Homeostats for the 21st Century? Simulating Ashby Simulating the Brain

Stefano Franchi • Texas A&M University, USA • stefano.franchi/at/gmail.com

> Context • W. R. Ashby's work on homeostasis as the basic mechanism underlying all kinds of physiological as well as cognitive functions has aroused renewed interest in cognitive science and related disciplines. Researchers have successfully incorporated some of Ashby's technical results, such as ultrastability, into modern frameworks (e.g., CTRNN networks). **> Problem** • The recovery of Ashby's technical contributions has left in the background Ashby's far more controversial non-technical views, according to which homeostatic adaptation to the environment governs all aspects of all forms of life. This thesis entails that life is fundamentally "heteronomous" and it is conceptually at odds with the autopoiesis framework adopted by Ashby's recent defenders as well as with the primacy of autonomy in human life that most of the Western philosophical tradition upholds. The paper argues that the use of computer simulations focused on the more conceptual aspects of Ashby's thought may help us recover, extend and consequently assess an overall view of life as heteronomy. **> Method** • The paper discusses some computer simulations of Ashby's original electro-mechanical device (the homeostat) that implement his techniques (double-feedback loops and random parameter-switching). **> Results** • First simulation results show that even though Ashby's claims about homeostatic adaptivity need to be slightly weakened, his overall results are confirmed, thereby suggesting that an extension to virtual robots engaged in minimal cognitive tasks may be successful. **> Implications** • The paper shows that a fuller incorporation of Ashby's original results into recent cognitive science research may trigger a philosophical and technical reevaluation of the traditional distinction between heteronomous and autonomous behavior. **> Constructivist content** • The research outlined in the paper supports an extended constructionist perspective in which agency as autonomy plays a more limited role. **> Key words** • **Second-order cybernetics, autonomy and agency, general homeostasis thesis, W. R. Ashby, homeostat, concept simulation.**

« 1 » This paper describes some early experiences with a virtual yet faithful replica of Ashby's original homeostat, as described in *Design for a Brain*, which was designed in order to explore Ashby's *generalized homeostasis thesis* in biologically and psychologically more salient settings. The first two sections discuss the theoretical and perhaps non-obvious rationale behind the project. The following section describes the implementation of the virtual homeostat and some early experiments in broad qualitative terms. The final section discusses some options in extending the homeostat to a virtual robotic environment that will be the main subject of future work in the project.

Simulating objects vs. simulating concepts

« 2 » The construction of computational devices embodying particular views of cognition may serve different theoretical and practical goals. On the one hand, researchers may use a computational device to validate the cognitive theory it embodies. For instance, the traditional AI approach pioneered in the 1960s by Herbert Simon and Allen Newell identified higher cognitive functions with searching an abstract symbolic space organized by goals and sub-goals. The early computational devices Simon and Newell built (e.g., the General Problem Solver, as well as their chess player) tried to validate the theory by simulating specific cognitive tasks carried out according to the theory and then measuring the machine's output by comparing it to human performances on similar tasks. The proximity between human and mechanical outputs provided a metric for measuring a cognitive theory's validity. Computer simulations of cognitive operations were thus the equivalent, to a certain extent, of hard science experiments. This approach to the use of computational technology is still widely practiced, even though the identification of the mind with computer software running symbolic manipulations has long gone out of fashion. A large part of current work in Artificial Life and evolutionary robotics, for instance, follows this approach: computer simulations are interpreted as abstract replicas of real world entities and researchers are interested in minimizing the distance between the simulated and the actually observed behaviors. When the theory the computer simulation embodies can be characterized as being "constructionist," we thus have an occurrence of computational constructivism.[1] We could call this use of computational device the "simulation of real objects," where the last term is understood broadly enough to encompass living beings as well as real world processes.[2]

« 3 » An alternative approach to the use of computational devices focuses in-

---

[1] See Egbert, Barandiaran & Di Paolo (2010, 2012) on metabolution for a recent example of such an approach.

[2] For a full articulation of the theoretical basis behind object simulations in the sense discussed above, see Gomila & Müller (2012).

stead on simulating *concepts*. In this case, computational devices are used to produce more concrete implementations of abstract ideas that allow the researchers to gain further insights into their structure. Simulating software becomes "a tool to think with" that gives its users the ability to play with an abstract concept independently of its possible applications. Seymour Papert's theoretical and concrete work provides the best known example of how artifacts can be used to simulate conceptual articulations. Logo's *microworlds* (including, but not limited to, its famous "turtle geometry world") are environments simulating the behavior of abstract mathematical concepts, often at a very considerable level of theoretical complexity.[3] Concept simulations are intrinsically exploratory. As a consequence, the proximity between the simulated behaviors and their equivalents in the real world is not an important metric. Concept simulations are successful if they allow the researchers to get a better grasp of the concept at issue and, especially, if they allow them to understand its possible articulations. They are elaborate thought experiments.

« 4 » The two approaches I just sketched are not necessarily incompatible.[4] In fact, I would claim that any successful project in the cognitive sciences that has relied on extensive use of computational devices has been forced to use both approaches. I will refer again to the classic AI example mentioned above: Newell and Simon's simulations were also thought experiments. They started from the view that a cognitive agent is first and foremost a (symbolic) problem solver and tried to examine what it would be like for such an agent to deal with a variety of cognitive challenges. For instance, which problem solving strategies would be more effective in a closed-world well-defined scenario such as chess? Which strategies would bring more benefits to an open-world fuzzy scenario such as human language interaction? More recent work in Artificial Life and evolutionary robotics follows the same paradigm, mixing object and concept simulations in the same research projects.[5] Nonetheless, I think it is important to keep the two kinds of simulation separate at the analytic level. Not only do they have different concerns and different metrics, but the exclusive focus on the simulated objects' behavior that we find prevalent in critical discussions and in the researchers' presentations of their work risks obfuscating the conceptual work. To put it bluntly: sometimes, seemingly technical computationally-intensive work in the cognitive sciences is (also) philosophy in disguise. It is philosophy presented in a technical garb more suited to a positivistic-minded age that shies away from abstract conceptual frameworks unless they are presented under scientific-looking covers. I use here the term "philosophy" in its precise technical meaning, and not as a synonym for "abstract thinking" or, worse, as a shorthand for "critical reflection upon empirical results." Since its not so recent foundation, philosophy has been the pursuit of a general and necessarily abstract and self-reflective knowledge about "being as such."[6] In other words, it is the search for an overall perspective on life at all its levels. There is no doubt that such an overall perspective is present in the enaction paradigm, from its inception in Francisco Varela's work to its most recent reincarnations. An overall philosophical framework was also present in classic AI (Franchi 2006), as well as in several of the research paradigms that emerged after its prolonged crisis in the 1980s. Keeping concept simulation distinct from its object counterpart helps us isolate the philosophical component of these works and may guide us toward a better appreciation of a research program's scope, its challenges, and its unfinished business. This task is particularly important when dealing with the work of Ross W. Ashby.

« 5 » Ashby's major work, *Design for a Brain* (Ashby 1952b), contains a detailed description of an electro-mechanical device, the *homeostat*, which, he claims, can concretely demonstrate some essential features of the nervous system. Most readers understand the homeostat as a proxy object: it is an artificial, extremely simplified version of a complex, real world object, namely the brain. Performing experiments on the homeostat is thus equivalent to simulating an object's behavior in the sense discussed above. Consequently, the proximity between the homeostat's behavior and the brain's, its real world counterpart, would provide a measure of the correctness of Ashby's theory about cognitive and sub-cognitive functions. Ashby's numerous references to the then current psychological and neuro-physiological literature, as well as his efforts to show how the basic structure of some relevant sample of psychological behavior could be mimicked by his device, support the view of the homeostat as simulating objects. Yet, in my view, this interpretation, while correct, fails to capture the fuller extent of Ashby's thought. As I will argue in the next section, it is only by *also* considering the homeostat as a *proxy concept* that we can appreciate it.

The homeostat as a concept

« 6 » Ashby's theory can be reduced to a small number of theses that I will sum up as follows:

1 | A complex system running to equilibrium generates complex behavior.
2 | A system is in equilibrium when it keeps its essential variables at stable levels.
3 | A system *runs* to equilibrium through ultrastability; when its essential variables are unstable, it rearranges the connections between its subparts.
4 | A system *achieves* ultrastability through a stochastic step function that operates on the internal connections of the system by randomly stepping through the parameters connecting together its various parts.[7]

---

[3] See, for instance, the early mathematical explorations by Abelson & DiSessa (1982), and the more recent work based on Uri Wilensky's *NetLogo* (An & Wilensky 2009).

[4] For a more detailed analysis of various uses of computer simulations, see Asaro (2011).

[5] Among the many examples, I will just mention the 2008 special issue of BioSystems on "Modeling autonomy," which was meant to strengthen and move Varela's enactive cognition paradigm further on. As the editors' introduction makes clear (Barandiaran & Ruiz-Mirazo 2008), the simulation of autonomy is really a conceptual enterprise based on a very strong philosophical outlook on life.

[6] Aristotle's definition in *Metaphysics*, book Γ, substantially accepted by the Western philosophical until after Hegel.

[7] The first thesis is a compact formulation of chapter 4 ("Stability") of Ashby's *Design for a*

« 7 » The homeostat embodied these four theses in its very construction. On the basis of the experiments he carried out with his device, Ashby eventually derived the following further set of claims:

5 | A system with many parts (subsystems) that are fully connected to each other will need an exceedingly long time to reach equilibrium.

6 | A real system will then need weak connections among its parts – the system will be *dispersed*.

7 | Typically, the subsystems will not be mutually connected on the organism side: (its interactions with) the environment will mediate the communication among subsystems.[8]

« 8 » Let me focus on the first four core points. The first one – *the generalized homeostasis thesis* – is Ashby's most original and, at the same time, least warranted idea.[9] It embodies Ashby's fundamental claim about the behavior of any complex system – be it a physiological, psychological, or even a social one. Ashby's remaining theses and the homeostat itself make the generalized homeostasis thesis plausible by demonstrating the insights we can gain through their broad application and by virtue of the conclusions stemming from them.

« 9 » The second thesis – the equilibrium of an organism equals the stability of its essential variables – is Ashby's least original but most solid one. It amounts to a restatement of the principle of homeostasis that was to become widely accepted in the physiological community by the late 1940s. The third and fourth theses, instead – the principle of ultrastability as a double feedback that we achieve through random step functions – represent the core of Ashby's *technical* contribution. In other words, they are the basic components of the homeostat as an object (or, more precisely as a proxy object for the real object, the brain).

« 10 » Ashby wanted to prove point 1, which constitutes a very general and broad thesis about life, by effectively building a simplified system (the homeostat) that illustrates its plausibility. It is easy to see why the first thesis is so unorthodox, and why it encountered such resistance in the cybernetic community when the homeostat was first presented at the Macy foundation's meetings, where Ashby claimed that life is fundamentally a process (or series of processes) aiming at minimizing the impact of external disturbances on the organism.[10] The nickname devised by William Grey Walter seems quite apt: Ashby's homeostat is a *machina sopora* [sleeping machine], he claimed, because its ideal function is to go back to sleep as quickly as possible.[11] Walter's characterization is rather superficial, though, because it does not take into account Ashby's insistence on the complexity of behavior that can be produced by the systems he discusses. Walter's remark, however, does indicate where the main source of resistance to Ashby's thesis about life lies: in its effort to derive action from non-action. In other words, in Ashby's homeostatic organism, philosophical and theoretical preeminence is accorded to non-action – or to sleep, in Walter's mocking words. This does not mean that the homeostat will always be sleeping. Rather, it means something philosophically far more disturbing: the homeostat will act as much as is needed in order to *go back to sleep*. It will go to extraordinary lengths, it will take whatever action it can – just in order to go back to rest (i.e., to equilibrium). Moreover, its always possible failure to do so will necessarily result in the homeostat's death. In other words, the Ashbian organism will always be trying to accommodate itself to its environment by whatever means necessary: it is essentially a *passive* machine whose activity is a by-product of its search for non-action. It is also a *contingent* machine insofar as its "search" for equilibrium will involve essentially random processes. All life forms (life itself) turn out to be made out of contingent, dispersed, embedded, equilibrium-seeking beings that may (or may not), among the many devices they use to achieve quiescence, include the generation of goals and their related consciousness.[12]

Brain. A shorter formulation is presented at the end of the book (Ashby 1960: ch. 14/1: 184); the second thesis sums up Ashby's discussion of biological systems in chapter 5 of the same work (see especially ibid: ch. 5/3: 58; ch. 5/7: 64); the third thesis corresponds to Ashby's principle of double feedback (ibid: ch. 7/2–7/8: 80–84); finally, the fourth thesis about the biological necessity of stochastic step functions is defended in chapter 7 (ibid: ch. 7/13–7/25: 87–98) as well as in a number of other publications (Ashby 1947, 1952a, 1953, 1956b).

