adaptive behavior as purely reactive, in the sense of a behaviorist psychology. After all, the system's behavior depends on the existence of internal state. Nevertheless, like behaviorism it locates the causal locus of activity in the environment alone: On this view it is only because of contingent external factors that an organism displays behavior at all.

3.3 Ashby's Critique of Self-organization

It might be thought that we can rescue Ashby by appealing to the cybernetics of self-organizing systems, especially since there is indeed a sense in which such systems could be considered self-individuating (Collier, 2004). However, this option is quickly dispelled by Ashby's systematic rejection of the concept of self-organization. In brief, his view is that any isolated component that was shown in figure 1 cannot be called *self*-organizing because its operation depends on being embedded within the complete configuration, and the latter configuration cannot be meaningfully called self-

organizing either, since as a whole it is an unchanging system that is simply organiz*ed*. In other words, in order for any system to be properly called self-organiz*ing* we would require its organization to include a function that not only determines the next state of the system, but where that function itself, and thereby the system's defining organization, is determined by the system's current state. Only if there is such a peculiar form of self-dependence within the system could the system itself change its organization in relation to itself.

However, this organizational self-dependence appears to entail an illogical mathematical self-reference, and is therefore unacceptable to Ashby: "To allow *f* to be a function of the state is to make nonsense of the whole concept" (Ashby, 1962, p. 268). For him, "it is clearly illogical to talk of *f* as being a function of *S*, for such talk would refer to operations, such as $f_a(b)$, which cannot in fact occur" (Ashby, p. 268). This is why the selector of the random step-function in the homeostat is implemented as a function that is not dependent on the state of the system that breaks. Accordingly, Ashby pronounces:

> Since no system can correctly be said to be self-organizing, and since use of the phrase "self-organizing" tends to perpetuate a fundamentally confused and inconsistent way of looking at the subject, the phrase is probably better allowed to die out. (Ashby, 1962, p. 269)

Note that while Ashby's rejection of the notion of self-organization is in some respects justified (it is often misused and notoriously hard to define), his dismissal of using self-reference seems premature. The fact that his particular conceptualization turned out to be an illogical failure should not have been taken as a universal demonstration. For instance, since the equations of his system always affect each other instantaneously, the existence of temporally extended self-reference is simply collapsed into a paradoxical form of immediate self-coincidence. On this view, the apparent impossibility of self-organization appears to be a consequence of the abstraction underlying his systems theory because, as we noted before, it neglects the role of material properties and temporality.[7]

We now have a more detailed understanding of the constitutive assumptions underlying Ashby's way of thinking about biology. This has given us an initial indication of the kind of intellectual climate with which systems biology was infused during the time when the young Maturana started his research. In the context of the problems outlined in sections 1 and 2, it is particularly noteworthy that cybernetics explicitly resists the formulation of a principled distinction between the living and the non-living: both are treated in terms of a control theory that focuses on stability-as-stasis. The next step in our investigation will be to determine more precisely in what manner this general systems framework has been an influence on Maturana's notion

7. Similarly, in neuroscience the conduction delays inherent in the nervous system were long thought to be negligible side-effects contingent on its material realization, but in the end these might actually turn out to play an essential role in its capacity for self-organization (cf. Buzsáki, 2006).

of the circular organization of the living, which was a direct precursor to the concept of autopoiesis.

4. Maturana's Biology of Cognition

In this section we will compare Ashby's and Maturana's approach to biology in terms of some methodological and epistemological considerations, which show the general extent of their overlap (e.g., application of general systems theory to the domain of theoretical biology), as well as some crucial differences (i.e., the role of the observer). Once these similarities and differences have been clarified we then proceed to analyze Maturana's notion of the circular organization of the living in more detail. The guiding question is to determine whether this notion has managed to avoid some of the troubling premises that constitute Ashby's cybernetic framework.

4.1 Methodological and Epistemological Considerations

Let us recall the starting point for Ashby's investigations. He was dissatisfied with the vague terminology that had until then been used in science to describe and explain biological phenomena, in particular adaptation. Instead he proposed to model these phenomena in terms of the mathematics of dynamic systems theory which, as operational models, at the same time provided him with a formulation of the *generative mechanisms* that could give rise to these phenomena. Ashby was therefore a firm believer in what has been called the synthetic method (Cordeschi, 2008). Similarly, Maturana was convinced that a rigorous scientific explanation had to go beyond a mere description of the data obtained by an observer's interaction with a system to a generative theory which can give an account of the genesis of the phenomenon being investigated:

> Accordingly, the full explanation of the organization of the nervous system (and of the organism) will not arise from any particular observation or detailed description and enumeration of its parts, but rather like any explanation, from the synthesis, conceptual or concrete, of a system that does what the nervous system (or the organism) does. (Maturana, 1970, p. 47)

The first methodological point of agreement between Ashby and Maturana is therefore that biological phenomena are scientifically best explained in an operational manner by the proposal of generative mechanisms. Moreover, they both emphasize the point that if we want to study a natural phenomenon like a living system we are indeed forced to interact with it in some manner first. Only through interaction can we get a response profile which can become the target for our proposed generative mechanism.

From this perspective we can get a better sense of what Ashby meant when he said that the concept of equilibrium was essentially a "dynamic one" (Ashby, 1940, p. 479). In other words, it was not the actual mechanism to which the concept was referring that was said to be dynamic, but rather the interactive manner in which the concept itself becomes realized: "If we just look at the three bodies on our table and do nothing with them the concept of equilibrium can hardly be said to have any particular meaning. It is only when we disturb the bodies and observe their subsequent reactions that the concept develops its full meaning" (Ashby, p. 479).

The synthetic methodology consists of (1) observer-system interaction in order to obtain (2) descriptions of the system's response profiles for (3) the specification of a generative mechanism that could explain that profile. It is important to note that the (inter-)active role of the observer in this methodology entails some epistemological considerations about the status of a scientific explanation. In effect, it leads to Maturana's famous dictum: *Anything said is said by an observer* (Maturana, 1970b, p. 4). This still remains a much needed warning against the all too common confusion between knowing by interaction and knowing by representation. But before we look at this issue in more detail there is another confusion of which we must beware, namely that observers have a different perspective than the perspective of the observed:

> Anything said is said by an observer. … The observer beholds simultaneously the entity that he considers (an organism, in our case) and the universe in which it lies (the organism's environment). This allows him to interact independently with both and to have interactions that are necessarily outside the domain of interactions of the observed entity. … This latter domain lies within the cognitive domain of the observer. (Maturana, 1970, p. 8)

In other words, we observers can interact with the observed system and its environment from a perspective that is external to them, namely so as to make observations about the relations between the two, but our knowledge of these relations is constituted by our own interactions and do not pertain to the internal operations of the system itself. In other words, the system-environment correlations that we can establish as external observers (e.g., a representation) cannot explain why the system behaves in its environment as it does, because that knowledge cannot be a part of its internal generative mechanism. The generative mechanism operates on the inside of the target system and therefore does not have access to operations that are based on an external observation of the relations which this mechanism realizes.

Ashby was also aware that the observer played a crucial role in the way we come to understand observed systems. On the one hand, the models of adaptation which he had proposed clearly embody the epistemological consideration that was later formulated by Maturana: The realization of the model's internal operation and of the resulting adaptive behavior does not depend on any knowledge that an external observer might have about the model's relation to its environment. This is nicely demonstrated by the design of the homeostat. On the other hand, Ashby also tried to formalize the way we measure the entailment between different parts of an observed system by referring to the point of view of the observer. For example, we can try to measure the entailment of part A on part B in terms of the constraint imposed on events occurring at B in relation to events occurring at A. But notice that this measure can only have meaning in terms of the uncertainty of the observer, because it is this uncertainty which determines the total space of possibilities for events at B against which the constraint of the actual states can be compared:

> 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 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, p. 258)

This conclusion of the relational nature of systemic knowledge deals a considerable blow to any theory of knowledge that is based on the naïve assumption that observing is about internally representing observer-independent entities. Interestingly, the fact that this is an indication of the interactive and hence intrinsically relational nature of all knowledge 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, p. 257). But in what sense are those sciences more objective? Did the advances in 20th century physics not once and for all demonstrate the interactive and relational nature of knowledge? Nevertheless, it looks like Ashby might have had quite a naïve objectivist conception of the work of the “hard sciences.” Perhaps Newtonian mechanics represented for him the ideal scientific situation which he tried to replicate in his attempts to devise concepts of biological phenomena that are “purely objective” (e.g., Ashby, 1940, p. 483).

Ironically, the very insights which Ashby developed in his work on the organization of adaptive systems would later be used to soundly reject the possibility of attaining pure objectivity in any form of knowledge, including that of the hard sciences. As Maturana has repeatedly emphasized, the representational validity on which the objectivity of knowledge depends is not supported by the circular organization of the nervous system:

> Notions such as embodiment of representation express the correspondence that the observer sees between relations, or sets of relations, and different states of activity of the nervous system, and, as such, lie in his cognitive domain. … they do not characterize the nature of the functional subordination of the nervous system to its own states. This subordination is that of a functionally closed, state determined, ultrastable system, modulated by interactions [Cf. Ashby, 1960[8]]. (Maturana, 1970, p. 25)

In other words, Maturana interprets Ashby’s work to show that the nervous system is a functionally closed system that does not require internal representations of an external world as part of its operations in order to generate adequate conduct. Here we thus have a fundamental epistemological shift from Ashby’s cybernetics to Maturana’s biology of cognition, which inaugurates the radical constructivist turn in biology.

4.2 The Circular Organization of the Living

The next task will be to find out more precisely to what extent Maturana’s interpretation of ultrastability makes a difference in terms of the actual generative mechanisms which he proposes to explain biological phenomena. The most notable

8. Ashby (1960) refers to the second edition of Ashby (1954).

conceptual innovation in Maturana's early work is that of the circular organization of the living. We will examine this concept in detail since it is a direct precursor of autopoiesis and will therefore give us important clues about the latter concept's relationship to the work of Ashby.

First of all, we can note that both Maturana and Ashby posit the existence of a machine as the starting point of their investigation. The point of this concept is that we are dealing with a deterministic system whose behavior is defined by its configuration and its current state, that is, it is a state-determined system (Ashby, 1947, p. 48). Using slightly different terminology Maturana has always insisted that "a *structure determined system* is a system such that all that takes place in it, or happens to it at any instant, is determined by its structure at that instant" (Maturana, 2002, p. 15). Note, however, that the concept of state- or structure-determinism does not get us far in defining the organization of the living since we can treat all material (physical and biological) entities as belonging to the class of these systems. This is because purely reactive systems exist as abstract mathematical entities only (e.g., a system of non-recursive functions). The general applicability of the concept of state- or structure-determinism robs it of all explanatory power when it comes to accounting for what is characteristic of the living.[9]

A more exclusive concept is clearly needed if Maturana is to succeed in delimitating what is special about the class of living systems. We have already seen that Maturana (1970, p. 25) accepts Ashby's characterization of the organization of the nervous system in terms of the concept of ultrastability. The next crucial step is that Maturana (p. 50) generalizes this conception of the nervous system by applying it to the organization of a living system as a whole: "A living system is not a goal-directed system; it is, like the nervous system, a stable state-determined and strictly deterministic system closed on itself and modulated by interactions not specified through its conduct." As a consequence of this shift, the target system's operations no longer consist neurological activity, but of metabolic processes: "this circular organization constitutes a homeostatic system whose function is to produce and maintain this very same circular organization by determining that the components that specify it be those whose synthesis or maintenance it secures" (Maturana, p. 9). Thus, whereas Ashby was only concerned with the *maintenance* of a stable organization, as demonstrated by the homeostat, Maturana is adding the idea that the operation of living systems can also be characterized as the *production* of a stable organization. More specifically:

> The living organization is a circular organization which secures the production or maintenance of the *components* that specify it in such a manner that the product of their functioning is the very same organization that produces them. Accordingly, a living system is an homeostatic system whose homeostatic organization has its own organization as the variable that it maintains constant through

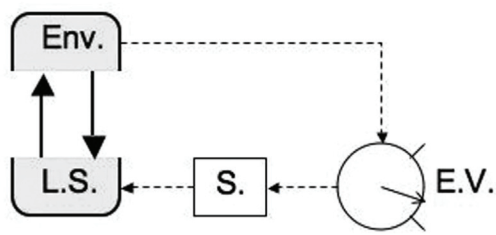
9. The concept of state- or structure-determinism should not be confused with the notion of biological autonomy. This point seems to have been lost on Maturana who proceeded to build a whole worldview on the fact that physical systems have state and are deterministic (as did Ashby, cf. Pickering, 2009, p. 225).

> the production and functioning of the *components* that specify it, and is defined as a unit of interactions by this very organization. (Maturana, 1970, p. 48)

The circular organization of the living appears to be a reference to metabolic activity: It consists of organized processes of component production whose product is the very same organization that produces them. It has often been thought that Maturana's notion of the circular organization must therefore denote a reciprocal relation between the local and global levels of description, namely a circular relationship between the components (the processes of production) and the system as a whole (its organization). This is the sense of reciprocal causality that Kant (1987) had proposed long ago. But this is apparently not at all what Maturana had intended to convey, at least not according to his most recent retrospective interpretation:

> I am not saying, as Kant and others have said, that the parts exist for the whole and the whole exists for the parts. That is a commentary of an observer in relation to what he or she thinks; it does not reveal what happens in the molecular dynamics of a cell or organism. … Molecules interact with other molecules in a way in which the result of their interactions does not participate at any moment in the genesis of that result. … The components of any system exist as local entities only in relations of contiguity with other components, and any relation of the parts to the whole proposed by an observer can only be a metaphor for his or her misunderstanding, and has no operational presence. (Maturana, 2002, p. 9)

Figure 2: Schematic of How Maturana Can Generalize Ashby's Notion of Ultrastability to the Organization of the Living



The living system (L.S.) gives rise to interactions with its environmental niche (Env.) such that the outcome of those interactions leads to the maintenance of the identity of the living system that gives rise to those interactions. If an essential condition is not maintained at any point in time, that is, if the essential variable (E.V.) exceeds the boundaries of viability, then the identity of that particular living system, as that living system, disintegrates, that is, its organization is changed by a random step-function (S.).

According to Maturana's own interpretation of his work the circularity of the living must exclusively happen on the level of components without any reference to the system as a whole, and without any reference to the result of the component interactions. But what kind of circularity can exist within the domain of components that could fulfill these conflicting criteria? Given Ashby's rejection of material and energetic considerations, it may well turn out that this question is asking the impossible; Maturana may be using a conceptual framework that fails to adequately capture the phenomenon of material self-production. But let us give Maturana the benefit of the doubt and first consider some other interpretations of the concept of circularity which follow from conceiving a living system as an ultrastable machine. This generalization of the concept of ultrastability is illustrated in figure 2.

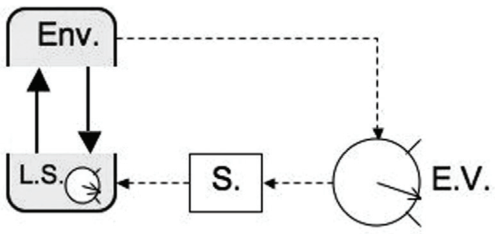
At first sight it might look like figure 2 includes two types of circularities that could fit Maturana's demand of a circular organization of the living that does not appeal to any kind of reciprocal causality:

1. We have the homeostatic loop that controls whether the identity (organization) of the system is changed. However, this loop is only activated when the living system exceeds its boundary of viability and thus ceases to be a living system. While it can be used as a generative mechanism that explains the spontaneous origin and death of a living system, it cannot be the kind of circularity we are looking for to define the living as such since the loop is strictly external to any living system that might arise.

2. A living system interacts with its environment in a circular manner via its sensory-motor loop. The idea here is that how an organism gets perturbed depends on what it does, and what it does depends on how it gets perturbed. This is indeed an important kind of circularity in cybernetics because it can ground a theory of perception and knowledge that is not founded on representations (cf. von Foerster, 1976). However, some version of this parametrical interaction would remain even if the living system disintegrates and becomes a different kind of structure-determined system since that new system would still be coupled to other systems. Structural coupling is a property of all material (physical and biological) systems and cannot help us to define the specific circularity of the living organization.

These two circularities are compatible with Maturana's demand that there is no reciprocal interaction between the components and the system as a whole, but they fail to capture the notion of self-production. Are there any other possibilities that could satisfy both of the requirements? There is one final possibility which can be derived from Maturana's claim that "a living system is an homeostatic system whose homeostatic organization has its own organization as the variable that it maintains constant" (Maturana, 1970, p. 48). Given Maturana's reservations against reciprocity we can ask: How can the organization of the system also be a component variable of that same system? Does this not violate the condition that the circularity must remain on the level of components? Note that this is only a problem if we assume that Maturana is here making an operational statement that pertains to the domain of components. If, on the other hand, we interpret his statement as part of a functional discourse, that is, an external commentary that is based on observations pertaining to the system-environment relationship, then the reciprocity disappears. The living system is observed to maintain a certain variable constant and, from our external perspective, we can tell that this variable is linked to the existence of its identity as a whole. But this correlation does not entail that the variable represents the system's identity to the system itself. This possibility is illustrated in figure 3.

Figure 3: Figure 2, Updated



The schematic shown in 2 has been updated so that the living system now has its own organization as an essential variable.

Maturana does not need a representative relation because of the way in which the living system is embedded in its environment: either it operates with its essential variable in a manner such that it happens to maintain its identity, or it simply disintegrates. In other words, the link between the essential variable and the organization of the system as a whole is contingent on the fact that the living system exists in the first place. No epistemic access to an external state of affairs is necessary. In fact, the correlation itself is entirely inaccessible to the operations of the system, as is illustrated by Maturana's metaphor of the submarine:

> What occurs in a living system is analogous to what occurs in an instrumental flight where the pilot does not have access to the outside world and must function only as a controller of the values shown in his flight instruments. … In terms of their functional organization living systems do not have inputs and outputs, although under perturbations they maintain constant their set states. (Maturana, 1970, p. 51)

Of course, this metaphor should not be taken too seriously, especially because it would otherwise commit Maturana to the problems associated with the postulation of an internal homunculus that does all of the work. A living system might be an observer, but there is no observer inside the living system. In any case, with this functionalist interpretation of the claim that the system has its own organization as the essential variable we satisfy Maturana's rejection of reciprocity. But now we are once again faced with the problem that Maturana no longer has any principled manner of distinguishing living from non-living systems. As we have seen, even the homeostat maintains its own organization in precisely this manner. Ashby would certainly be happy with this conclusion since he does not think that there is any difference in kind between living and non-living systems, but for Maturana, who clearly wants to demarcate the living in terms of their processes of self-production, this is unacceptable. It seems that we have hit another dead end in our search for a consistent meaning of the circularity of the living organization.