[8] For Ashby's theses 5–7 see, respectively: Ashby 1960: ch. 11 "The Fully-Joined System," ch. 16/11: 214; 1960: sec. 16/5–16/6 208–209. Ashby could provide empirical confirmation for the fifth thesis when he conducted a series of experiments to estimate the connectance threshold value that would separate a stable system from a hopelessly unstable one. His data indicated that even small systems become immediately unstable when their connectance goes beyond a relatively low value of 13% (Ashby 1970). The sixth thesis is the principle of Dispersion that Ashby discussed in the first edition of *Design for a Brain*, where he claimed that it was applicable to sense-organs and to the brain (Ashby 1952: 103ff)). Unfortunately, the later edition drops the effective label, "dispersed," while keeping the concept: "evidence has been given… that the actual richness [of connections within a system] is by no means high. [We] have shown that when it is not high, adaptation by ultrastability can occur in a time that is no longer impossibly long."

[9] The label is Ashby's. It provides the title of section 5–6 of the second edition of *Design of a Brain*. Ashby often used it to refer to his overall view of the brain as regulator.

[10] See, for instance, Julian Bigelow: "It may be a beautiful replica of something, but heaven only knows what" in Foerster, Mead & Teuber (1953: 95).

[11] See Walter (1961: 111). Walter concluded his brief review of Ashby's homeostat by saying that "judging *Machina sopora* entirely by its behavior, the naturalist would classify it as a plant" (ibid: 112).

[12] In Franchi (2011b), I presented an extended argument in defense of my "7-theses" interpretation of Ashby's project, advanced in this paragraph on the basis of a detailed examination of the history and evolution of the concept of homeostasis from Claude Bernard to the classical works by Ernest Henry Starling, John Scott Haldane and Walter Cannon, and the mid-20th century work by Norbert Wiener and Julian Bigelow, up to the recent rediscovery of neural homeostasis suggested by Gina Turrigiano and other researchers, as well as the alternative conceptions advanced by Peter Sterling and Jay Schulkin. Even though space constraints prevent me from repeating the entire argument here, there is one point I would like to stress: my distinction between

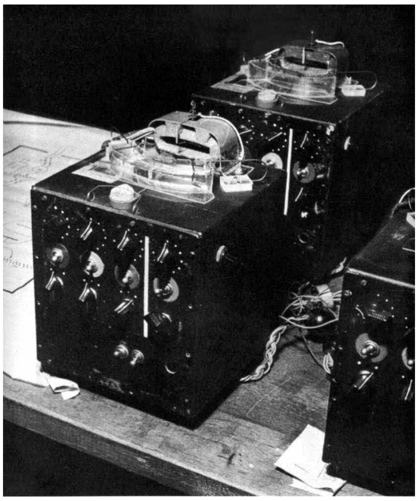


Figure 1: Two of the homeostat's four units (from *Design for a Brain,* Ashby 1960: 101).

« 11 » This necessarily brief discussion of Ashby's project allows us to reinterpret the homeostat as (also) a concept simulator. Ashby's device is a proxy for a view of life as generalized homeostasis, whose principles it embodies. If we keep in mind Grey Walter's characterization (which captures the content of theses 1 and 2) and the homeostat's reliance on ultrastability through stochastic step processes (which captures the content of theses 3 and 4), we could define the homeostat as *a simulator of passive contingent life*. The latter term is the *concept* that the homeostat articulates and Ashby explores throughout his book. The numerous claims Ashby defends in the second half of *Design of a Brain* – from the concepts of polystable and multistable systems to the introduction of the principles of dispersion and localization – are conceptual explorations of the theoretical consequences of adopting a view of life as passive and contingent. In other words, Ashby tries to address the question: "What would it be like to live a passive/contingent life"? Or, more precisely, "What would the necessary internal functional organization of a passive/contingent organism be like?" The various experiments he carries out on the homeostat provide evidence in support of the feasibility of the concept. Take, for instance, the last experiment in the book, presented in chapter 16 (Ashby 1960: 210–212). The experiment shows how a system composed of individually ultrastable subsystems can achieve overall stability if (and only if) its components are poorly joined. I find Ashby's choice of words when introducing the experiment extremely interesting: "the process," he says, "can be illustrated on the homeostat" (1960: 210). The homeostat "illustrates" a process: it paints a picture of what the process would be like *if* life were to follow the general homeostasis thesis. The homeostat is a complex thought experiment used to carry out a sophisticated conceptual exploration.

« 12 » This conclusion leads me to a second important issue. From the concept-simulation standpoint, the most important thesis Ashby defends is the first one: "A complex system running to equilibrium generates complex behavior." All the other claims he advances in *Design for a Brain* are either further explications of this thesis (as are the definitions of equilibrium in terms of essential variables and of ultrastability as random rearranging of internal parameters), or its direct theoretical consequences (e.g., dispersion and environment-side connection among subparts). Yet, the critical discussion of Ashby's work has tended to focus on the derived claims, and particularly on the ultrastability mechanism. This is not a recent phenomenon. Already Norbert Wiener, while praising Ashby's work, had demoted homeostasis to a particular instantiation of the principle of feedback at the physiological level.[13] We can observe a similar tendency in recent re-elaborations of the Ashbian paradigm. Often, it is ultrastability as a double-feedback system that receives most of the attention, especially when its basic mechanism is grafted upon a well-known neural network approach (Di Paolo 2000). By focusing their attention on homeostasis's mechanisms instead of the general homeostasis thesis that precedes them, old and new interpretations privilege the objective component of Ashby's simulation at the expense of its conceptual component. The reason may lie in the still-theoretically controversial status of Ashby's general homeostasis thesis. The thesis's insistence on life's essential heteronomy, to use Kantian language that has now become fashionable again, is very much at odds with a philosophical tradition that has traditionally distinguished human from non-human life precisely on the basis of the opposition of autonomous and heteronomous behavior.[14]

« 13 » The recent emphasis on autonomous behavior in cognitive science, the philosophy of mind, and evolutionary robotics has tried to reject the classical dichotomy by pushing the threshold identifying philosophical autonomy toward less and less organized forms of life. The best known example of this trend is most likely Hans Jonas's (2001) view of any form of metabolism, from bacteria "upward," as containing the essential kernel of philosophical freedom in Kant's sense: the possibility to establish the rules governing one's own behavior. Jo-

---

"technical" vs. "non-technical" theses (resp.: 3–4 vs. 1–2) does not imply that the latter are somewhat more important, original or creative than the former. It simply denotes their different logical status. Thesis 1, as I briefly mention here and discussed at length in the previously-mentioned paper, is a philosophical thesis – it properly belongs to metaphysics. Thesis 2 belongs to biology. Taken together, they introduce ways to think about the world in general and about biological life in particular. Theses 3 and 4 explain how those concepts could be realized in biological or physical systems. Which of the two groups of theses is more important depends on the assumed perspective. My claim is that, historically speaking, Ashby's readers have been mostly interested in how to build artifacts or in how to provide biological explanations for real biological systems. Hence, they (legitimately, given their interests) focused their attention on theses 3–4 and let theses 1–2 slide to the background. The shift in perspective from "simulating objects" to "simulating concepts" I am advocating here would allow us to focus primarily on theses 1–2.

13| Wiener (1948: 135). See also the later discussion of homeostasis in Wiener & Schadé (1963: 1), now in Wiener (1976: 400). Ashby's view could not be further from this, as he explicitly declares in *Introduction to Cybernetics*: "there has yet to be written the book […] that shall show how all the organism's exteriorly-directed activities – its "higher" activities – are all similarly regulatory, i.e., homeostatic" (Ashby 1956b: 196, emphasis added).

14| It will suffice to mention that the mutual isomorphisms between autonomy/heteronomy, human/non-human, and passions/reason have played a largely unchallenged foundational role in Western philosophy since Plato's times.

nas's view, in turn, has played an important role in the theory of life as autopoiesis (Varela (1979; Weber & Varela 2002) that often provides the initial conceptual framework adapted by those researchers who have successfully recovered the technical parts of Ashby's conception (Di Paolo 2005; Ikegami & Suzuki 2008).

« 14 » In my view, Ashby's conceptual effort goes in the exactly opposite direction. While Jonas pushed the boundary between autonomy and heteronomy downward to turn bacteria into proto-autonomous creatures, the generalized homeostasis theory pulls it up so much that even the traditionally highest cognitive functions become fully heteronomous (e.g., playing chess, as in Ashby 1952a and 1953).[15]

« 15 » I think one possible approach toward a better understanding of Ashby's broader claim is to refocus on the original device in order to further explore the view of passive contingency he advocated. In other words, refocusing on the original homeostat as an exercise in concept simulation is a strategy aimed at asking again the following question: what is it like to live a passive, fully homeostatic life? Putting back at the center of our theoretical attention the *conceptual* simulation of life as general homeostasis entails starting from (but certainly not finishing with) a modern version of Ashby's homeostat. This is not because we want to preserve somehow the original and therefore allegedly more authentic flavor of Ashby's work. Rather, it is because the effort to recover the whole original package would preempt the temptation to focus on the mechanical, object-oriented aspects of his simulations.

« 16 » Thus, a research program consistent with the previous discussion would start by replicating Ashby's original setup, and would then proceed to replicate some of the experiments detailed in *Design for a Brain*. Then, it would conduct some preliminary investigations of general characteristics of the homeostat, and finally, and most importantly, expand the scope of Ashy simulations to other, open-ended domains (i.e., with "real" environments) by embedding homeostats in sensor- and motor-endowed "virtual creatures" acting in a real "environment" – virtual robots or, to put it differently, virtual minimally cognitive agents, instead of the self-contained original homeostat.

« 17 » It is perhaps worth stressing that while Ashby claimed repeatedly that the structure of his device's abstract *performances* bore some similarities with living beings' actual behavior, he never claimed, nor could he reasonably do so, that the *device* itself had any biological plausibility. From the point of view of concept simulation I am advocating here, the homeostat's lack of biological reality is actually an advantage. It allows the researcher to focus on its abstract features while leaving in the background important, yet premature questions about the possible biological realizations of general homeostasis. For this methodological reason, and not for any archaeological interest, my suggested recreation of the homeostat will follow as closely as possible Ashby's original implementation, as I will describe briefly in the next section.

A simulated homeostat

The original device

« 18 » Ashby's original device (see Figures 1 and 2) is so well-known that my description will be very brief. The original homeostat is a fully-connected collection of 4 double-feedback units, each including a random step mechanism capable of altering the weight of its 4 connections when triggered. This mechanism (which Ashby calls a "uniselector," from the name of the actual physical device he used to implement it) is capable of randomly selecting one out of 25 possible equally spaced values for each input connection. The step mechanism aside, a homeostat unit j is a simple dynamic device whose value y_j in an N-unit network is described by the following equation (Ashby 1960: 246ff.):

$$m\ddot{y}_j = -v\dot{y}_j + \sum_{i=1}^{N} w_{ij} y_i$$

where v is the viscosity of the medium in which the unit's needle works, m depends on the unit's needle's moment of inertia, and w_{ij} is the weight of the connection from unit i

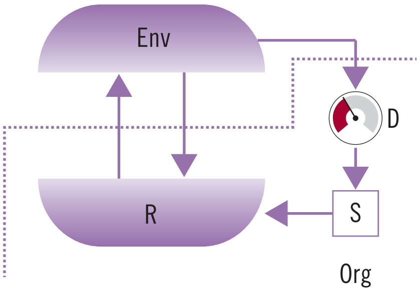


Figure 2: Diagram illustrating Ashby's principle of ultrastability as a double-feedback loop in a typical homeostat unit (Ashby 1960: 83). *Env* is the Environment (in the homeostat, the other three units); *R* is the part of the organism that interacts with it; the box *S* represents the set of parameters affecting the organism's behavior; the dial *D* stands for the "essential variable" monitor whose out-of-bound (unfavorable) value triggers the switch to a new set of randomly chosen parameters in *S*, ensuring "ultrastability." The thick arrows stand for the sensory and motor channels. The area within the dotted line is on the organism side, the area outside is on the environment side. The shaded areas in Env and R represent the "fine networks of linkages" existing within each of them. The first-order feedback loop occurs between *R* and *Env*, as the reactive part *R* constantly acts on the environment *Env* on the basis of the input it receives from it. The second-order feedback loop occurs between the *R-Env* and the *D-S* subsystems. *D* periodically measures the value of the essential variable it is meant to protect (e.g., sugar or CO_2 levels, etc.) and changes *S* (i.e., *R*'s parameters) accordingly. Note that Ashby identifies *R* in purely functional terms and it does not need to correspond to any anatomical or histological structure. See Figure 5 for an illustration of Ashby's ultrastability mechanism in a simple conceptual robot.

---

[15] I discuss why Ashby's overall philosophical framework is at odds with autopoiesis and the enactive paradigm followed (among others) by DiPaolo in his recovery of Ashby's ultrastability in Franchi (2011b: 41–45).

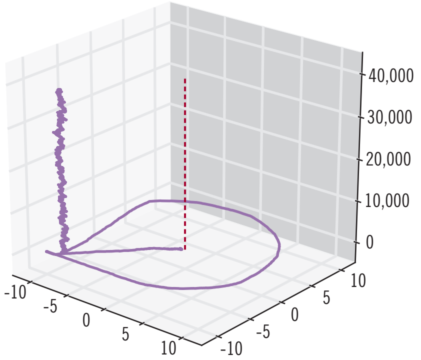


Figure 3: Temporal phase-space of a 2-unit homeostat reaching a local minimum (scenario S_2 in the text). The x-axis represents the value of the self-connected unit (or "agent"); the y-axis represents the value of the non self-connected unit (or "environment); the z-axis is time. Perfect equilibrium ($x = y = 0$) would be represented by a vertical line at the center of the chart, while the experimental homeostat reaches a local equilibrium at $y \approx \max$, $x \approx 8.5$. The tolerance interval was $-10 < x < 10$.

to unit j. There are no external inputs, as the homeostat is self-contained (mechanical analogies aside, m obviously represents a unit's inertia and it is similar to a neuron's bias, while v is a unit's damping factor). A single, negatively self-connected unit (Ashby's base case) turns into a harmonic oscillator, since the summation term reduces to $-w_{jj}y$. The oscillator is damped if viscosity $v > 0$.

« 19 » A unit has a critical oscillation band that, when exceeded, will trigger a random reset of its weights until the value is stable within the oscillation band. This means that a single self-connected unit will necessarily fall back on a negative self-connection weight and therefore behave as a damped harmonic oscillator. It also means that, as Ashby's experiments show, the damped oscillator is the homeostat's basic building block. A unit is just a building block, though, since a 1-unit homeostat has no conceptual meaning, being unable to interact with the environment. In other words, the minimally significant homeostat includes 2 units (Agent–Environment).

General features of a homeostat simulator

« 20 » The simulator has been built accordingly, with the main parameters for a unit being: mass, viscosity, oscillation band, noise, uniselector timing and uniselector function. The step mechanism aside, the behavior of a unit is determined by the equation above, modified to include noise, both on the connections and internally generated, and with values clipped within the unit's oscillation band.

« 21 » The units' displacement values are not computed using a numeric integrator – although that is planned for future versions – but using a discrete time simulation, using standard equations of motions from elementary physics. Force is computed as the weighted, noise-affected sum of inputs minus a velocity-proportional drag determined by viscosity.

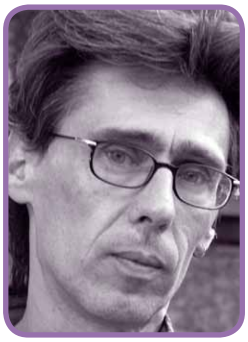
Single unit setup

« 22 » As previously mentioned, a homeostat comprising a single unit does not have any physical or biological correspondence. In particular, a 1-unit homeostat is not equivalent to a 1-neuron neural network (of whatever kind) or, more generally, to a 1-neuron organism. In Ashby's setup, the minimal homeostat modeling a real-world state of affairs includes two interacting units, corresponding to an agent and an environment. The goal of a study of a 1-unit homeostat is simply to determine the basic minimum conditions that result in a stable unit. Units exploiting those conditions will later be used as building blocks for "real" homeostats.

« 23 » The main parameters affecting a single unit are the presence of a positively or negatively weighted and potentially noisy self-connection, its general characteristics (mass, viscosity and internal noise) and the timing of its step mechanism. The main parameters affecting a unit's stability are its "mass" and "viscosity." Obviously, the lower the mass, the higher the sensitivity to force, and therefore the higher the instability. Conversely, a low viscosity value will lower or eliminate all possibility of reaching stability.

« 24 » Leaving the step mechanism (the "uniselector," in Ashby's language) aside for a moment, and focusing on the self-connection (or lack thereof), we can see that a positively self-connected unit will exhibit a runaway behavior. Conversely, a negatively self-connected unit will behave as a harmonic oscillator in the absence of viscosity, and as a damped oscillator otherwise. Finally, the value of a non-connected unit will tend to drift depending on noise. If the ultrastability mechanism – the "uniselector" – is activated, it would therefore seem likely that a unit will rapidly converge on a negative self-connection. This is indeed the case in experimental runs, provided that the timing of the ultrastability mechanism is sufficiently slow. As Ashby himself noted, the ultrastability mechanism must not check the value of the essential variable too often. After connection weights have been rearranged, a new behavior must be free to develop and possibly move the unit toward equilibrium. The ultrastability mechanism timing is partially dependent on the unit's "mass," that is, its inertia. Units with higher mass will react more slowly to change and require higher timing values. Experimental runs indicate that a minimum effective value is around 1 order of magnitude bigger than the basic execution cycle.

« 25 » There is one final element of the homeostat's behavior that deserves a brief discussion. Consider what may happen when a unit's value reaches its maximum or minimum value. In Ashby's physical device, this happens when the unit's needle hits one of the two walls delimiting the trough it moves in. What happens to the needle's velocity? Should it go to zero or should it just change sign? In physical terms, the first alternative is equivalent to hitting a perfectly inelastic wall and the second one is, conversely, equivalent to hitting a perfectly elastic one. The first solution is obviously what Ashby's device, in virtue of its construction, necessarily entails. As a virtual homeostat is not constrained by physical details, we are free to choose either alternative on conceptual grounds. Let me exploit the homeostat's physical metaphor to illustrate why I think Ashbian behavior (i.e., velocity reaching 0) is more consistent with the homeostat's overall setup. If we think of a unit's displacement as directly corresponding to the intensity of its actions on the external environment, and a unit's velocity, consequently, as the rate of change of the intensity of its actions over time, zeroing velocity when a unit's essential variable goes out of bounds means that its effective actions will continue



STEFANO FRANCHI
is Associate Research Professor in the Department of Hispanic Studies, Texas A&M University. He has published extensively on twentieth century European philosophy, the history and philosophy of science in the twentieth century, and the history of artificial intelligence and cognitive science. He has edited several volumes on the philosophy of cognitive science and AI, including (with Güven Güzeldere) *Mechanical Bodies, Computational Minds* (2005) and (with Francesco Bianchini) *The Search for a Theory of Cognition. Early Mechanisms and New Ideas* (2011). Most of his publications are available on his website: http://www.cleinias.org/stefano

unabated. They will not change in intensity or direction. In other words, this is another instance of the essentially passive behavior embodied in the homeostat: it will continue acting the way it normally does until an outside force compels it to change course of action (from this particular perspective, the second-order step mechanism can be considered as being external to the unit's core first-level feedback setup).

« 26 » In conclusion, in its default setup, a homeostat's unit will be negatively self-connected and act in a viscous fluid with a suitably high ultrastability timing. In practice, it will act as a damped harmonic oscillator. This conclusion only applies to self-connected units, though, which would represent agents in Ashby's model. It is difficult to envision a unit representing the particular aspect of the environment the agent is connected to as self-regulating though negative feedback. Such a unit would not be self-connected and would therefore exhibit an indifferent (drifting) behavior. I will come back to this issue in the next section.

Two-unit setups

« 27 » A 2-unit homeostat is the minimally significant setup in Ashby's model. It represents an agent interacting with an environment, where the latter term refers to any other part of the overall system under examination, be it the ecosystem external to the organism or a different subsystem within the organism itself. Although the difference is irrelevant from the systemic point of view Ashby advocates, it has some important consequences on the setup of a minimal agent–environment system. As noted above, an elementary agent's default configuration is a negative, damped self-connection, for that will be the outcome of its double-feedback mechanism. Should we assume the environment the agent interacts with to be similarly self-regulating? That possibility may make some sense when the "environment" in question is another sub-system within the organism, but seems to be less plausible when it is the external ecosystem. In *Design for a Brain,* Ashby insists on the necessity to recognize that the "terrestrial environments" are fundamentally stable and largely disjointed. As he puts it, "the mammal stands on ground that is almost always immobile; tree-trunks keep their positions" (Ashby 1960: 193). Yet, do the overall constancies of the environment translate into active negatively-damped behaviors? The answer to this question will substantially affect the equilibrium conditions of a 2-unit homeostat. There are only two basic configurations of such a system: it will either have a pair of identically configured self-regulating units or it will have a self-connected and a non self-connected one. All other options, besides being biologically meaningless, will rapidly run out of equilibrium. The first configuration, as Ashby correctly states, is stable and very robust. That is, an alteration of either connection's polarity will rapidly trigger the step mechanism and restore equilibrium.

« 28 » The second configuration – where one of the two units is not self-connected – exhibits a more interesting behavior. There are three possible scenarios:

S_1 The system will reach equilibrium with both units at around 0;
S_2 The self-connected unit will reach a stable equilibrium at a fixed value different from 0 but within its admissible interval, while the other one will reach its maximum or minimum values (see Figure 3);
S_3 Neither unit will reach equilibrium, with both units constantly oscillating within their admissible intervals.

In experimental runs, scenario S_1 is more resilient to external perturbations than S_2, which is more resilient than S_3.

« 29 » In other words, both S_2 and S_3 act as local minima within the general landscape, with S_2's being relatively lower, as it were, than S_3's. The next section will discuss a fairly abstract physical interpretation of the three cases in terms of elementary robots (i.e., Braitenberg-like vehicles).[16]

16| In *Vehicles*, Valentino Braitenberg (1984) introduced several kinds of very simple robots as conceptual tools for the discussion of many issues in psychology and the philosophy of mind, such as memory, bilateral symmetry and the crossing of neural pathways, emotions and so forth. Braitenberg's work has been so successful that his simpler models (e.g., Type 1 and Type 2) are routinely used in AI and robotics to introduce students and readers to non-symbolic, purely reactive cognitive architectures that determine behavior on the basis of un-mediated connections between sensors and actuators. A Type-1 Braitenberg vehicle is a simple device possessing a single sensor (a light detector) and a single actuator (a motor connected to a single wheel). The input from the sensor is transmitted directly to the motor. The result is that the vehicle will move back and forth depending on the intensity of the light stimulus it receives and it will actively "try" to run away from light sources (since a higher light intensity will immediately translate into higher motor speeds). The example robot discussed in the next section and illustrated in Figure 5 adopts this Type 1 architecture but replaces Braitenberg's original direct connection between sensor and actuator with the homeostatic unit just discussed.

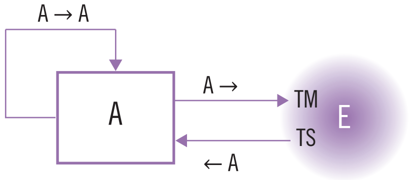


Figure 4: Physico-biological translation of a 2-unit 1-,2x homeostat with a negatively self-connected unit *A* and a non-self-connected unit *E*. *A* and *E* stand for Agent and Environment and are connected to each other by sensory (*A*←) and motor (*A*→) transducers (TS and TM, respectively).

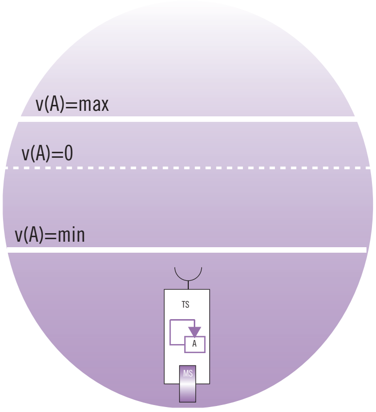


Figure 5: A Braitenberg-like Type 1 vehicle embedded in an environmental gradient. The two solid lines delimit the minimum and maximum acceptable values of the variable gradient for the vehicle, while the dotted line represents the optimal equilibrium. The half-circle in front represents a surface sensitive to some kind of input (light, for instance), which the sensor transducer conveys to the robot. The motor transducer, conversely, transmits output from the robot to the environment by operating the motor connected to the robot's (only) back wheel. The unit A inside the robot controls the connection between sensory input and output. With reference to Ashby's diagram of a minimal homeostat, which I reproduced in Figure 1, the sensor transducer corresponds to the *Env* to *R* thick arrow, the motor transducer corresponds to the *R* to *Env* thick arrow and unit A corresponds to the organism and includes the *D-S* subsystem. The gray circled area is the environment *Env*.

Agent–environment

« 30 » Figure 4 shows a physico-biological translation of the 2-unit homeostat setup with only one unit self-connected. Let us now translate the various scenarios sketched above into their "real world" counterpart. Figure 4 shows a functional diagram, while Figure 5 translates it into a virtual setting with an environmental gradient for some variable to which the unit is assumed to be sensitive.