4.3 Ashby's Concept of Closure

At this point it is helpful to have another look at Ashby's work. If the circularity has to do with the way that a living system is "closed on itself" (Maturana, 1970, p. 50) perhaps we should understand it in terms of the systemic concept of closure? Ashby provides us with the following definition:

> *Closure.* When an operator acts on a set of operands it may happen that the set of transforms obtained contains no element that is not already present in the set of operands, i.e. the transformation creates no new element. … When this occurs, the set of operands is **closed** under the transformation. (Ashby, 1956, p. 11)

The meaning of this concept of closure is best illustrated by a simple example: The set of all Boolean states (i.e., true and false) is closed under the transformation of negation. The negation of true is false and the negation of false is true. The operation of negation never results in any element other than true or false and is therefore always applicable again to the operands it produces. There is in principle no limit to the number of iterations that the operation can be repeated.[10] Operational closure is therefore a desirable property in the domain of systems engineering because of its inherent stability, whereas "the unclosed transformation is thus like a machine that takes one step and then jams" (Ashby, 1956, p. 18). We already know that for Ashby the concept of closure is equally applicable to organisms: the set of all states during which a living system is alive is closed under the transformation of living. In this sense the concepts of biological survival and closure (a stable set) can be brought into an exact relationship (Ashby, p. 197). The property of operational closure guarantees that the identity (organization) of the living system is well defined and that it will persist throughout its operation.

If we now return to Maturana's notion of circularity in relation to the organization of the living, it is clear that he has Ashby's concept of closure in mind:

> A living system defines through its organization the domain of all interactions into which it can possibly enter without losing its identity, and it maintains its identity only as long as the basic circularity that defines it as a unit of interactions remains unbroken. (Maturana, 1970, p. 9-10)

> The closed nature of the functional organization of the nervous system is a consequence of the self-referring domain of interactions of the living organization; every change of the state of the organism

10. Future researchers may also want to trace Maturana's notions of structure determinism and the circular organization of the living to his collaboration with Warren McCulloch (e.g., Lettvin, Maturana, McCulloch, & Pitts, 1959). In this context it is especially noteworthy that McCulloch and Pitts (1943) already had a formal concept of nets with circles, by which they essentially meant Boolean systems closed under recursive transformation. But Maturana's work appears to be more closely related to Ashby's because the latter combines the concept of closure with the belief that "Man does not think logically—he thinks dynamically" (Ashby, 1951, p. 7).

> must bring forth another change of state, and so on, recursively, always maintaining its basic circularity. (Maturana, 1970, p. 25)

The possibility that Maturana's notion of the circular organization of the living is merely an appropriation of the concept of systemic closure is further supported by the problems which would be entailed by this identification. For instance, operational closure is a purely logical property that does not depend on any concrete realization or activity: A mathematical system that is never put to use can still be a closed system. It should therefore come as no surprise that recent work has encountered difficulties in using the insights of the autopoietic tradition to devise a biological account of embodied action. In terms of precariousness, moreover, we want a definition of the living for which the possibility of disintegration (death) is a necessary condition of existence. But strictly speaking the concept of systemic closure provides us only with the opposite: It rules out in principle that the system can potentially transform itself into another (previously undefined) system. On this view, there is simply no room for autonomously generated possibilities of becoming, development, and death.

Most importantly, if Maturana's circularity of the living organization is essentially the same as Ashby's notion of closure under the transformation of living (with or without ultrastability), then we are again faced with the same fundamental problem as before. Ashby would have been happy to point out that there are many trivial examples of operationally closed systems that are not instances of living systems. In fact, any complete dynamic system or machine will satisfy the mathematical criteria required for closure, which means that this type of circularity cannot be used as definitional of what makes a living system a different kind of system. Ashby believed that this was actually a key advantage of his cybernetic approach, but for Maturana it threatens to dissolve the domain of biology into that of general systems theory.

4.4 Competing Motivations: From Lobotomy to Love

Before we move on to consider Maturana and Varela's concept of autopoiesis in the next section, it is of interest to emphasize just how unsuitable Ashby's work is to account for biological autonomy and the self-assertion of the living. Despite Maturana's best efforts to the contrary, it is simply impossible to get a grip on these phenomena in terms of a cybernetic framework focused on stability-as-stasis. Of course, Ashby never intended his work to do what Maturana had wanted to achieve, but the extent of the discrepancy is still remarkable. Perhaps this incompatibility can be partly understood in relation to Ashby's professional background.

Interestingly, Ashby's inspiration for conceiving the notion of ultrastability appears to have been diametrically opposed to Maturana's own motivations. It is not known whether Maturana was aware of it or not, but Ashby was active in a very mechanistic school of psychiatry, where he was closely involved in the development of brutal techniques for intervening in the brains of his patients, including lobotomy, chemical shock, and electroshock therapy (a.k.a electroconvulsive therapy, or ECT).

That there is in fact a deep relationship between his discoveries in cybernetics and his psychiatric practice is made clear in his notebooks, as Pickering has shown:

> In November 1958, he made some notes on how his psychiatric thinking was informed by them. Under the heading 'blitz-therapy,' he began: 'Treating a patient is an imposition of the therapist's will on the patient's; it is therefore a form of war. The basic principles of war are therefore applicable.' Various practical morals followed from this, including 'Not waiting to "understand the patients pathology" … but hitting hard & seeing what happens, and, as a concrete suggestion, use of techniques in combination, simultaneously. E.g. LSD, then hypnosis while under it, & ECT while under the hypnosis'. (Pickering, 2009, p. 223)[11]

Given Ashby's perspective on therapeutic intervention it becomes difficult not to see the concept of ultrastability, that is, the application of extreme external perturbations leading to a series of system breaks eventually resulting in equilibration, as a cybernetic explanation of the efficacy of lobotomy and electroshock therapy. Nothing could be further removed from the biology of autonomy, responsibility, and love, which Maturana will later try to develop on these same conceptual foundations (e.g., Maturana, 1988).

The upshot of this comparative analysis is that both Ashby and Maturana make use of the same underlying dynamical systems framework for their thinking, but that Maturana tries to use it in a different way than Ashby had originally intended it. In particular, Maturana goes against Ashby when he tries to apply the concepts of closure and ultrastability in a way that would demarcate the living from the non-living in terms of their material self-production. This is a distinction that Ashby never intended to achieve with these systemic concepts; on the contrary, he explicitly insisted that his principles are equally applicable to living as to non-living systems, with the difference being merely a matter of degree and not of kind. It is therefore highly doubtful whether the circularity of the living organization, understood as systemic closure under transformation, can be made to serve as a useful definition of the living.

5. Maturana and Varela's Concept of Autopoiesis

The analysis of the previous section has shown that Maturana unwittingly characterizes the organization of living systems in essentially the same non-distinct manner as was proposed by the cybernetics of Ashby, although Maturana tries to enhance this general systems approach with an interpretation based in constructivist epistemology. Maturana's appropriation of the cybernetic framework in the service of an organism-centered biology is therefore misleading, and the actual significance of Maturana's early biology of cognition lies mainly in its epistemological shift. This conclusion is also supported by Francisco Varela's analysis of this period:

> The direct antecedent to the gestation of autopoiesis is the text that Maturana wrote in mid-1969, originally entitled 'Neurophysiology of Cognition'. [The text was written] in order to clarify what

11. For a fuller treatment of these issues, see Pickering (2010).

> until then he alluded to as the self-referred nature of living beings, and to definitively identify the notion of representation as the epistemological pivot which had to be changed. … This article marks an important jump, and to this day I still believe that it was the indisputable beginning of a turn in a new direction. (Varela, 1996, pp. 411-412)

We agree with Varela's assessment that this early phase of Maturana's work marks an important advance in terms of epistemology, but our analysis has also revealed that in terms of ontology Maturana has retained Ashby's commitment to traditional Newtonian mechanics. The fundamental conceptual building block of the cybernetic framework has simply remained unchanged: intrinsically static (passive and stable) systems.

Given our many reservations about the adequacy of this approach for capturing what is unique about the phenomenon of life, the next task will be to determine to what extent the concept of autopoiesis, which Maturana later developed in collaboration with Varela (1973) as a reformulation of the circularity of the living organization, differs from the cybernetics of stasis. The hope is that autopoiesis, taken literally as *self-production*, will add a novel dimension that takes it beyond Ashby's interest in equilibration.

5.1 The Phenomenon and the Method

What can be noticed in Maturana and Varela's collaborative work is that the descriptions of the essential characteristics of the phenomenon of life have taken on a more refined quality. The central observation that is to be explained is the appearance of *biological autonomy.* And instead of insisting on the importance of stability and conservation there is talk of an organism's *active self-assertion*, and this is recognized as precisely where the central problem of biology is rooted: "Yet, autonomy, although continuously revealed in the self-asserting capacity of living systems to maintain their identity through the active compensation of deformations, seems so far to be the most elusive of their properties" (Maturana & Varela, 1973, p. 73). How can this peculiar nature of life to actively assert itself be accounted for in a scientific manner? Their novel proposal is that "the notion of *autopoiesis is necessary and sufficient to characterize the organization of living systems*" (Maturana & Varela, p. 82).