« 31 » If we look at the relevant parameters of the homeostat and translate them into a vehicle's behavioral aspects, we see that the value of unit A, $v(A)$, refers to the value of A's essential variable. Since A is directly connected to a motor transducer, any $v(A) \neq 0$ will indicate movement in either direction, while $v(A) = 0$ implies the vehicle is at rest. The connection between the agent and the environment ($A \rightarrow E$) represents the agent's capacity to affect (or manipulate) its environment; its weight w is simply the intensity of the motor action expressed as a fraction of the unit's current value $v(A)$ at time t. In other words, $w(A \rightarrow E)$ expresses how "quickly," or perhaps how "intensely," the agent responds to the environment. More generally, the connection represents the agent's capacity to manipulate its relationship to the environment, including its location in it. Hence it stands for the agent's motor skills (of whatever kind and very broadly conceived). In perhaps the simplest possible case, it may stand for the agent's capacity to move around and therefore change its relative position within a given environment.[17] Conversely, the $E \rightarrow A$ connection stands for the sensory input and its weight w expresses the "receptivity" of the vehicle to its input source.

« 32 » What is the value of unit E, $v(E)$? It is not the value of some absolute, agent-independent, environment variable, for $v(E)$ always changes in response to the agent's actions $A \rightarrow E$, actions which do not necessarily alter the environment per se. $v(E)_{max}$ thus represents the maximum value that could be possibly sensed by E. In other words, $v(E)$ is the raw value of *the environment relative to the agent*. Thus, $v(E)_{max}$ and $v(E)_{min}$ would indicate that the input source has reached its maximum and minimum possible level, respectively, from the agent's standpoint. On this basis, let us now translate the three scenarios S_1 to S_3 above into the physical setting represented in Figure 5. When $v(A) \approx v(E) \approx 0$ (scenario S_1), the vehicle has reached an optimal equilibrium point and it is resting around the dotted line in the figure. When, conversely, the value $v(A)$ oscillates with $v(A)_{min} < v(A) < v(A)_{max}$ (scenario S_3), the vehicle periodically overshoots its equilibrium point and keeps moving back and forth between the two solid lines. Finally, when the $v(A)$ is stable at a value different from 0 when $v(E)$ is either at a minimum or at a maximum, the vehicle is continuously moving while keeping its essential variable within tolerable limits. In other words, the vehicle has reached a stable yet dynamic equilibrium: its constant motion keeps it within a tolerance zone without allowing it to climb it all the way to the top.

Two experiments from Ashby's *Design for a Brain*

« 33 » The setup just discussed bears a direct relationship to the first concrete appearance of the device in Ashby's work, which he labels "The homeostat as an adaptor." Ashby shows the robustness of the ultrastability mechanism by starting with a setup perfectly analogous to the one we just saw. He then reverses the polarity of the connection between agent and environment and observes how the ultrastability mechanism quickly restores equilibrium under different conditions. Ashby's text does not specify whether both units are self-connected or just the first one (the "Agent"). Experimental runs behind the previous discussions, however, show that we can reach the level of robustness and perfectly restored equilibrium exhibited by the book's device only with the relatively implausible setup that includes two self-connected units. This does not mean that the homeostat's capacity to adapt to mutated conditions is in any way reliant on such an implausible assumption. On the contrary, the adoption of the more

---

[17] This definition is so general as to actually include any kind of relative movement, whether the agent is endowed with locomotion or not. For instance a plant leaf's capacity to turn in the direction of incoming light during phototropism is as valid a "motor action" as a bacterium's flagellate action during chemotaxis.

likely setup described above leads, I believe, to an even more interesting conclusion. It shows that the agents modeled by Ashby's device are "adaptors" that follow the environment even more closely and "sleepily" (to use Grey Walter's image) than Ashby envisioned. They can get stuck in partial, dynamic equilibria representing local minima of their adaptivity function, as it were, and only repeated perturbations in their relationship to the environment will eventually lead them to a more restful position.

« 34 » Similar considerations apply to the second experiment carried out in *Design for a Brain*, a three-unit setup that can be interpreted as modeling a more sophisticated agent with separate motor (*M*) and sensor (*S*) sub-parts connected to an environment (*E*) in a circular fashion: *E* → *S* → *M* → *E*. Ashby's experiment shows that even a dispersed, poorly joined system lacking direct mutual connections among all its subparts can restore equilibrium by indirect communication through the environment, as it were (see Figure 6). As in the first experiment just discussed above, though, the results Ashby reports could only be replicated exactly by assuming an implausible self-connection in the unit representing the environment. The more likely setup, without a self-connected "environment," leads nonetheless to equilibrium, as shown in the figure, even though the path to it is less direct than the trace in *Design for a Brain* would imply.

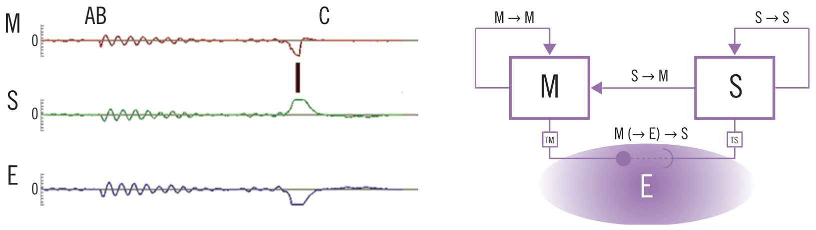


Figure 6: A replica of Ashby's second experiment in *Design for a Brain* (Ashby 1960: 106). Three units (roughly interpretable as Agent–Motor, Agent–Sensor, and Environment) are connected circularly (*M* → *E* → *S* → *M*). *M* and *S* are self-regulated, *E* is not. The system is originally stable. At *A*, the value of *M* is manually decreased, provoking a local instability and resulting in oscillating behavior until equilibrium is restored. At *B*, the polarity of the *E* → *S* connection is switched (as if a sensory transducer's mode of functioning had been reversed) and a downward manual displacement of *M* is repeated. The system becomes unstable until three actions of the step mechanism restore stability at *C*.

Conclusion

« 35 » As discussed in the first two sections, the goal of this project is to recover as much as possible the intuition behind Ashby's generalized homeostasis thesis and extend it further while keeping it focused on the conceptual side of his simulations and leaving intentionally in the background the neuro-physiological (or psychological, or simply chemical) equivalents. The first steps in the project involved the construction of a virtual yet faithful replica of Ashby's homeostat, an abstract preliminary and still-qualitative analysis of its behavior in standard configurations and some measure of validation from a replica of Ashby's original experiments as described in his works. The analyses in the previous two sections have provided the initial confirmation that was sought and suggested that a modest yet potentially relevant relaxation of Ashby's original setup may still lead the system to equilibrium, and perhaps in more interesting ways. I think these results help to clarify a perspective of life as a passive-contingent phenomenon, to use the terminology I introduced earlier. We can undertake the move from mere plausibility toward a more substantive view in a number of ways. One possibility, already mentioned, is to extend the homeostat model to a simulated robotic environment and to test basic homeostatic configurations as controllers of creatures engaged in minimal tasks such as different forms of mono- and multi-sensor taxis behavior. Of particular interest would also be the simulation of an elementary multi-tasking goal involving two non-necessarily cooperative sub-tasks such as, for instance, a simplified form of obstacle avoidance carried out together with a taxis behavior, or, alternatively, forms of taxis involving multiple sensory modalities. It is well-known, although often forgotten, that Ashby claimed his homeostat could achieve progressively higher forms of adaptation even if it lacked any memory, be it explicit or implicit (i.e., as a form of Hebbian learning). A complex homeostatic system immersed in a similarly complex semi-stable environment, he claimed, can reach a reasonable level of adaptation in biological (as opposed to "geological") time. Indeed, the last third of *Design for a Brain*, and especially the whole of chapter 16, are devoted to a detailed discussion of how a "multistable" poorly joined and therefore dispersed system can occur "in a time that is not impossibly long" (Ashby 1960: 214). It is fair to say, however, that the concrete evidence he presented in support of this claim (the last experiment on the homeostat described in the pages immediately preceding this claim) has appeared to many readers as being rather contrived and generally underwhelming. Ashby himself must have shared his readers' perplexity, for it is known that in the years after the publication of the first edition of *Design for a Brain*, he was working on a more sophisticated device (Dams, *Dispersive and Multistable System*, Ashby 1952b: 171), which he was eventually unable to develop fully (Husbands & Holland 2008: 125f.). I think that concrete work in evolutionary robotics settings carried out along the lines I suggested in this paper may provide the evidence Ashby failed to deliver – in large part, undoubtedly, because of the primitive computational technology he had available – and reopen a theoretical discussion on the philosophically most daring aspects of his theory of life as general homeostasis.

Received: 18 June 2013
Accepted: 19 September 2013

Open Peer Commentaries

on Stefano Franchi's "Homeostats for the 21st Century?"

Standing on the Broad Shoulders of Ashby

Inman Harvey
University of Sussex, Brighton UK
inmanh/at/gmail.com

> Upshot • It is a mistake to characterise Ashby's view of life (from the 1950s) as passive, abstractly modelled in part by the homeostat; one should distinguish the stasis of homeostasis from the activity of the (model) organism. Likewise mistaken is the accusation of contingency; one should distinguish the purposeless mechanism from the purposeful (model) organism. There is no basic conflict between Ashby's view and later developments in a similar tradition; technical advances are not the same as foundational gaps.

« 1 » I had some trouble recognising the nature of Stefano Franchi's criticisms of Ross Ashby, since they seemed largely to ignore those aspects of his work that make me rate him a giant of the twentieth century. I am a critical fan of Ashby, since I think he got quite a few technical issues wrong; but in my view, as a man of his time he asked the right sort of questions, phrased in novel form by framing cognition in terms of dynamical systems, and provided inspiration for many who built on and extended his ideas. At first sight, Franchi's reading of Ashby's writings focusses on some heteronomous-autonomous distinction (§12), whereas Ashby, as far as I was aware, made little or no obvious comment on this. Ashby's focus on adaptation seems orthogonal to that distinction.

« 2 » It can be a revelation to see how some complex, many-layered piece of writing can be interpreted differently through different eyes. Though my response here is directed at the target article, some extra background to Franchi's perspective is given in Franchi (2011b); with these two articles in mind, I welcomed the excuse to re-read Ashby's *Design for a Brain*.

« 3 » I was immediately struck, as always, by the directness and honesty of his writing; the somewhat dated style reflects the date it was written, 1952 for the first published version. This placed it two years after Turing's (1950) *Mind* paper on "Computing Machinery and Intelligence," and four years before the 1956 Dartmouth conference usually thought of as defining (the GOFAI version of) AI based on a computational theory of mind. This dates and places Ashby's book at a major fork in the cognitive road, pointing clearly down the left branch (relatively rather neglected for some decades thereafter) defining life and cognition in terms of biology and adaptivity, using concepts from cybernetics and dynamical systems theory. In contrast, the right branch (soon to become the multi-lane highway of classical AI or GOFAI) defines life and cognition in terms of logic and rationality, using concepts from computational theory.

« 4 » In his review, Franchi wants to contrast Ashby with "the autonomy-autopoiesis framework" that we can tag with the name Francisco Varela. Now all would agree that Ashby and Varela are somewhere along the same left-hand branch of this major fork, and most would agree that even where not explicitly constructivist this branch is at least constructivism-friendly. I would claim that the main differences between Ashby and Varela relate basically to their earlier and later positions on this branch, the latter extending and going beyond the ideas of the former but with no fundamental contradictions between them. However, Franchi claims some unbridgeable and foundational gap.

« 5 » Franchi starts (§3) by positing a distinction between simulating objects and simulating concepts. I cannot make sense of "simulating concepts," unless perhaps this can be rephrased as "simulating objects at a rather abstract level." In any modeling exercise, one chooses a subset of all the properties of an object, one simplifies according to the needs and context. So when modelling traffic flow, I might choose – depending on the motivation for the simulation – to simulate the specific properties (and idiosyncracies) of my car, or of any car of its class, or of some more abstract vehicle-in-general. These are shades of grey, a spectrum from less to more abstract rather than some black-and-white step-change from object to concept, but we can agree with Franchi that Ashby's models are at the more abstract end of the scale. At the end of Chapter 2 of *Design for a Brain* Ashby says:

> "…this book will attend closely to certain idealised cases… Maybe it will be found eventually that not a single mechanism in the brain corresponds *exactly* to the types described here; nevertheless the work will not be wasted if a thorough knowledge of these idealised forms enables us to understand the workings of many mechanisms that resemble them only as approximations." (Ashby 1960: 29)

« 6 » Franchi's main issue (§6) with Ashby is with what he calls Ashby's first thesis: "(1) A complex system running to equilibrium generates a complex behavior." As this stands, this makes no claims about life or cognition or autonomy or learning or adaptation; it seems (if "generates" is altered to "may generate") an innocuous uncontroversial statement about a class of mathematical models. But Franchi interprets this as a "very general and broad thesis about life" and sees the Homeostat as an exemplar for this that has non-action, or "*going back to sleep*" as its core principle.

« 7 » An initial problem with Franchi's criticism is that Ashby (1960) does not present the homeostat, or homeostasis, as representing life or cognition in general. He makes it very clear that he is concerned very specifically with *adaptation*; what type of mechanism could explain how a curious and naive kitten, on first putting its paw into the fire and getting burnt, learns not to do this next time? But if we allow this extension to life-in-general, the central issue that Franchi raises is that in his eyes the homeostat (and by extension the Ashbian organism) is (§10) "…essentially a *passive* machine whose activity is a byproduct of its search for non-action. It is also a *contingent* machine insofar as its "search" for equilibrium will involve essentially random processes."

« 8 » First let us demolish the *passivity* issue. Franchi repeats Grey Walter's jibe that the homeostat is a *machina sopora*, and casts Ashby as claiming that organisms seek quiescence; but this is to misattribute the entity to which the *stasis* in *homeostasis* refers. Within Ashby's framework, what the homeostat tends to maintain, in the face of perturbations, is a continuance of the agent-environment system in a form where the essential variables (EVs) remain within their bounds. But the stasis of that property – a property of the agent-environment system – does not imply any stasis or quiescence in the organism itself. As an example, a bicycle with a homeostat-inspired control mechanism might wobble along, perhaps lurching from side to side, whilst maintaining its EV (e.g., angle from vertical) within acceptable bounds. The stasis of the EV is not the stasis of the moving bicycle; indeed, the continued active control of the bicycle is essential for that stasis – it is no passive machine. An Irish peasant in the 19th century potato famine may have taken the bold decision to emigrate to America; the very active planning and travelling (to be described at the human level) is in no way contradicted by the associated (homeo-)stasis of the peasant's viability as a living organism.

« 9 » Second, let us examine the *contingency* issue. Ashby makes clear that his focus is on adaptation, as with the kitten and fire example; and also makes explicit that he will use no teleological explanation for behaviour – to avoid circular argument. It follows that his mechanism must rely on randomness and contingency in the sense of being non-teleological. But once again, we must be careful to distinguish the different levels of explanation: contingency and lack of purpose at the mechanism level is shown to enable purposive behaviour at the organism level. In the homeostat, with little sense of history and no social dimension, this will be the basic low-level purposiveness equivalent to the described adaptive behaviour of a kitten; but this lays the foundation for potentially understanding much more sophisticated purposiveness, for instance extended through personal history and social interaction. It is essential to Ashby's argument that the mechanism be not purposive yet that the organism demonstrate purpose; to castigate this as describing a contingent machine is to misunderstand his argument, to miss the subtlety of describing the machine-organism at two different levels.

« 10 » Franchi goes on to replicate some aspects of Ashby's homeostat (§21), but the points raised there do not particularly affect the mistaken central assertion that Ashby was presenting a perspective of life as a passive-contingent phenomenon; he was doing no such thing.

« 11 » Though my own work in evolutionary robotics and cognition (e.g., Harvey et al. 1997) has been immensely influenced by Ashby, I have my own criticisms of him. I think it unfortunate that he glossed over a crucial distinction between two distinct meanings of "essential variables" (i.e., any one specific EV may be interpreted from two different perspectives, giving rise to two different values): a first (EV-*sensu*-viability) meaning that relates directly to the chances of survival of the organism, and a second (EV-*sensu*-signal) meaning that relates to some sensory perception used by the homeostatic mechanism as the second form of feedback. Necessarily, the second variable cannot be identical to the first, although some correlation between them is necessary. A related mistake was that Ashby made both these EVs in step-function form (indeed made them one and the same step-function), whereas at least one, if not both, needs to be graded rather than stepped, for instance by equating the EV-*sensu*-viability to the current probability-rate of dying, thus incorporating noise. With the benefit of hindsight, with knowledge gained in catastrophe theory and the notion of phase transitions that has now reached wider public consciousness as "tipping-points," we now realise that step transitions can arise without step-functions (Harvey 2008). A further claim of Ashby's (that Franchi lists as claim 5) is, I believe, unduly and grossly pessimistic; he suggests that a homeostat of 1000 units or more would be almost certainly unstable whereas I believe the converse is true when the interactions are nonlinear and the variables refer to physically bounded quantities (Harvey 2011).

« 12 » Despite these (and many other) technical disagreements, I see Ashby standing as a giant from the mid-twentieth century, with subsequent advances building on his achievements. The directness of his writing on the dynamical systems approach to cognition remains unsurpassed. He was clear in distinguishing between different levels of description. His work necessarily used explanations with circular causation, which can easily be misinterpreted by those who demand that these be recast in terms of linear causation. Varela and others have gone beyond his position, but to suggest this means some foundational gap seems as far-fetched as suggesting a foundational gap between Darwin in the mid-nineteenth century (who knew nothing of genetics) and the twentieth century neo-Darwinian synthesis.

« 13 » Extensions to Ashby's framework include moving from homeostasis to homeorhesis (i.e., stability of process or trajectory, rather than stability of state). One can take an autopoietic perspective on life as a self-maintaining organisation of flows of matter, driven by energy and entropy gradients while maintaining itself in a steady state far from equilibrium despite noise and perturbations. There is a close and natural fit with notions of viability and multiple feedback loops in Gaia Theory and Daisyworld models (Harvey 2011). All these later developments go beyond Ashby's relatively modest homeostat, but are very compatible with the philosophy behind it and fit well with Ashby's analysis of adaptation. Technical advances that imply discarding or revising some of his statements should not be equated with unbridgeable philosophical gaps. Ashby's shoulders are broad and well worth standing on.

Inman Harvey, with an academic background in mathematics, philosophy and social anthropology, was a founding member of the Evolutionary and Adaptive Systems (EASy) group at the University of Sussex. His interests include evolutionary robotics, artificial life, a dynamical systems perspective on cognitive science, and Gaia theory.

Received: 3 October 2013
Accepted: 8 October 2013

From Fechner, via Freud and Pavlov, to Ashby

Helge Malmgren
University of Gothenburg, Sweden
helge.malmgren/at/filosofi.gu.se

> Upshot • Ashby's view of the organism as an essentially passive machine is not quite as original as the target article may suggest, since it can be traced to Freud's pleasure principle and from there back to Fechner's ideas about different kinds of stability in deterministic systems. A modification of the author's distinction between "simulation of real objects" and "simulation of concepts" is also suggested, and it is argued that the main aim of Ashby's simulations was to explore the possibility of high-level (structural) explanations of real phenomena, especially learning and memory.

From Fechner via Freud

« 1 » Stefano Franchi's interesting paper has many merits but I will focus on three issues where I think it is somewhat misleading. My first point is not really an objection to what the author says but rather an extension of it.

« 2 » In the abstract, the author emphasizes that Ross Ashby's view of the organism as an essentially passive machine "is conceptually at odds with the autonomy-autopoiesis framework adopted by Ashby's recent defenders as well as by most of the Western philosophical tradition." This may well be so, but in order to complement the image of Ashby as an isolated thinker that this passage may provoke, I would like to point to another tradition. Henri Ellenberger (1970) was the first to highlight Gustav Theodor Fechner's early contributions to system theory and the influence they had on Sigmund Freud's theorizing. Fechner distinguished different kinds of tendencies towards stability in deterministic systems, including what we would now call point attractors and limit cycles. Freud, on the other hand, was the first neural networks theorist, and the desire of the organism to return to equilibrium plays a leading role in his speculations about the brain's subsystems of neural elements as well as in his later psychological thinking about the tripartite mind.

« 3 » Ashby was, of course, a psychiatrist, not an engineer nor a mathematician nor a philosopher. He was a frequent contributor, mostly with book reviews but also with original papers such as Ashby (1954b), to the *British Journal of Psychiatry*. I have not been able to find any statement by Ashby about a direct connection between Freud's and Ashby's first principles. But there is solid evidence that Freud influenced and inspired Ashby. A commentator to an early paper by Ashby on the physiological basis of neurosis finds that "Dr. Ross Ashby has given us Freud in terms of the Gestalt school" (Clifford Allen in Ashby 1933: 1460), and in Ashby (1954b: 122), a closer study of the relation between the basic instinctual forces of psychoanalysis and the "essential variables" of his own systems is recommended. Another telling – and at the same time moving – piece of evidence was found by in his private notebook by Alan Pickering. Early in 1951 Ashby writes:

> "Others may talk of freewill and the individual's power to direct his life's story. My personal experience has convinced me over and over again that my power of control is great – where it doesn't matter: but at the important times, in the words of Freud, I do not live but 'am lived'." (Pickering 2010: 112).

What is "I'm lived" if not an expression of the view that the organism is essentially passive?

From Pavlov to Ashby

« 4 » I find the paper by Franchi a little imbalanced in that it (to use an understatement) does not place any emphasis on *learning* and *memory*. Indeed, the only time these words occur is when the author quotes Ashby as saying that the homeostat "could achieve progressively higher forms of adaptation even if it lacked any memory, be it explicit or implicit (i.e., as a form of Hebbian learning)" (§35). This may seem to imply that the forms of adaptation that interested Ashby most were not connected to learning or memory. But clearly one of Ashby's expressed main aims, if not *the* main aim, was to explain learned behavior: "Our problem may now be stated in preliminary form: what cerebral changes occur during the learning process, and why does the behaviour usually change to the better?" (Ashby 1960: 4). And for me, "explicit or implicit memory" and "cerebral changes that occur during the learning process" are almost synonymous terms.

« 5 » Below, I will sketch a way to reconcile the quote on memory from the author and the quote on learning from Ashby. But let us first note that Ashby a few pages earlier (Ashby 1960: 2) refers to Ivan Pavlov's experiment on conditioned reflexes. Ashby was actually ambivalent to Pavlov as a scientist. An early short review reveals that he was "thrilled by Pavlov's discoveries" (Ashby 1951). In a later, longer review (Ashby 1957) he however distinguishes sharply between Pavlov the experimenter and Pavlov the thinker. He does not hold the latter in high esteem and finds his attempts to *explain* the phenomena of conditioning severely lacking. In Ashby's opinion (Ashby 1957: 252), nobody has yet been able to explain the more complex phenomena that Pavlov describes in a plausible way, but he thinks that his own methods (especially those introduced in Ashby 1956b) are very promising in this respect.

« 6 » How shall we reconcile this expressed ambition of Ashby's to explain learning with Franchi's de-emphasis on learning and memory? Here is one possibility. I could agree with the author's suggestion that Ashby thought that his machines did not need any memory to adapt, if by "memory" is meant *memory as conceptualized in current theories.* Ashby was indeed discontent not only with the explanations of memory and learning current at his time, but also with the way in which the basic concepts used in these explanations were defined, or (most often) not defined. The 1957 review of Pavlov provides several examples, another one is his rather furious comment about another author:

> “He attempts a synthesis, but in fact only shows, how confused are our present ways of thinking about the subject. Memory is at one moment an image, then a power of recall, then a trace, an activation of a trace or a process and so on – all used wherever they happen to fit, with practically no attention to how all these different concepts are related.” (Ashby 1971: 574)

But of course this does not mean that Ashby wants to eliminate *memory* from his own theory! "Design for a Brain" is not about memory images or memory traces; however, to conclude that it is not about memory and learning would be plainly wrong.

Simulation of concepts vs. simulation using (revised) concepts

« 7 » So, Ashby is critical of the concepts used in contemporary learning theory. How shall one characterize the conceptual work that Ashby is doing to remedy the situation? What role does it play in his endeavour as a whole? And what is the connection between the conceptual work and his simulations? I must confess that, for purely linguistic reasons, I do not find "simulation of concepts" to be a good term. According to most definitions of "simulation," what can be simulated is a chain of events, a process, or a device that typically goes through a certain process: you can simulate (the time course of) a hurricane, or a house being hit and destroyed by the hurricane, or the trajectories of a network in the brain. But a concept is not a chain of event or a process, and does not typically go through a process. It cannot be simulated in this sense. (Of course, the *development* of a concept over time could be simulated, but that is not the issue.)

« 8 » In my opinion, the author's formulation "simulation of concepts" covers (and partly hides) three things that Ashby does with concepts:

1 | He *revises* the current concepts of learning and memory, making them much more general; and
2 | he *illustrates* the revised concepts by
3 | *simulating certain physical processes* that by necessity lead to learning and memory (in the revised sense).

« 9 » In short, he is doing simulations *using* revised concepts. This explication of Ashby's conceptual work fits nicely with his explanatory ambition and therefore lessens the contrast between the two major components of his work that Franchi describes. The connection is that the simulations point to a form or structure that an adequate low-level explanation in terms of (for example) memory traces or modifications of synapses might have. Namely, the same structure as the simulated physical processes. I would like to speak about high-level, "structural" explanations here. To be sure, the structural explanations may look strange to us at first sight since the concepts used in them are unfamiliar. And to be sure, they do not explain all details. These we have to fill in by describing the workings of the concrete machines (e.g., brains) that we are interested in.