Maturana and Varela are quite explicit about the terms in which they intend to frame their proposal: "Our approach will be mechanistic: no forces or principles will be adduced which are not found in the physical universe" (Maturana & Varela, 1973, p. 75). Given the many problems that a Newtonian mechanicism faces in the context of an organism-centered biology (e.g., Sheets-Johnstone, 2000; Bickhard, 2008), this does not appear to be a promising start. A more charitable interpretation would be that the appeal to a mechanistic approach reads like a general commitment to some form of scientific naturalism, that is, that no magical notions like those associated with vitalism are allowed. In other words, if we interpret the notion of physical universe in a broad sense, then their approach could indeed include special biological principles that are not to be found in the academic discipline of physics as such. Unfortunately, this is clearly not the case. In fact, their approach does not even care about the basic principles of physics, either.

> It is our assumption that there is an organization that is common to all living systems, *whichever the nature of their components*. … This mode of thinking is not new, and is explicitly related to the very name of mechanicism. We maintain that living systems are machines … and, hence, that [their organization] can be explained as any organization is explained, that is, in terms of relations, *not of component properties*. (Maturana & Varela, 1973, p. 76; emphasis added)

Considering that the main topic will be an attempt to explain biological autonomy in terms of chemical self-production, this might appear like an odd starting point for their work on autopoiesis. The problem is that if you abstract the organization of the living from its components to such an extent that you simply do not need to consider their nature at all (e.g., whether they are material, biological, or social; whether they are organs, processes, or particles; whether they are thermodynamically stable or far-from-equilibrium; etc.), then you end up with a general systems concept that is no longer specifically applicable to the class of living systems. Here again we see the influence of Ashby's style of cybernetics:

> Cybernetics started by being closely associated in many ways with physics, but it depends in no essential way on the laws of physics or on the properties of matter. Cybernetics deals with all forms of behaviour in so far as they are regular, or determinate, or reproducible. The materiality is irrelevant, and so is the holding or not of the ordinary laws of physics. … *The truths of cybernetics are not conditional on their being derived from some other branch of science.* Cybernetics has its own foundations. (Ashby, 1956, p. 1)

Thus, like Ashby before them, Maturana and Varela explicitly limit their framework to a small subset of abstract principles for which materiality is irrelevant. What they are interested in, of course, is to define life in terms of cybernetic concepts, that is, the relational properties of machines such as a "control plant" (Maturana & Varela, 1973, p. 75). And the laws that define and govern the operation of such a machine are given by its organization, a logical form which is independent from how it is concretely realized as a physical structure. The advantages of this methodological choice are clear: If it can be shown that biological phenomena are best understood by means of cybernetic principles, then the domain of biology can no longer be reduced to that of physics. Material and energetic properties of the components alone cannot explain system level properties.[12]

At the same time it should be evident that by retaining Ashby's mechanistic ontology, which identifies the domain of life with an abstract formalism that is devoid of material and energetic considerations, there will be problems in adequately dealing

12. Similarly, the computationalist theory of mind which grew out of cybernetics as well, compare the work of McCulloch and Pitts (1943), has protected itself against the worst excesses of materialist reductionism by claiming that all mental functions bottom out in two-valued logic, no matter the physical basis. Both the computational and the dynamical hypothesis in the cognitive sciences got caught in a similar trend toward systematic abstraction, with only the choice of mathematical framework being different.

with the concrete manifestation of the living as materially embodied beings. We can therefore already expect that Maturana and Varela's mechanistic framework will lead to difficulties when it is applied to the complexities of biological autonomy.

5.2 Defining Autopoiesis

What is autopoiesis? It is surprisingly hard to give a precise answer to this question. The primary literature of the autopoietic tradition gives a variety of definitions with different flavors and emphases,[13] and the secondary literature accordingly abounds with an even wider range of interpretations (cf. Bourgine & Stewart, 2004). It follows that there is no one correct definition of autopoiesis. Whether we accept a definition as valid depends on the context of our investigation. At the very least we must distinguish whether we want a scholarly definition of autopoiesis, that is, a definition that is based on an interpretation of the primary literature, or a scientific definition of autopoiesis, that is, a definition that to the best of our current knowledge captures the self-producing organization of the living. Our aim in this paper is to develop a scholarly interpretation, because this will help to clear up some of the prevailing confusion about the concept. Along the way we will highlight hidden tensions in the traditional definition of autopoiesis that need to be resolved if we want to develop a more scientific concept.

Let us begin by recalling one of Maturana's first attempts to characterize the essential organization that defines a living system: "a living system is an homeostatic system whose homeostatic organization has its own organization as the variable that it maintains constant" (Maturana, 1970, p. 48). In the last section we argued that this statement is just another way of expressing Ashby's idea that we are dealing with a closed system, and that this closure is homeostatically linked to the system's survival (cf. Ashby, 1956). It should therefore come as no surprise that the concept of autopoiesis is first introduced as a special case of homeostasis:

> There are machines which maintain constant, or within a limited range of values, some of their variables. … Such machines are homeostatic machines and all feedback is internal to them. If one says that there is a machine *M*, in which there is a feedback loop through the environment so that the effects of its output affect its input, one is in fact talking about a larger machine *M'* which *includes the environment and the feedback loop in its defining organization*. Autopoietic machines are homeostatic machines. (Maturana & Varela, 1973, p. 78; emphasis added)

This passage confirms that autopoiesis is a peculiar form of homeostasis, where its own organization is the essential variable that it maintains constant (Maturana & Varela, 1973, p. 79). It also throws up a new puzzle: the idea of including feedback loops through the environment in the definition of the machine. One possibility that has recently been suggested by Virgo, Egbert and Froese (in press) is that the traditional concept of autopoiesis actually includes aspects of the organism and some parts of its environment as well. The argument is as follows: since (i) any living being

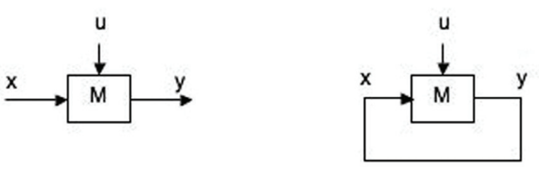
13. This has finally lead to the point where Maturana (2002) and Varela (Weber & Varela, 2002) are actually giving directly opposing interpretations of the concept of autopoiesis. We will return to this issue later.

M necessarily has feedback relations with its environments, and (ii) it is claimed that homeostatic systems only have feedback that is internal to their organization, and (iii) it is also claimed that autopoietic machines are homeostatic machines, it follows that (iv) autopoiesis necessarily refers to the organism-environment unity *M'* as a whole.

This interpretation is supported by the fact that Ashby uses a similar reasoning to argue that an animal and its environment together form a complete machine and that therefore his principle of ultrastability applies: "It should be noted that we now make no essential distinction between the animal and its environment: both contribute to the organisation of the whole, both act dynamically on themselves and on the other, any equilibrium must stabilise both" (Ashby, 1947, p. 56). However, while this approach might have worked for Ashby, it would pose a significant difficulty for Maturana and Varela. For if there is no longer any essential difference between an organism and its environment, then on what basis can we propose autopoiesis as the defining organization of living systems?

There is, however, another possible interpretation of this complicated passage which is also grounded in cybernetics, namely that Maturana and Varela are just making use of yet another way of expressing the idea that living systems are operationally closed systems. More specifically, we already know from Ashby that a system is closed if its operation of transformation only produces outputs that it could operate on in principle. But in order for the system to be actually able to operate on these outputs in practice, we have to introduce a feedback loop from the outputs to the inputs of the transformation (see figure 4). It is only when with this feedback loop is in place that the system has an internal state which can be stabilized, and which determines the outcome of the next operation of the system. On this view, an autopoietic system is a certain kind of feedback system which entails state-determinism and stability.

Figure 4: Schematic of Operationally Open System (Left) and Operationally Closed System (Right)



The open system has an input *x*, an output *y*, a transformation function *M*, and a control parameter *u*. The organization of this system can be defined as: y = M(x, u). The closed system, on the other hand, still has a transformation function *M* and a control parameter *u*, but it no longer has clearly defined inputs or outputs. The variables *x* and *y* are now defined in relation to each other: $x_{t+1} = y$. The organization of this system can now be defined as: $y = x_{t+1} = M(x, u)$. In fact, variable *y* is no longer needed at all in this organization: $x_{t+1} = M(x, u)$. The closed system includes the feedback loop as part of its defining organization.

We now have a better understanding of what Maturana and Varela could have meant by the statement that autopoietic machines are homeostatic machines. They are operationally closed in the sense of feedback systems or recursive functions. However, these kinds of systems are well known in the cybernetic literature (Von Foerster, 1970),[14] and are broadly applicable. They cannot constitute the defining organization of the living. Maturana and Varela acknowledge this by noting that the peculiarity of autopoietic machines does not lie in their homeostatic organization as such, but "in the fundamental variable which they maintain constant" (Maturana & Varela, 1973, p. 78). They then proceed to provide the full definition of autopoiesis:

> An autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through the interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network. (Maturana & Varela, 1973, pp. 78-79)

This is quite a jump indeed. From the cybernetic concept of operational closure, which is already satisfied by as simple a system as $x_{t+1} = M(x, u)$, to the complex notion of a self-producing network of processes of production. It is especially the idea of self-production that adds a new quality to the discussion. So far we have been mainly discussing systems which maintain their identity (organization) by being operationally closed with respect to their transformation. But the identity that is being maintained by these systems is not a product of their operation, rather it is externally defined. Yet in the case of autopoiesis this is precisely what Maturana and Varela suggest:

> It follows that an autopoietic machine continuously generates and specifies its own organization through its operation as a system of production of its own components, and does this in an endless turnover of components under conditions of continuous perturbations and compensation of perturbations. (Maturana & Varela, 1973, p. 79)

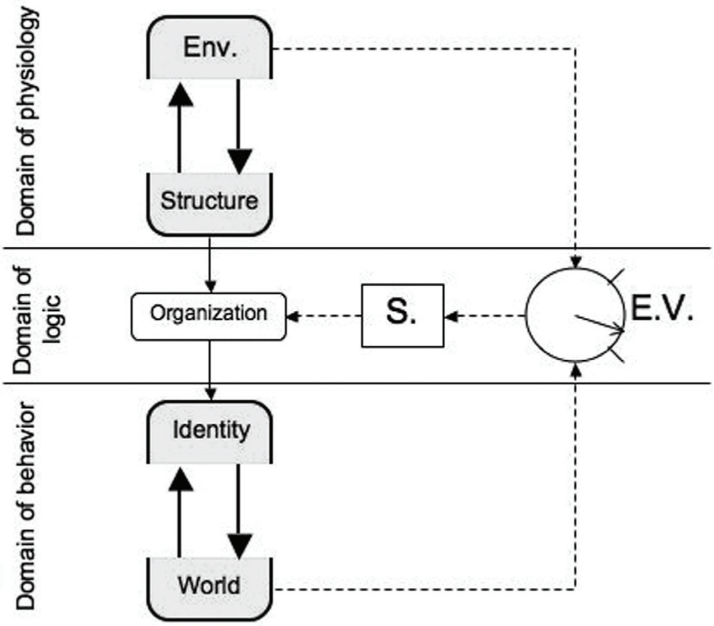
The realization that self-production is a defining aspect of the organization of the living might well be considered as the core insight of the autopoietic tradition, even if its full implications are obscured by an inherited discourse that prevents them properly grasping the phenomenon. The concept of autopoiesis, as self-production, made it conceivable for the first time that a scientific explanation of biological autonomy is indeed possible: a living system is autonomous because it operates so as to specify its own organization with its operations.

Of course, related ideas were also developed in cybernetics, for example in terms of the notion of *double closure*, whereby a system was allowed to operate on its operators (e.g., Von Foerster, 1973). But there remains a decisive difference: double

14. In the language of Von Foerster we could say that we have established the distinction between trivial and non-trivial systems.

closure entails autonomy by self-regulation and not self-production. In other words, the organization of operations of a doubly closed system can be parametrically regulated by its own operations, but precisely how this can be done is not specified by the operations of the system itself. Its organization remains externally specified. The autopoietic system, on the other hand, can specify its own organization because it is able to produce itself as a system. No equivalent notion of self-production exists in Ashby's work. Thus, by drawing their inspiration from the circular metabolic activity of cellular organisms, Maturana and Varela have managed to discover an essential biological phenomenon that goes beyond the standard cybernetic approach.

Figure 5: A Schematic of the Autopoietic Framework



There are three distinct domains about which we can make observations about organisms. In the *domain of physiology* we distinguish the organism as a physical structure (network of components) which exchanges matter and energy to its environment. This structure realizes a particular organization (network of relations) in the *domain of logic*, which we can distinguish as the autopoietic organization. This organization defines an identity (point of reference) for distinguishing sensory-motor interactions in the *domain of behavior*. Both structural and behavioral changes can lead to the disintegration of the autopoietic organization, at which point the organism ceases to exist.

The concept of autopoiesis has implications for all biological phenomena because the particular way in which an organism's autopoietic organization is realized will determine its identity as a living being and the domain of interactions in which it can engage its environment without disintegrating. We can therefore distinguish three distinct domains of discourse in the autopoietic tradition, namely (i) the physiological level of components, (ii) the logical level of relations, and (iii) the behavioral level of interactions. The advantage of these distinctions is that it is no longer possible to reduce behavior to structure and structure to physics. That would be to commit a category mistake. This framework is shown schematically in figure 5.

The concept of autopoiesis thus extends the cybernetic interest in self-maintenance with the biological phenomenon of self-production. However, what still needs to be carefully assessed is to what extent the definition of autopoiesis given by Maturana and Varela actually succeeds in accounting for that most elusive property of the living, biological autonomy. Is their conception of autopoiesis sufficient to explain "the self-asserting capacity of living systems to maintain their identity through the active compensation of deformations" (Maturana & Varela, 1973, p. 73)? More specifically, the crucial issue will be to determine in what sense the self-production of autopoietic systems, as described by Maturana and Varela, can be said to be self-asserting and active. Of course, we know that an account of self-asserting activity is precisely what they are trying to achieve:

> An organization may remain constant by being static, by maintaining its components constant, or by maintaining constant certain relations between components otherwise in continuous flow or change. Autopoietic machines are organizations of the latter kind: they maintain constant the relations that define them as autopoietic. (Maturana & Varela, 1973, p. 81)

In other words, the reason why Maturana and Varela consider autopoietic systems to be actively self-asserting is because these systems maintain their organization even though their components are in continuous flow and change. However, what is the reason for this ongoing turnover of components? We have encountered this idea of self-maintenance in the face of continuous perturbations before in Ashby's (1940) discussion of the vibrating cube as an illustrative example of a system in stable equilibrium. The problem with Ashby's account was that this continuous activity in response to the ongoing stream of perturbations is contingent on external factors. It was still possible in principle to place such a system in a perturbationless environment and thereby reduce it to passive stasis. In other words, the assertive activity was not intrinsic to the system. But is the process of self-production intrinsic to the existence of an autopoietic system?

We can imagine a robot that automatically repairs any damage that it suffers with its interactions with the world, and in this sense it would be an autopoietic system: the damage prompts it to regenerate its structure (cf. Froese & Ziemke, 2009). If the damage and repair is continuous it would appear to be a self-producing system. But what if we place it in an environment in which it suffers no damage? Then its generative activity would simply cease to take place. This thought experiment demonstrates that the concept of autopoiesis does not manage to account for intrinsically generated self-production, and it is therefore not a sufficient condition for the phenomenon of active self-assertion.

To be sure, Maturana and Varela are at pains to make it sound as if autopoiesis is both a necessary and sufficient condition, but they are constrained by their commitment to the abstract discourse of Ashby's general systems theory, which excludes considerations of component properties. This prevents them from resolving this issue by specifying that the components themselves must be unstable such that they would disintegrate if left isolated from the autopoietic system. Indeed, they indicate that this is what they mean to say:

> Since the relations of production of components are given only as processes, if the processes stop, the relations of production vanish; as a result, for a machine to be autopoietic, its defining relations of production must be continuously regenerated by the components which they produce. (Maturana & Varela, 1973, p. 79)

This qualification almost gets them off the hook. But the question remains: Under what conditions do the processes actually stop? In order to specify these conditions we have to appeal to component properties, namely issues of materiality and energy. Of course, this would normally not pose any in principle difficulties, except that Maturana and Varela are explicitly committed to a framework that separates such considerations from the level of organization in which autopoiesis is defined. In such an abstract domain of relations it is not clear how the physical conditions of a process could be incorporated (Froese et al., 2007). In sum, it turns out that the original concept of autopoiesis cannot be used to distinguish between self-production that involves (i) passive repair of an otherwise static system, and (ii) active creation of an intrinsically dynamic system. Maturana and Varela do not manage to accomplish what they set out to do, namely to capture the essence of biological autonomy which lies in an organism's active capacity for self-assertion.

In this failure lies the very root of the extensive list of problems that we have identified in the introduction to this article. Despite Maturana and Varela's convincing marketing of their position, it has become clear that their blind adherence to the principles of general systems theory has made their conception of autopoiesis unsuitable for the foundation of an organism-centered biology, bio-semiotics and, more recently, of Varela's enactive approach to cognitive science. Moreover, now we can also understand why autopoiesis has so readily been appropriated as a universal systems concept (e.g., Luhmann, 2002), and why Varela (1979) himself was able to generalize the concept into the abstract notion of organizational closure.[15] In fact, since the activity of self-production is not an intrinsic necessity of Maturana and Varela's autopoietic system, but only a contingent response to external environmental factors, we can simply abstract away from the requirements of the structural level altogether. It should therefore come as no surprise that it is not even clear what is lost

15. Varela put forward the notion of organizational closure so as to take "the lessons offered by the autonomy of living systems and convert them into an operational characterization of *autonomy in general*, living or otherwise" (Varela, 1979, p. 55). More precisely defined: "We shall say that autonomous systems are organizationally closed. That is, their organization is characterized by processes such that (1) the processes are related as a network, so that they recursively depend on each other in the generation and realization of the processes themselves, and (2) they constitute the system as a unity recognizable in the space (domain) in which the processes exist" (Varela, p. 55).

by abstracting away from autopoiesis itself to a purely behavioral domain constrained by an externally defined viability constraint, as has been done to great effect in dynamical and evolutionary approaches to robotics (e.g., Beer, 1997).