« 10 » Maybe this is not the only possible way of reading Ashby, but it is a reading that makes his endeavour clearly understandable as a coherent whole. And to add a personal note, it is this reading that the reviewer has himself used when trying to carry Ashby's work a little further (Malmgren 2006).

Helge Malmgren is a Ph.D and senior professor in theoretical philosophy. He is also an MD and is doing interdisciplinary work at the interface between philosophy, medicine and psychology. This work includes philosophy of mind, concept formation in psychiatry and neural networks for medical decision support.

RECEIVED: 19 OCTOBER 2013
ACCEPTED: 21 OCTOBER 2013

Designing for Emergent Ultrastable Behaviour in Complex Artificial Systems – The Quest for Minimizing Heteronomous Constraints

Robert Lowe
University of Skövde, Sweden
robert.lowe/at/his.se

> Upshot • The target article has addressed core concepts of Ashby's generalized homeostasis thesis as well as its relevance to building complex artificial systems. In this commentary, I discuss Ashby-inspired approaches to designing for ultrastable behaviour in robots and the extent to which complex adaptive behaviour can be underdetermined by heteronomous constraints.

« 1 » In §35, Stefano Franchi discusses the possibility of utilizing Ross Ashby's homeostat model in a simulated robotic environment in order to engage in not just "minimal" tasks but those that require subtasks or parallel tasks (e.g., phototaxis whilst avoiding obstacles). In this commentary, I discuss existing Ashby-inspired robotics work and methodologies that consider this problem of *scaling*, a problem that Franchi suggests the DAMS project of Ashby failed to address adequately.

« 2 » Ultrastability offers an approach to understanding cognition without recourse to *strong* design by exploiting random processes that, given enough time, permit (re-)establishment of agent-environment equilibrium. Ashbyan ultrastable agents, including Ashby's exemplar, the homeostat, are nevertheless not immune to the effects of design decisions – they are heteronomously influenced. For the homeostat, design concerns set the internal homeostatic oscillatory ranges within which negative-feedback-induced operation errors are absent (when the system is in equilibrium). For Ashby-inspired robotic agents purported to be imbued with ultrastability, the design typically concerns set ranges within which essential variables (e.g., battery level) are "comfortable" (cf. Di Paolo 2003; Avila-Garcìa & Cañamero 2005; Pitonakova 2013), similarly signifying that the system is in equilibrium. Design decisions for ultrastable-like robots may further be categorized according to their strictness of adherence to the notion of ultrastability:

- *Random self-configurable approaches* (Di Paolo 2000; Pitonakova 2013): for example, individual neurons within continuous time recurrent neural networks (CTRNNs) modulate weighted connections in a (Gaussian) random manor until neural activation homeostasis is (re-) established.
- *Non-random self-configurable approaches* (Manicka & Di Paolo 2009; Lowe et al. 2010): where parameters, including those of neural network transfer functions, may be directionally modulated.
- *Non self-configurable approaches* (Avila-García & Cañamero 2005; Montebelli et al. 2010): where internal parameters are not changed but nevertheless, as in (a) and (b), internal and behavioural homeostasis (the basic, double feedback systems – see §12) based on the use of essential variables may be achieved.

Furthermore, Ashby-like robotics approaches may be classified into those that use one essential variable type, e.g., energy or neural activity, and those that use two or more essential variable types, e.g., fuel and water levels. The above classifications can be linked to the design-informing concepts of Franchi outlined in §6 and §7, of which there are a total of seven.

« 3 » Of particular interest to the above-identified ultrastable-like designs for robots is point 4, which can be described as follows: In an ultrastable system, parameter changes, induced by operational error (essential variable disequilibrium), that permit new agent-environment interactions are modulated by a (non-directed) random process that necessitates trial-and-error agent-environment interactions. Ashby-inspired robotics approaches are faithful to at least the first two of the four "core" concepts given by Franchi (§6). Here, the focus is on producing robots that produce stable internal (essential variable) and behavioural (interactive) activity cycles. However, points 3 and 4 are not always faithfully adhered to (point 3 concerning *self-reconfiguration*). These points represent design decisions critical to the issue of scaling from minimalist to complex robotics design.

« 4 » Of the robotics designs most faithful to the Ashbyan concept – in relation to core concepts 1–4 – many suffer from the problem alluded to in Franchi's concept 5: systems with many self-reconfiguring components can take an age to achieve equilibrium or discover an adaptive process.

« 5 » Where core concept 4 is implemented, concept 5 tends to follow. This is apparent in Ezequiel Di Paolo's (2003) simulated phototactic robot, which is required to maintain its battery level (an essential variable) homeostatically. Only when parameters that permit the re-establishment of agent-environment interactive equilibrium (behavioural stability, cf. McFarland & Bösser 1993) are discovered is ultrastability achieved, but this requires a drawn out trial-and-error interactive process. It has been pointed out that such trial-and-error behaviours, even those that chance upon an equilibrium (re-)establishing behaviour, cannot be considered adaptive as small changes in the environment may render the particular behaviour insufficient to maintain equilibrium (cf. Pickering 2010).

« 6 » The problem with trial-and-error processes is not just their inefficiency and non-adaptivity in the context of a mobile biological or artificial agent (cf. Manicka & Di Paolo 2009, though also see Pitonakova 2013). In a complex, dynamic and hazardous environment, purely trial-and-error driven behaviour is *non-viable* – the agent risks encountering damage or even destruction if it repeatedly tries out the "wrong" behaviour. Usefully incorporating the ultrastable system concept in designing for complex artificial systems, thus, requires compromising on the purity of the Ashbyan vision whilst acknowledging the need to minimize the extent of design of the equilibrium (re-)establishing process.

« 7 » Key to *designing for* ultrastable behaviour – where, minimally, robots are able to achieve a double feedback loop – is determining whether such behaviour is "adaptive." Adaptive behaviour is not *designed* in ultrastable systems. The trial-and-error nature of the agent-environment interactive process precludes it from being *designed for* either (cf. Pickering 2010). As referred to at the beginning of this debate, design decisions affect internal homeostasis. Therefore, homeostats, and "concept 1–4" ultrastable robots, in this sense, are *heteronomous* in relation to their internal (essential variable) equilibrium; they are non-heteronomous in relation to their behavioural stability but also non-adaptive.

« 8 » Common to the pursuits of complex robotic behaviour grounded in adaptive interactive processes is the notion of *emergence*. Designing for emergence, at first glance, appears to be a misnomer – how can a phenomenon be emergent if it is *a priori* designed for? In robotics, designing for emergence is central to an approach that views intelligent behaviour as an embodied phenomenon. Rodney Brooks (1991) and Rolf Pfeifer & Christian Scheier (1999) are leading advocates of the notion that intelligent behaviour can emerge from the interaction of (in themselves potentially non-adaptive) constituents. An emergent behaviour may simply be described as one the presence of which depends on a mathematically intractable process requiring simulation and experiment for appropriate analytic description. The field of *evolutionary robotics* (cf. Nolfi & Floreano 2000) concerns itself with the identification of novel adaptive behaviours by exploiting the principle of emergence. Referred to in §13 in relation to identifying autonomy with minimalist design, as a methodology it can simply be considered as providing a means of grounding the agent's adaptive processes in its task environment. Through the use of computer simulations, and behaviour-shaping fitness functions that prioritize satisficing (survival) rather than optimization of a pre-given survival strategy, novel and adaptive patterns of behaviour can be identified. As with any simulation-based approach, evolutionary robotics entails heteronomous influence, in this case design of: fitness function, parameters subject to evolutionary modulation, task set-up, choice of evolutionary algorithm. A satisficing perspective in evolutionary robotics, however, reduces strong design of adaptive behaviour to adaptive/adapted behaviour that is *designed for*.

« 9 » Evolutionary robotics also provides a means for hierarchically grounding an agent's adaptive processes. In Lowe et al. (2010), we adopted an evolutionary robotics approach in order to ground artificial metabolic processes in an artificial neural network controller for a simulated e-puck robot. The robot's task was to survive/satisfice via selecting, over a number of episodes, between two variable replenishing resources. The robot-environment could be compared to an ultrastable system:

- Essential variable values were given by the level of "substrate" and "water" at the anode and cathode, respectively, of a simulated microbial fuel cell (cf. Melhuish et al. 2006).
- Thresholds, set by the genetic algorithm (GA), determined essential variable monitor nodes' "comfort" zones.
- Chemical node activation of the network adapted the slope of nodes' electrical activity output, a function that *directly* altered the robot's sensorimotor activity interaction with its environment.

« 10 » In this set-up, however, the GA determined the direction of the modulation of the output function slope – it was a *non-random self-configurable approach*. However, the directedness (e.g., action selection) of the *ultrastable* behaviour was not explicitly designed. It emerged from the evolutionarily *designing for* satisficing. Furthermore, this evolved satisficing was facilitated by making anticipatory use of active vision (a mobile low resolution *retina* superimposed on a two-dimensional camera view) to minimize the Euclidean distance of motor trajectories to the selected resource. By compromising on the fourth principle/concept identified in §6, the agent was adapted to its environment through use of its directed adaptive processes.[1] Adaptive/adapted behaviour was relatively complex, and we argued that it was the emergent product of (ultrastable) motivated decision-making grounded in energy constraints. The sensitivity to grounding was noted when energy constraints were varied to either higher or lower levels than those for which the robot-controller was evolved – in this case, the active vision was not in tune with the energy-constrained motor actuation and the robot would "saccade," either too early or too late relative to its motor movement. In other words, sensory, motor and energy constraints were intimately connected to the grounded non-random self-configurable ultrastable process.

« 11 » Alternative Ashby-inspired frameworks exist that seek to minimize heteronomous constraints whilst producing ever more complex behaviour. Such approaches typically fall into the category of *non self-configurable approaches* and do not adhere to Franchi's identified core concepts 3 and 4. David McFarland (2008) postulates that intelligent robots should be imbued with three core levels of autonomy: *energy*, *motivational*, *mental*. Autonomy for McFarland simply means *free from outside control*. In robot architectures that incorporate *motivational autonomy* processes, McFarland & Spier (1997), and also Avila-Garcìa & Cañamero (2005), utilize variables that are provisioned with homeostatic limits, where values beyond these limits comprise "physiological drive" errors. They are essential variables. In these works, there is a basic sense in which essential variables *non-randomly* influence *motivated* decision making. In line with McFarland's (2008) perspective, Ziemke & Lowe (2009, see also Lowe & Ziemke 2013) have suggested that McFarland's levels of autonomy should be hierarchichally, or perhaps bidirectionally, grounded in each other in order for the robot system to be considered adaptive. This is somewhat consistent with the approach in Lowe et al. (2010), where *motivational autonomy* processes (drive-based selection for navigating towards an external essential variable replenishing resource) are grounded – behaviourally and internally homeostatically constrained – by artificial metabolic activity (see also Montebelli et al. 2010 for related work). However, in that work, essential variable set points are not modulated by motivational autonomy processes – they are control variables. Lowe & Ziemke (2011) have considered that *mental autonomy* may involve a top-down regulatory influence on basic Ashby-like homeostasis, implementing a notion of bidirectional grounding from which *ultrastable* behaviour may emerge. This perspective is consistent with the modern notion of *allostasis* (cf. Sterling 2004), where individual essential variables' homeostatic regulation may be temporarily compromised in the service of the agent's adaptive/adapted behaviour. Biological agents are often required to operate outside their comfort zone – to be not merely *passive-contingent* (§11) – in order to be viable. This requires something of a shift away from one of the core conceptions of the homeostat – trial-and-error behaviour rooted in random processes. Complex adaptive systems are predictive and regulate their internal and behavioural processes accordingly. A given behaviour sequence may be entirely non-adaptive unless completed – this may require short-term discomfort (e.g., dangerously low energy levels) for long-term benefit (access to a large food supply). The agent must persist to meet long-term demands and resist, at least some of the time, opportunistic stimulation (cf. Avila-Garcìa & Cañamero 2005, Gaussier et al. 2000). The notion of allostasis has elsewhere been specifically applied to robotics scenarios in which the robot is required to produce non-equilibrium short-term behaviour for long-term benefit (e.g., override obstacle perception thresholds in order to navigate over ramps, see Muntean & Wright 2007).

« 12 » The allostasis notion of bidirectional predictive regulation reduces the effects of any heteronomous (explicit, e.g., intelligent, or implicit, e.g., evolutionary) design on robot behaviour. Designing for emergence in this sense serves to underdetermine the resulting complex adaptive behavioural repertoire. Where processes imbuing robots with levels of autonomy, in the McFarland sense, are hierarchically (e.g., evolutionarily) grounded, heteronomous constraints to complex adaptive behaviour may be minimized by bidirectional regulatory means. Future incorporation of Ashby's ultrastable concept satisfying complex adaptive behaviour must consider that internal and behavioural changes relevant to (re-)establishing system level equilibrium require non-random self-configurable approaches.