6. Beyond Maturana and Varela's Autopoiesis

We have localized the failure of the original conception of autopoiesis to properly do justice to the phenomenon of life in the concept's origins within Ashby's mechanistic framework of general systems theory. After their initial proposal Maturana and Varela have each in their own way tried to address some of the tensions that arise from this mismatch. We will consider these modifications at least briefly so that we have a better understanding of the potential options for reformulating autopoiesis. It turns out that Maturana and Varela have actually ended up with interpretations of autopoiesis that are mutually incompatible with each other. Maturana (2002) has become more interested in material considerations but still sticks with an Ashbyan approach that explicitly rejects Kant's notion of reciprocal causality, while Varela has tried to address some of the earlier shortcomings by taking on board phenomenology, semiotics, and Kant's notion of natural purpose (e.g., Weber & Varela, 2002).

6.1 Maturana on the Physical Necessities of Ashbyan Autopoiesis

In Maturana's later work we find an attempt to relate the concept of autopoiesis with precisely the two factors that Ashby (and previously Maturana and Varela as well) had abstracted away to arrive at his general systems formulation of cybernetics, namely material properties and thermodynamics. We now find him insisting on the importance of including material considerations when talking about the possibility of autopoiesis:

> an autopoietic system (a living system) exists only in the molecular medium in which it can operate as a totality in the conservation of its autopoietic dynamics through the continuous change of its molecular architecture through the *spontaneous thermal molecular dynamics*. (Maturana, 2002, p. 8; emphasis added)

This description of autopoiesis as a form of molecular self-production does indeed enable Maturana to exclude certain ambiguous organized structures like tornadoes and social systems from the class of autopoietic (and therefore of living) systems. Still, it remains doubtful that the simple addition of material specifications to a general systems concept, which was originally derived in abstraction from materiality as such, can do the job. It still remains a largely arbitrary restriction of the concept's applicability. In fact, the absolute distinctions between organization (identity) and structure (components), as well as between relational (interactional) and constitutive (physiological) phenomena (a.k.a. the doctrine of non-intersecting domains), still leave it absolutely mysterious as to why the material being of the autopoietic system should be operationally relevant for what it does as a systemic unity and vice versa.

If the concept of autopoiesis is to account for the active self-assertion characteristic of biological autonomy, then we must find a way for the process of self-production to be an internal necessity of the livings system rather than the result of external contingency. In addition, if we remain committed to the idea of developing a systemic account of biological autonomy, namely one that is focused on the relational level of organization, it is meaningless to introduce additional material specifications on the component level if these specifications are not necessarily entailed by the organization itself. Otherwise that organization could still in principle be realized in some other manner that does not satisfy those component specifications, but which retains the same organization.

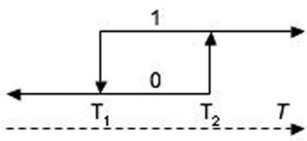
We need a phenomenon for which the components are not just realizing their systemic organization, for this is true of any physical system, but for which they only exist in this way because of this organization itself. In other words, we require a situation in which the components and the organized system as a whole mutually depend on each other for their existence. Indeed, despite Maturana's protestations to the contrary, this would be a rigorous way of cashing out more precisely what could be meant by his statement that autopoiesis depends on spontaneous thermal molecular dynamics.

One way of addressing this issue would be to make the activity of the component processes dependent on their organization, such that an organization of a self-producing system directly necessitates self-production on the level of components. In 5 we would accordingly have to add another arrow going back down from the system's organization (identity) to the underlying structural realization. However, this move is not supported by Maturana who, perhaps still under the influence of Ashby's (1962) "proof" of the apparent impossibility of self-organization, still denies the possibility of any such organizational entailment by explicitly rejecting a Kantian interpretation of autopoiesis (Maturana, 2002, p. 9). It seems that Maturana has made a move in the right direction by conceding the importance of material and energetic requirements for autopoiesis, but this realization has not been accompanied by any conceptual advances.

6.2 A Kantian Interpretation: Dissipative Structures and Beyond

Let us consider an illustrative case that could lend support to the possibility of a Kantian interpretation of autopoiesis. A paradigmatic example of the kind of system we are looking for is a dissipative structure such as a candle. We know that the candle's activity is not merely externally contingent because when we decrease the temperature to a point below the threshold at which the flame had originally emerged (the point of spontaneous emergence) the phenomenon still continues to persist in its organized form until some lower threshold is finally reached. Under these conditions, for instance when a source of fire is removed, the component processes continue to exist in a configuration of mutual regeneration that would be impossible if it were not for the existence of the structure as a whole. This state of affairs is represented by a hysterisis diagram, as shown in figure 6.

Figure 6: Diagram Representing the Property of Hysterisis That Is Characteristic of Dissipative Structures



A candle flame, for instance, will only emerge when the temperature of the system exceeds T_2 and remains present as T increases. But once this emergent structure is present (state 1) then it will continue to persist even if the temperature is decreased below T_2. The ongoing regeneration of the flame is thus internally contingent on the structure as a whole. Only when the temperature of the environment is further decreased below T_2 does the phenomenon disintegrate (no state, or 0).

Dissipative structures provide us with the existence proof that the organization of the system as a whole can have entailments on the level of components, and that forms of self-production that are intrinsically active are therefore possible. Moreover, since it is the case that living beings are dissipative structures on some level of description, and since we do not observe the spontaneous formation of living beings in our environment, we can be certain that they are realizing their existence under precarious conditions. This would be akin to the presence of a flame in a situation where $T_1 < T < T_2$. The study of far-from-equilibrium systems such as dissipative structures could open up an interesting area of research at the transition point between physics and biology (cf. Bickhard, 2009; Collier, 2008; McGregor & Virgo, in press).

Of course, this is not to say that the phenomenon of life is exhausted by a consideration of dissipative structures (Ruiz-Mirazo et al., 2004). The next crucial challenge will be to show how such structures can give rise to adaptive behavior (e.g., Barandiaran & Moreno, 2008). This conceptual situation is illustrated in table 1.

Table 1: Summary of Conclusions

Activity:	Self-maintenance	Re-generation	Adaptive behavior
External	Ultrastability	M&V's autopoiesis	
Internal		Dissipative structures	Living systems

Ashby's concept of ultrastability characterizes a machine that can *self-maintain* when prompted by external perturbations. Maturana and Varela extend this idea with the concept of autopoiesis which denotes a machine that can *self-repair* when prompted by external factors. In the case of dissipative structures we find an intrinsic need for *self-production* underlying the appearance of self-maintenance. Finally, in living systems we observe the additional capacity of *adaptive behavior*. Ashby's framework remains on the *external* row, while a Kantian interpretation shifts us to the *internal* row.

A more detailed discussion of the possibilities of bridging the gap between dissipative structures and living systems is beyond the scope of this paper, but a few brief pointers to relevant areas of research are in order. We suggest that the best strategy would be to work on both sides of the gap. On the one hand, it has been demonstrated by simulation that dissipative structures, in this case a reaction diffusion spot, can be made to follow external chemical gradients (Virgo & Harvey, 2008). However, the diffusion spots are normally immobile and so this capacity for spontaneous motility is a passive response to an environmental stimulus. Further research must be conducted to determine the conditions in which internally driven behavior starts to appear. Experiments with the self-motion of oil droplets might provide further insights in this regard (e.g., Hanczyc, Toyota, Ikegami, Packard, & Sugawara, 2007). And it is also important to determine whether we can make motion dependent on sensitivity to internal conditions (Egbert, Di Paolo, & Barandiaran, in press).

On the other hand, it is also essential that we clarify what the target of these concrete investigations actually is, namely the flexible and adaptive behavior displayed by living beings. Of particular importance will to be to clarify the conceptual relationship between adaptivity and autopoiesis (Di Paolo, 2005) and the manner in which adaptive behavior is realized even by the most simple organisms, in particular its internally differentiated infrastructure (Barandiaran & Moreno, 2008). Finally, it should be noted that another major hurdle remains: How do we account for the development of abstract cognition from adaptive behavior? In this regard it will be essential to consider the mediating role of the nervous system (Barandiaran & Moreno, 2006), the constitutive role of social interaction for behavioral capacities (Froese & Di Paolo, 2009), including the supportive structures of the socio-cultural environment (Steiner & Stewart, 2009).

6.3 Varela's Turn to Complexity Theory and Phenomenology

It is important to emphasize that this rejection of the cybernetic machine metaphor of life should not simply be equated with a dismissal of the systemic approach to biology. On the contrary, there have been many developments in dynamic systems theory that try to do justice to phenomena where the emergent properties of the whole are interdependent with the properties of the underlying components. This branch of mathematics is often referred to as complexity theory. And indeed this is precisely the approach that Varela chose to adopt for his later biological work. In a paper written two years before his death he implicitly admits that his earlier approach with Maturana was misguided:

> These *complex systems linking global to local are quite unrelated to the classical mechanical views* of causality and explanation. Unlike mechanical systems, most of what occurs is determined by the internal regulation of these systems, which keeps them within a domain of viability instead of at a preset point attractor. (Varela, 2001, p. 226; emphasis added)

In many ways complex systems theory thus promises to provide better scientific concepts for capturing the phenomena that cybernetics was clearly interested in, namely flexible adaptation, self-organization and active self-assertion, but which could not be adequately grasped other than by emphasizing that there was contingent activity due to continuous fluctuations and environmental perturbations. There is an important lesson to be learned here, because in the case of Ashby and Maturana this conceptual inadequacy was even elevated to a reductive principle against nature, whereby their limits of conceivability coincided with the supposed limits of reality. Contrary to their claims, there are material phenomena which exhibit global-to-local determination and it is possible to address their reciprocal relationship in rigorous manner through the mathematics of complexity theory. This theoretical development has opened the door for a systematic re-evaluation of the possibility of self-organization (Collier, 2008). We should therefore always beware of claims of impossibility that are derived from a formal level.