Robert Lowe is an Assistant Professor in Cognitive Science and Cognitive Robotics at the University of Skövde where he works at the Interaction Lab. His principle areas of interest concern cognitive-affective architecture for robots, reinforcement learning models of emotion and evolutionary robotics approaches to the study of adaptive behaviour.

Received: 11 October 2013
Accepted: 18 October 2013

---

[1] Di Paolo (2005) provides a more stringent notion of "adaptive" than is intended in this commentary. An adaptive process here simply refers to the robot's ability to regulate the essential variables that constrain the robot's sensorimotor capabilities.

Ashby's Passive Contingent Machines Are not Alive: Living Beings Are Actively Goal-directed

Tom Froese
Universidad Nacional Autónoma de México
t.froese/at/gmail.com

> Upshot • Franchi argues that Ashby's homeostat can be usefully understood as a thought experiment to explore the theory that life is fundamentally heteronomous. While I share Franchi's interpretation, I disagree that this theory of life is a promising alternative that is at odds with most of the Western philosophical tradition. On the contrary, heteronomy lies at the very core of computationalism, and this is precisely what explains its persistent failure to construct life-like agents.

Introduction

« 1 » Stefano Franchi's article is focused on a neglected aspect of Ross Ashby's theoretical framework, which characterizes the phenomenon of life in terms of *passivity* and *contingency*. In other words, the primary condition of the organism is conceived of as an equilibrium of inactivity, achieved by random convergence. Franchi argues that this puts Ashby's cybernetics in tension with contemporary trends in philosophy of biology, which emphasize intrinsic teleology, autonomy, agency, and enactive perception (e.g., Weber & Varela 2002; Thompson 2007). Given Ashby's explicit influence on these trends (e.g., Di Paolo 2003; Ikegami & Suzuki 2008; Froese 2009), his theory of life and this implicit tension deserve a closer analysis.

« 2 » Although I concur with Franchi's interpretation, I strongly disagree with his conclusions about how we should resolve the tension he has identified. Whereas Franchi wants a fuller development of Ashby's theory, I see an opportunity for exploring genuinely alternative possibilities. Briefly, aspects of Ashby's passive contingent theory of life have already been implicitly tested in computationalism (passive, non-contingent machines) and dynamical systems approaches (passive, contingent machines). For example, one passive non-contingent mechanism that is currently receiving a lot of high-profile interest is predictive coding based on Bayesian inferences. Perhaps expectedly, following William Grey Walter's joke about the homeostat's "sleepiness" (§10), this approach is faced by a so-called "dark room" problem, i.e., why an agent should be motivated to do anything at all if it can just shut itself away (Clark 2013). Similar problems of overcoming passivity are also encountered by approaches that follow Ashby's approach more closely, such as mobile robot designs that replace non-contingent cognitivist architectures with embodiment and situatedness (Dreyfus 2007).

« 3 » I think that any theory that sees the phenomenon of life as essentially a quest for eternal stasis, be it in a contingent or non-contingent manner, is misguided. At best such a theory only accounts for pathological behavior (Froese & Ikegami 2013). Organisms are intrinsically active and their behavior is non-contingent. The real challenge therefore lies in the development of a formal framework that can do justice to a conception of life as a non-equilibrium, self-producing and self-transforming phenomenon that is guided by its own emergent goals (Froese et al. in press).

A tension between Ashby and enaction

« 4 » Franchi states that, "the Ashbian organism will always be trying to accommodate itself to its environment by whatever means necessary: it is essentially a *passive* machine whose activity is a by-product of its search for non-action" (§10). I have reached a similar interpretation of Ashby's ideas in their scientific and historical context (Froese 2010). Moreover, I have also found that Ashby's theory creates an implicit tension in the writings of later proponents of systems biology. For example, his general systems theory significantly influenced the development of the theory of autopoiesis by Humberto Maturana and Francisco Varela (1980), but since Ashby had in fact viewed life as essentially *passive*, his influence undermined Maturana and Varela's intended aim to provide a systematic account for the self-asserting autonomy of the living (Froese & Stewart 2010).

« 5 » It is therefore no surprise that Maturana (2011) has distanced himself from my Ashbyan interpretation of his work, although he still fails to overcome properly Ashby's implicit influences. By rejecting Varela's eventual turn toward a Kantian interpretation of autopoiesis (Weber & Varela 2002), Maturana stays committed to a theory of life as passive-contingent. Indeed, classical autopoietic theory can be interpreted as a well-developed instance of what Franchi has called Ashby's generalized homeostasis thesis (Froese & Stewart 2010). The growing tension between Maturana's autopoietic theory and ongoing developments in enactive cognitive science (Villalobos 2013) can therefore be usefully understood as an echo of the tension originally provoked by Ashby's work.

« 6 » Nevertheless, despite these frictions, I believe that the enactive approach can provide crucial help in defending Maturana's guiding intuition that there is a qualitative difference that distinguishes the living from the non-living (Froese & Stewart 2013), a difference that does not even exist from Ashby's point of view.

Practical failures, theoretical shortcomings

« 7 » Why is it implausible to think of life as essentially passive? We can derive some insights from the practical failure of attempts to engineer artificial agents on the basis of this principle. Already, Ashby's own failure to take the homeostat work further with his follow-up DAMS project could be taken as an indication of the theory's shortcomings (Pickering 2010). Subsequently, all of symbolic AI shared his view of life as a passive phenomenon (albeit *non-contingent* passivity, since goals are explicitly represented). Just like Ashby's homeostat, a computer only reacts to commands, either from the user or from software triggers, until it reaches yet another resting state. This fundamental heteronomy at the heart of symbolic AI can be used to explain its well-known practical failures to construct life-like artificial agents, and thus motivates an enactive approach to AI that places autonomy at the core of life and agency (Froese & Ziemke 2009).

« 8 » Proponents of the enactive approach agree that the homeostat's double feedback architecture made an important contribution, but at the same time they struggle to overcome its limitations (Ikegami & Suzuki 2008). Franchi briefly refers to evolutionary robotics models inspired by this ultrastability mechanism, but he does not mention that further progress has been difficult. Ezequiel Di Paolo (2003) showed that implementing Ashby's mechanism is a significant step toward more organism-like robots, but the *contingent* link between (internal) homeostatic mechanisms and (external) behavior is a continuing source of problems.

« 9 » Switching the connections of an ultrastable system may in some cases lead to an adaptive re-organization of original behavior, as Franchi's model of two self-connected homeostat units shows, but it is far from guaranteed in more complex systems. It is possible that the switch is not sufficient to "break" the essential variables, thereby failing to provoke an adaptive response at all, or, if adaptation does occur, the new equilibrium may not reestablish the original behavior. From Ashby's perspective, this failure to recover the original behavior is not a problem, since he simply equates all equilibriums with survival – without asking how they might differ in terms of desirability (an idea that is later echoed by Maturana's claim that autopoiesis does not admit gradations).

« 10 » However, this neutrality is not found in nature. For example, inverted goggle experiments can be safely conducted without any irreversible consequences to the participants. How then is the appropriate reconfiguration achieved? There have been a couple of attempts at addressing the practical problems stemming from contingent dependencies with improved evolutionary robotics models, in particular by more closely tying the desired behavior to the satisfaction of the homeostatic conditions (Iizuka & Di Paolo 2008; Iizuka et al. 2013). However, these attempts have met with only partial success.

« 11 » We need to move beyond the traditional fixation on equilibrium dynamics in order to make real progress on these issues. According to Franchi, an Ashbyan organism is continuously required to behave so as to go to equilibrium, since "its always possible failure to do so will necessarily result in the homeostat's death" (§10). However, at the level of its physical body, an organism is always far-from-equilibrium with respect to its environment. Falling into an equilibrium is the same as dying, because the organism would lose its ability to do the work of self-producing its own material identity, i.e., the very process that ensures that the double feedback loop between behavior and internal homeostasis is intrinsically connected within a whole. Only non-living matter can be in physical equilibrium with its environment. Franchi conflates the living with the non-living in another way when he notes that the homeostat "will continue acting the way it normally does until an outside force compels it to change course of action" (§26). This is precisely the way in which Newton's first law of motion describes the behavior of objects. However, living beings – subjects – can change their behavior even in the absence of a change in external conditions.

Overcoming the autonomy/heteronomy dichotomy

« 12 » Franchi enticingly hints at the possibility of a "reevaluation of the traditional distinction between heteronomous and autonomous behavior" (Abstract). Is perhaps the *relational* view of life defended by the enactive approach (e.g., Di Paolo 2009) such a reevaluation? As I see it, there is a natural development of ideas from first-order cybernetics via second-order cybernetics toward enactive cognitive science. The early approach was too heteronomous, while the second-order approach was too autonomous, but the relational stance of the enactive approach formulates a dialectic between these two extremes (Froese 2011; Froese & Stewart 2013). Relatedly, it aims for a middle ground between the extreme of computationalism, which explicitly represents goals, and the extreme of Ashby's (and Maturana's) insistence on absolute non-contingency. Crucial for these two theoretical shifts is to go beyond traditional theories that portray life's ultimate purpose, no matter whether by active or passive means, as equilibrium, stability, or survival. As Hans Jonas recognized, only when we face up to the essential precariousness of living existence can we hope to understand the meaning of life:

> “the survival standard itself is inadequate for the evaluation of life. If mere assurance of permanence were the point that mattered, life should not have started out in the first place. It is essentially precarious and corruptible being, an adventure in mortality, and in no possible form as assured of enduring as an inorganic body can be. Not duration as such, but 'duration of what?' is the question.” (Jonas 2001: 106)

Tom Froese graduated with a D.Phil. in cognitive science from the Univ. of Sussex. He is currently a postdoctoral researcher at the National Autonomous University of Mexico. He investigates the relationship between life, mind and sociality using a variety of methods, including evolutionary robotics, systems theory and phenomenology.

Received: 2 October 2013
Accepted: 18 October 2013

• • • • • • • • • • • • • • • • • • • •

For Biological Systems, Maintaining Essential Variables Within Viability Limits Is Not Passive

Matthew Egbert
University of Hertfordshire, UK
mde/at/matthewegbert.com

> Upshot •The target article proposes that Ashby's investigations of the homeostat and ultrastability lead to a view of living systems as heteronomous, passive "sleeping" machines and thus are in fundamental conflict with concepts of autonomy developed by Jonas, Varela and others. I disagree, arguing that (1) the maintenance of essential variables within viability limits is not a passive process for living systems and (2) the purpose of Ashby's investigations of the homeostat was to investigate adaptivity, a subject that is related to, but clearly distinct from, autonomy. As such, I find Ashby's work on adaptivity to be neither in opposition to nor in direct support of modern concepts of biological autonomy and suggest that a productive way forward involves the investigation of the intersection between these two fundamental properties of living systems.

« 1 » Stefano Franchi distinguishes between *models of objects*, which are used to evaluate the accuracy of our understanding of real-world systems, and *models of concepts,* which are used to refine, explore or explain theoretical constructs. He observes, rightly, that the two approaches are not incompatible and can in fact be complementary, provided that they are not conflated.[2] I agree with Franchi that the homeostat is most productive when it is considered as a model of a concept rather than a model of the brain. However, I may disagree with Franchi over what concept the homeostat is intended to be a model *of.*

« 2 » I see the homeostat as a system for investigating adaptivity. "How does the brain produce adaptive behaviour?" is the opening sentence of Ross Ashby's *Design for a Brain* (1960), and, in my view, the goal of the book. Franchi appears to have a different interpretation, seeing the homeostat as targeting a broad view of life: "Ashby's device is a proxy for a view of life as generalized homeostasis, whose principles it embodies" (§11). I will argue below that if the homeostat is seen as targeting "just" adaptivity, and not a broader concept of life in general, a more consistent view emerges of Ashby's research, and of the subsequent research that integrates Ashby's work with concepts of biological autonomy.

« 3 » Franchi's argument that the homeostat is passive and heteronomous begins with William Grey Walter's comment that the homeostat is a sleeping machine "because its ideal function is to go back to sleep as quickly as possible" (§10). Franchi observes that this does not take into account the complexity of the behaviour that the homeostat can generate but, other than that, he seems to agree:

> "the main source of resistance to Ashby's thesis about life lies […] in its effort to derive action from non-action [and this] means something philosophically far more disturbing: the homeostat will act as much as is needed in order to *go back to sleep*. It will go to extraordinary lengths, it will take whatever action it can – just in order to go back to rest (i.e., to equilibrium)." (§10, emphasis in original)

« 4 » Here, Franchi is suggesting that because the homeostat "derives action from non-action," it is passive; but there is some confusing language used here that I believe leads to an incorrect view of the homeostat as passive. In particular, there is a conflation with the idea of (1) a variable being at a steady state (a concept that is more general than that of equilibrium and that allows for stability, i.e., an absence of change, at conditions that are far away from equilibrium) and (2) the system being inactive, passive, or "asleep." The homeostat is capable of the former, but is described in the quote above as doing the latter. It is not the same thing to say:

A: "The homeostat will act as much as is needed in order to go back to sleep." (Franchi §10)

B: The homeostat will act as much as is needed in order for an essential variable to remain (roughly) the same. (My interpretation of Ashby)

« 5 » In *Design for a Brain*, Ashby clearly presents a system that performs as described in (B), but this system is being described by Franchi according to (A). The error in (A) is apparent when we recognise that the maintenance of an essential variable within viability limits can require substantial dynamic activity, or as Ashby puts it: "the constancy of some variables may involve the vigorous activity of others" (Ashby 1960: 67). We can consider, as a metaphor, the act of balancing a vertical pole on your finger tips – the pole remains roughly vertical, but it would be a mistake to describe the system balancing the pole as passive or asleep – constant effort is needed to counteract the system's inherent tendency *not* to remain at that state.