Accordingly, one way to interpret Varela's transformation of the concept of autopoiesis in his later work is that he eventually complemented Maturana's epistemological shift with a corresponding ontological shift: "what is unique about modern dynamical tools within science is a radical reconfiguration of various domains of ontology" (Varela, 2001, p. 225). It should also be mentioned in this context that his earlier radical constructivist epistemology has been refined as a consequence of his turn toward the phenomenological tradition in philosophy (e.g., Varela et al., 1991), though an analysis of the differences between these two philosophical viewpoints is beyond the scope of this paper.[16]

We are now in a position to summarize the historical development that took place from Ashby's cybernetics to the biological foundations of the enactive paradigm (cf. Froese, 2010) as shown in table 2.

Table 2: Summary of Historical Development

	Ontology:	**Epistemology:**	**Key concept:**
1. Ashby	Newtonian mechanics	Representationalism	Ultrastability
2. Maturana	Newtonian mechanics	Constructivism	Autopoiesis
3. Varela	Complexity theory	Phenomenology	Autonomy

Summary of the major historical developments in ontology and epistemology from the end of the cybernetics era, throughout the autopoietic tradition, and to the beginning of the enactive paradigm in the cognitive sciences.

The conceptual shift from passive stasis to dynamic activity has been made conceivable by replacing the traditional view of the living as mere mechanical machines with a more nuanced understanding of the far-from-equilibrium nature of biological autonomy (e.g., Ruiz-Mirazo & Moreno, 2004). Still, it remains to be seen whether the turn toward a phenomenologically informed approach to living and lived embodiment does not reveal additional characteristics of the phenomenon of life that we have not yet properly addressed (e.g., Sheets-Johnstone, 2000; Barbaras, 2002).

16. For an accessible introduction to phenomenology in the context of recent developments in the cognitive sciences, see Gallagher and Zahavi (2008).

Interestingly, even Ashby himself was aware of the important challenge that subjectivity, in its phenomenal expression as consciousness or lived experience, presented to the sciences of mind and behavior. First of all, he acknowledged that awareness of the first-person perspective is an irreducible fact, since

> the fact of the existence of consciousness is prior to all other facts. … there is no evidence in existence that could persuade me that my awareness itself was mistaken—that I had not really been aware at all. This knowledge of personal awareness, therefore, is prior to all other forms of knowledge. (Ashby, 1954, p. 11)

Nevertheless, Ashby chose to exclude the evidence of our lived experience from his work in cybernetics, at least for the time being, because he could not conceive how this insight could make a difference to what he was trying to achieve.

> As Ashby was later to conclude, cybernetics merely sheds some light on the *objective* aspects of behavior, namely those considered "from the point of view of an observer *outside* the system," without telling us anything bout the *subjective* aspects of the system itself. (Cordeschi, 2008, p. 224)

What Ashby was missing at the time, and what has only recently emerged alongside the enactive approach to cognitive science (e.g., Lutz, 2002), was an understanding of how to interweave the insights of first- and third-person perspectives in a mutually enlightening way. Our lived perspective needs to be brought to bear on the requirements for a scientific theory of life, and this theory, in turn, can inform the development of an embodied phenomenology. In fact, the synthetic method may play an essential role in this interweaving of methods and perspectives (Froese & Gallagher, 2010).

In any case, we have to give credit to Ashby for acknowledging the irreducible existence of the first-person perspective at a time when the topic of consciousness was still taboo in scientific circles. His existential insight provided the backdrop for the development of a rigorous science of the living that has thankfully managed not to fall into the temptation of trying to reduce subjectivity into the third-person view. This is in marked contrast to the computationalist tradition in the cognitive sciences which still continues to do precisely that with its insistence on the notion of an inner representation (Froese, 2010). By demonstrating the possibility of a science of mind and behavior devoid of a homuncular subject, Ashby has freed our perspective to look for subjectivity where it naturally belongs, namely in the first-person perspective.

7. Conclusion

We have argued that the original conception of autopoiesis is best understood from the perspective of Ashby's cybernetic framework. Our scholarly investigation has revealed that this original conception is fundamentally incapable of accounting for the active self-assertion that is characteristic of biological autonomy. On the basis of this diagnosis we have proposed to go beyond Ashby. The requirement of intrinsic self-production can only be met by far-from-equilibrium systems such as dissipative structures. This shift does not entail a rejection of the dynamic systems framework as such, because these far-from-equilibrium phenomena can still be researched in a systemic manner by means of recent developments in complexity theory. This suggestion is in line with the recent Kantian interpretation of autopoiesis favored by the enactive paradigm.

Since the Ashbyan and Kantian approaches to autopoiesis are incompatible with each other, we recommend that future work on autopoiesis should clearly specify which of these two interpretations is meant. Moreover, we suggest that the Ashbyan interpretation of autopoiesis should be avoided altogether. It fails to capture the phenomenon for which the concept of autopoiesis was originally intended, the phenomenon of life. Researchers who want to make use of the notion of autopoiesis in the Ashbyan manner, namely as a universally applicable abstract systems concept, can instead have recourse to Ashby's concepts of closure and ultrastability. The added incentive for this conceptual clarity is that Ashby's concepts are well founded in the mathematics of dynamic systems.

Of course, the scholarly analysis offered in this paper only clarifies the beginnings of an organism-centered biology of autonomy and sense-making. The Kantian conception of autopoiesis in terms of intrinsically active regeneration is only a necessary but not a sufficient condition for the phenomenon of life. An outstanding challenge is therefore to gain a better understanding of the transition from dissipative structures to living beings.

Acknowledgements

This article has benefited immensely from discussions held in the context of the Life and Mind seminars at the University of Sussex, especially Nathaniel Virgo. We would also like to thank Roger Harnden and two anonymous reviewers for their comments. Large parts of this article where written during Froese's research visit to Paris in 2009, which was financially supported by the EuCogII network.

References

Arnellos, A., Spyrou, T. & Darzentas, J. (2010). Towards the naturalization of agency based on an interactivist account of autonomy. *New Ideas in Psychology*, *28*, 296-311.

Ashby, J. (2009). W. Ross Ashby: A biographical essay. *Intentional Journal of General Systems*, *38* (2), 103-110.

Ashby, W. R. (1940). Adaptiveness and Equilibrium. *Journal of Mental Science*, *86*, pp. 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. (1951). Statistical machinery. *Thales*, *7*, 1-8.

Ashby, W. R. (1954). *Design for a brain: The origin of adaptive behavior.* New York: John Wiley & Sons Inc.

Ashby, W. R. (1956). *An Introduction to cybernetics*. London: Chapman & Hall.

Ashby, W. R. (1962). Principles of the self-organizing system. In H. von Foerster & G. W. Zopf (Eds.), *Principles of self-organization: Transactions of the University of Illinois symposium* (pp. 255-278). London: Pergamon Press. (Recently reprinted in: *E: CO*, *6* [1-2], 102-126, 2004)

Barandiaran, X., Di Paolo, E. A. & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, *17* (5), 367-386.

Barandiaran, X. & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, *14* (2), 171-185.

Barandiaran, X. & Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive Behavior*, *16* (5), 325-344.

Barbaras, R. (2002). Francisco Varela: A new idea of perception and life. *Phenomenology and the Cognitive Sciences*, *1*, 127-132.

Beer, R. D. (1997). The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems*, *20* (2-4), 257-289.

Bickhard, M. H. (2008). Emergence: Process organization, not particle configuration. *Cybernetics & Human Knowing*, *15* (3-4), 57-63.

Bickhard, M. H. (2009). The biological foundations of cognitive science. *New Ideas in Psychology*, *27*, 75-84.

Bourgine, P., & Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life*, *10* (3), 327-345.

Brier, S. (1995). Cyber-semiotics: On autopoiesis, code-duality and sign games in bio-semiotics. *Cybernetics and Human Knowing*, *3* (1), 3-14.

Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press

Cariani, P. A. (2009). The homeostat as embodiment of adaptive control. *International Journal of General Systems*, *38* (2), 139-154.

Christensen, W. D. & Hooker, C. A. (2000). Autonomy and the emergence of intelligence: Organised interactive construction. *Communication and Cognition – Artificial Intelligence*, *17* (3-4), 133-157.

Collier, J. (2004). Self-Organisation, individuation and identity. *Revue Internationale de Philosophie*, *59*, 151-172.

Collier, J. (2008). A dynamical account of emergence. *Cybernetics & Human Knowing*, *15* (3-4), 75-86.

Cordeschi, R. (2008). Steps toward the synthetic method: Symbolic information processing and self-organizing systems in early artificial intelligence modeling. In P. Husbands, O. Holland, & M. Wheeler (Eds.), *The Mechanical Mind in History* (pp. 219-259). Cambridge, MA: The MIT Press.

Di Paolo, E.A. (2003). Organismically-inspired robotics: homeostatic adaptation and teleology beyond the closed sensorimotor loop. In K. Murase & T. Asakura (Eds.), *Dynamical Systems Approach to Embodiment and Sociality* (pp. 19-42). Adelaide, Australia: Advanced Knowledge International.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4* (4), 429-452.

Di Paolo, E. A. (2009). Extended life. *Topoi*, *28* (1), 9-21.

Di Paolo, E. A., Rohde, M. & De Jaegher, H. (in press). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne & E. A. Di Paolo (Eds.), *Enaction: Towards a new paradigm for cognitive science*, Cambridge, MA: The MIT Press.

Dreyfus, H. L. (1991). *Being-in-the-world: A commentary on Heidegger's* Being and Time*, Division 1*. Cambridge, MA: The MIT Press.

Egbert, M., Di Paolo, E. A. & Barandiaran, X. (in press). Chemo-ethology of an adaptive protocell: Sensorless sensitivity to implicit viability conditions. In *Proc. of the 10th Euro. Conf. on Artificial Life*. Berlin: Springer Verlag.

Emmeche, C. (1998). Defining life as a semiotic phenomenon. *Cybernetics and Human Knowing*, *5* (1), 3-17.

Fleischaker, G. R. (1988). Autopoiesis: the status of its system logic. *BioSystems*, *22* (1), 37-49.

Froese, T. (2010). From cybernetics to second-order cybernetics: A comparative analysis of their central ideas. *Constructivist Foundations*, *5* (2), 75-85.

Froese, T. & Di Paolo, E. A. (2009). Sociality and the life-mind continuity thesis. *Phenomenology and the Cognitive Sciences*, *8* (4), 439-463.

Froese, T. & Gallagher, S. (2010). Phenomenology and artificial life: Toward a technological supplementation of phenomenological methodology. *Husserl Studies, 26* (1), 83-106.

Froese, T., Virgo, N. & Izquierdo, E. (2007). Autonomy: A review and a reappraisal. In F. Almeida e Costa, L. M. Rocha, E. Costa, I. Harvey, & A. Coutinho (Eds.), *Advances in artificial life: Proc. of the 9th Euro. Conf. on Artificial Life* (pp. 455-464). Berlin: Springer-Verlag.

Froese, T. & Ziemke, T. (2009). Enactive Artificial Intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, *173* (3-4), 366-500.

Gallagher, S. & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. London: Routledge.

Hanczyc, M., Toyota, T., Ikegami, T., Packard, N., & Sugawara, T. (2007). Chemistry at the oil-water interface: Self-propelled oil droplets. *J. Am. Chem. Soc.*, *129* (30), 9386-9391.

Hanna, R. & Thompson, E. (2003). The mind-body-body problem. *Theoria et Historia Scientarum*, *7* (1), 24-44.

Hoffmeyer, J. (1998). Surfaces inside surfaces: On the origin of agency and life. *Cybernetics & Human Knowing*, *5* (1), 33-42.

Hoffmeyer, J. (2000), "The Biology of Signification", *Perspectives in Biology and Medicine*, *43* (2), pp. 252-268

Husbands, P., Holland, O. & Wheeler, M. (Eds.) (2008). *The mechanical mind in history.* Cambridge, MA: The MIT Press.

Jonas, H. (2001). *The phenomenon of life: Toward a philosophical biology.* Evanston, IL: Northwestern University Press. (Originally published in 1966 by Harper & Row of New York.)

Jonas, H. (1992). The burden and blessing of mortality. *The Hastings Center Report*, *22* (1), 34-40.

Kant, I. (1987). *Critique of Judgment* (W. S. Pluhar, Trans.). Indianapolis, IN: Hacket Publishing Company, 1987 (Originally published in 1790 as *Kritik der Urteilskraft*)

Klir, G. J. (2009). W. Ross Ashby: A pioneer of systems science. *International Journal of General Systems*, *38* (2), 175-188.
Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 43, 1940-1951.
Luhmann, N. (1984). *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Frankfurt, Germany: Suhrkamp Verlag. (Also available in English: Luhmann, N. (1995), *Social Systems*, Stanford, CA: Stanford University Press.)
Luhmann, N. (2002). *Einführung in die Systemtheorie*. Heidelberg, Germany: Carl-Auer Verlag.
Lutz, A. (2002). Toward a neurophenomenology of generative passages: A first empirical case study. *Phenomenology and the Cognitive Sciences*, *1* (2), 133-167.
Maturana, H. R. (1970). Biology of Cognition. In H. R. Maturana & F. J. Varela (1980), *Autopoiesis and cognition: The realization of the living* (pp. 1-58). Dordrecht, Holland: Kluwer Academic.
Maturana, H. R. (1970b). Neurophysiology of Cognition. In P. L. Garvin (Ed.), *Cognition: A multiple view* (pp. 3-23). New York: Spartan Books.
Maturana, H. R. (1988). Reality: The search for objectivity or the quest for a compelling argument. *The Irish Journal of Psychology*, *1* (9), 25-82.
Maturana, H. R. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, *9* (3-4), 5-34.
Maturana, H. R. & Varela, F. J. (1973). Autopoiesis: The organization of the living. In H. R. Maturana & F. J. Varela (1980), *Autopoiesis and Cognition: The Realization of the Living* (pp. 59-140). Dordrecht, Holland: Kluwer Academic.
McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115-133.
McGregor, S. & Virgo, N. (in press). Life and its close relatives. In *Proc. of the 10th Euro. Conf. on Artificial Life*. Berlin: Springer Verlag.
Moreno, A. & Etxeberria, A. (2005). Agency in natural and artificial systems. *Artificial Life*, *11* (1-2), 161-175.
Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83* (4), 435-450.
Pickering, A. (2009). Psychiatry, synthetic brains and cybernetics in the work of W. Ross Ashby. *International Journal of General Systems*, *38* (2), 213-230.
Pickering, A. (2010). *The cybernetic brain: Sketches of another future*. Chicago: The University of Chicago Press.
Rosenblueth, A. N., Wiener, N. & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science*, *10* (1), 18-24.
Ruiz-Mirazo, K. & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life*, *10*(3), 235-259.
Ruiz-Mirazo, K., Peretó, J. & Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere*, *34*, 323-346.
Sheets-Johnstone, M. (2000). The formal nature of emergent biological organization and its implications for understandings of closure. *Annals of the New York Academy of Sciences*, *901*, 320-331.
Steiner, P. & Stewart, J. (2009). From autonomy to heteronomy (and back): The enaction of social life. *Phenomenology and the Cognitive Sciences*, *8* (4), 527-550.
Stewart, J. (1992). Life = cognition: The epistemological and ontological significance of artificial life. In F. J. Varela & P. Bourgine (Eds.), *Towards a Practice of Autonomous Systems: Proc. of the 1st Euro. Conf. on Artificial Life* (pp. 475-483). Cambridge, MA: The MIT Press.
Stewart, J. (1996). Cognition = life: Implications for higher-level cognition. *Behavioral Processes*, *35*, 311-326.
Stewart, J., Gapenne, O., & Di Paolo, E. A. (Eds.) (in press). *Enaction: Towards a new paradigm for cognitive science*, Cambridge, MA: The MIT Press.
Thompson, E. (2004). Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences*, *3* (4), 381-398.
Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: The Belknap Press of Harvard University Press.
Umpleby, S. A. (2009). Ross Ashby's general theory of adaptive systems. *International Journal of General Systems*, *38* (2), 231-238.
Varela, F. J. (1979). *Principles of Biological Autonomy.* New York: Elsevier North Holland.
Varela, F. J. (1996). The early days of autopoiesis: Heinz and Chile. *Systems Research*, *13* (3), 407-416.
Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34* (1), 72-87.
Varela, F. J. (2001). Why a proper science of mind implies the transcendence of nature. In J. Andresen (Ed.), *Religion in mind: Cognitive perspectives on religious belief, ritual, and experience* (pp. 207-236). Cambridge, UK: Cambridge University Press.
Varela, F. J., Maturana, H. R. & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, *5* (4), 187-196.
Varela, F. J. & Shear, J. (1999). First-person methodologies: What, why, how? *Journal of Consciousness Studies*, *6* (2-3), 1-14.
Varela, F. J, Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: The MIT Press.
Virgo, N., Egbert, M., & Froese, T. (in press). The role of the spatial boundary in autopoiesis. In *Advances in Artificial Life: Proc. of the 10th Euro. Conf. on Artificial Life*. Berlin: Springer Verlag.

Virgo, N. & Harvey, I. (2008). Reaction-diffusion spots as a model for autopoiesis (abstract). In: S. Bullock, J. Noble, R. Watson, & M. A. Bedau (Eds.), *Artificial Life XI: Proc. of the 11th Int. Conf. on the Simulation and Synthesis of Living Systems* (p.816). Cambridge, MA: The MIT Press.

Von Foerster, H. (1970). Molecular ethology: An immodest proposal for semantic clarification. In G. Ungar (Ed.), *Molecular Mechanisms in Memory and Learning* (pp. 213-148). New York: Plenum Press.

Von Foerster, H. (1973). On constructing a reality. In W. F. E. Preiser (Ed.), *Environmental Design Research* (vol. 2, pp. 35-46). Stroudsburg, PA: Dowden, Hutchinson and Ross.

Von Foerster, H. (1976). Objects: Tokens for (eigen-)behaviors. *ASC Cybernetics Forum*, *8* (3-4), 91-96.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds (C. H. Schiller, Trans.). In: C. H. Schiller (Ed.), *Instinctive Behavior: The Development of a Modern Concept* (pp.5-80). New York: International Universities Press. (Original work published as *Streifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten* in 1934)

Weber, A. (2001). The 'surplus of meaning'. Biosemiotic aspects in Francisco J. Varela's philosophy of cognition. *Cybernetics & Human Knowing*, *9* (2), 11-29.

Weber, A. & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, *1* (2), 97-125.