« 6 » Nowhere is the necessity of activity for the maintenance of essential variables more evident than in living systems. All organisms are far-from-equilibrium dissipative structures that require ongoing acquisition of energetic and material resources to counteract their degradation (Gánti 2003; Schrödinger 1944). If we think about archetypal essential variables, such as body temperature or the levels of sugar and oxygen in the blood, it is clear that these all require ongoing organismic *activity* if they are to be maintained within viability limits. To assume the maintenance of these within viability limits as being passive is a mistake.

« 7 » This leads to why I disagree with the proposition in §§13–14 that Ashby's work opposes the relatively recent frameworks developed for studying autonomy. Ashby developed a framework for defining, describing and studying adaptive behaviour. Hans Jonas, and later Francisco Varela and Humberto Maturana, developed conceptual frameworks for understanding how a system can be autonomous, how its own needs can emerge and how these can be satisfied by the system itself. There is overlap between these frameworks: for instance, essential variables play central roles in both frameworks; however, as observed by Ezequiel Di Paolo (2005), the concept of autopoiesis does not automatically entail adaptivity. Similarly, the concept of adaptivity developed by Ashby does not entail autonomy. Put another way, it is possible to conceive of an autopoietic or autonomous system that is adaptive, or one that is not adaptive; and conversely, it is possible to conceive of an adaptive system that is autopoietic or autonomous, or one that is not. The two properties are orthogonal despite their overlap.[3] In fact, the homeostat is an example of a system that is adaptive, but not autonomous. The "essential variables" of the homeostat (the state of its magnets) are not actually essential to its existence or operation. They are only labelled as essential

---

[2] As a side note, I do not entirely agree with Franchi's comment that "sometimes, seemingly technical computationally-intensive work in the cognitive sciences is (also) philosophy in disguise […] presented in a technical garb more suited to a positivistic-minded age that shies away from abstract conceptual frameworks unless they are presented under scientific-looking covers" (§4). I see *computational philosophy* in a less duplicitous light, where computational models do not hide philosophy, but are rather provide valuable new methods for philosophers to explain, criticise and investigate the implications of their own frameworks and those of others.

[3] Here, I am arguing that the concepts of adaptivity and autonomy should be kept as distinct. However, it may be that in practice, artificial and/or natural organisms must have some degree of adaptivity (or minimal dynamical robustness to a variable environment) if they are to persist, in which case, all real-world instances of autonomous systems would also be adaptive systems.

because Ashby simply assumes the existence of essential variables, and then investigates how they can be regulated, without delving into the details of the intrinsic dynamics of the essential variables (for example the tendency of blood sugar levels, in the absence of behaviour, to approach non-viable states). When a magnet of the homeostat leaves the predefined viability limits, the system continues to operate. For this reason, the homeostat itself is actually heteronomous, but because the homeostat is a model of an adaptive system and not a model of an autonomous or living system, the heteronomy of the homeostat says nothing about the heteronomy or autonomy of living systems.

« 8 » For Ashby's investigations, it sufficed to consider the essential variables in only very abstract terms – a variable that must be maintained within limits; nothing more. Because the essential variables were not included in detail, the homeostat is not particularly effective at demonstrating the dynamic nature of the maintenance of essential variables. It was therefore possible for the homeostat to be mistaken for a sleeping, passive machine that does everything it can "to do nothing." If essential variables were modelled in more detail and the intrinsic dynamics of the essential variables of dissipative structures such as life were included, it would have been more obvious that Ashbian adaptive regulation, whenever employed by a biological system, must be anything but passive.

« 9 » For reasons such as these, there needs to be more work modelling the homeostat and its interesting form of adaptation. In the latter section of the target article, Franchi pointed out that Ashby's "environments," when simulated as homeostat units, included an inappropriate, or at least odd, property of self-regulation. This is an important observation that raises questions that can be investigated using models such as that presented in the target article. Similarly, I have argued here that there are assumptions implicit in Ashby's work concerning the nature of essential variables that need to be made explicit and investigated. To understand how adaptive behaviour relates to autonomy and agency, we need to develop a more sophisticated understanding of essential variables, their intrinsic dynamics, the emergence of viability limits and how mechanisms of adaptivity can respond to essential variables to prevent catastrophic system failure. Some work in this area is already underway (Barandiaran & Egbert 2013; Egbert 2013; Egbert, Barandiaran & Di Paolo 2010; Egbert, Di Paolo & Barandiaran 2009), and further developments will not only help us to understand how life differs from non-life, but also how life could have originated (Ruiz-Mirazo, Pereto & Moreno 2004).

Acknowledgements

I would like to thank Kepa Ruiz-Mirazo for our discussion related to this commentary.

Matthew Egbert: "A prominent theme in my research is the recursive self-maintaining organisation of life, and how this relates to its profound adaptability. This draws me toward the investigation of synthetic protocells and the origin of life, where mechanisms of behaviour are at their simplest, and towards minimally cognitive robotics and simulated agents, where fundamental concepts underlying adaptive behaviour and autonomy can be clearly defined and examined. As a Research Associate at the University of Hertfordshire, I am currently investigating self-maintaining behaviours in autonomous robots and simulated agents." Homepage at http://www.matthewegbert.com

Received: 14 October 2013
Accepted: 18 October 2013

Interpreting Ashby – But which One?

David Vernon
University of Skövde, Sweden
david.vernon/at/his.se

> Upshot • The association of heteronomy with Ashby's work in the target article follows from a direct interpretation of the second edition of Ashby's book Design for a Brain. However, the first edition allows for an alternative – opposite – interpretation that is compatible with autonomy and autopoiesis. Furthermore, a more balanced perspective is suggested to avoid unintentionally giving the casual reader a misleading impression that the homeostat is Ashby's ultimate position on homeostasis and that it is an adequate model of the brain.

« 1 » The target article claims that Ross Ashby's generalized homeostasis thesis entails that living organisms are heteronomous rather than autonomous, being controlled by the environment rather than independently adapting to environmental perturbations. In the following, I will explain why such a conclusion is consistent with the second edition of Ashby's book (1960). However, I will also argue that an interpretation of the first edition (Ashby 1954) can lead one to the opposite conclusion, one that supports autonomy rather than heteronomy and one that is compatible with the principles of autopoiesis. I also wish to highlight the priority of multistability over ultrastability, and the associated limitations of the homeostat and simple ultrastability (which Ashby himself acknowledges, albeit in a more obvious manner in the first edition.).

« 2 » In the following, when referring to work in *Design for a Brain,* first edition (Ashby 1954), I will adhere to Ashby's convention of using, e.g., S. 3/9 to refer to Chapter 3, Section 9. I will add a leading superscript (^{1}S. 3/9 vs. ^{2}S. 3/9) to differentiate between the first and second editions, respectively. I have also retained Ashby's use of italics; words originally in bold face type are underlined.

Heteronomy vs. autonomy

« 3 » In the abstract, Stefano Franchi states that Ashby's thesis "entails that life is fundamentally 'heteronomous.'" While Ashby does not use the term "heteronomy" in either of his books, this conclusion follows naturally from Ashby's development of generalized homeostasis in the second edition (^{2}S. 5/6), according to which an organism and its environment form a single state-dependent system (^{2}S. 3/9, ^{2}S. 3/10), the variables of which include a set of essential variables, the value of which much be kept within certain bounds if the organism is to survive (^{2}S. 3/14).

« 4 » Homeostasis, the process of regulating the essential variables, requires the organism to adapt to its environment to achieve stability, i.e., to keep the essential variables within physiological limits (^{2}S. 5/3). A region in the system's field (i.e., the phase-space containing all the lines of behaviour) is stable "if the lines of behaviour from all points in the region stay within the region" ([2]S. 4/8). Thus "'*adaptive' behaviour is equivalent to the behaviour of a stable system, the region of stability being the region of phase-space in which all the essential variables lie within their normal limits*" ([2]S. 5/8).

« 5 » Ashby refers to the organism part of the system as the reacting part ([2]S. 4/8). This is significant and reflects the passive characteristic mentioned in the target article (§10).

« 6 » The organism and the environment interact in two different ways, through two feedback loops, referred to as a double feedback system (§12). One is through the usual sensory and motor channels ([2]S. 7/2). The second is the significant one in the context of heteronomy and requires some explanation. It is a second-order feedback loop (Figure 2) and comprises a chain of causal influence from environment to organism via the essential variables and a set of parameters *S*.

« 7 » Ashby states that the essential variables are "immediately affected by the environment only" ([2]S. 7/3). These essential variables do not affect the organism directly but do so via a set of parameters *S* that, by definition, are not variables in the organism but control the configuration of the field of the organism, and hence its behaviour. The parameters *S* are themselves affected directly by the essential variables. Thus, the second feedback loop works as follows: the environment changes the values of the essential variables, which in turn affect the parameters that cause a change in the field of the organism and, hence, a change in behaviour. This change leads, through a process of ultrastability, to the line of behaviour encountering a region in the field that is stable (i.e., that returns the values of the essential variables to the required bounds).

« 8 » The important aspect of this scheme is that the environment causally influences the behaviour of the organism, which adapts to re-establish the equilibrium. This is, as Franchi suggests, a quintessentially heteronomous process.

« 9 » In the abstract, Franchi states that Ashby's thesis of homeostatic adaptation "is conceptually at odds with the autonomy-autopoiesis framework." However, Ashby's first edition of *Design for a Brain* can be interpreted in a manner that is compatible with autopoiesis and Francisco Varela's definition of autonomy, conceptually and operationally. There are two aspects to this: the autonomy and self-construction of autopoiesis, and the interaction in which an autopoietic organism engages with its environment.

« 10 » Note that the first edition does not include the double-feedback mechanism: the organism and the environment interact through sensory and motor coupling alone. No explicit control of the organism by the environment is suggested, as it is in the second edition, and the organism adapts to the environment as a single *absolute* system. The qualification of being absolute – that nothing else impacts on the system – is dropped in the second edition (possibly to allow for the parameters *S* to affect the organism).

« 11 » Varela defines autonomy as follows: "Autonomous systems are mechanistic (dynamic) systems defined as a unity by their organization" (Varela 1979: 55). This is a form of biological autonomy – *constitutive autonomy* – that refers to the internal organizational characteristics of an autonomous system rather than the external behavioural aspects (Froese, Virgo & Izquierdo 2007; Froese & Ziemke 2009). In turn, constitutive autonomy is closely related to *organizational closure*, a generalization of autopoiesis, a form of self-producing self-organization. Autopoiesis implies "the subordination of all change in the autopoietic system to the maintenance of the autopoietic organization" (Maturana & Varela 1980: 97) and autonomy in general is "the condition of subordinating all changes to the maintenance of the organization" (Maturana & Varela 1980: 135).

« 12 » Autopoietic systems are autonomous, but they are structurally coupled with their environments (Maturana & Varela 1980; Maturana & Varela 1987) in a process of mutual perturbation between the organism (the autopoietic agent) and the environment. Consequently, structural coupling allows the agent and its environment to adapt to each other in a mutually compatible manner, a process referred to as *co-determination*. Thus, structural coupling is a matter of mutual interactivity (Riegler 2002) and the adaptation of the organism over its lifetime as it improves in this structural coupling constitutes the organism's ontogenetic development.

« 13 » Now, compare this characterization of autopoiesis and structural coupling with Ashby's treatment of adaptation in the first edition, bearing in mind that, for Ashby, adaptation has the very specific role of ensuring the survival of the organism. It will be clear that the two perspectives are compatible.

> "Every species has a number of variables which are closely related to survival and which are closely linked dynamically so that marked changes in any one leads sooner or later to marked changes in the others. … These … will be referred to as the essential variables." ([1]S. 3/14)

> "For survival, the essential variables must stay within some definite region in the system's phase-space. It follows therefore that unless the environment is wholly inactive, stability is *necessary* for survival." ([1]S. 5/9).

> "*A form of behaviour is adaptive if it maintains the essential variables … within physiological limits.*" ([1]S. 5/3).

> "A determinate 'machine' changes from a form that produces chaotic, unadapted behaviour to a form in which the parts are so co-ordinated that the whole is stable, acting to maintain certain variables within certain limits. … [This] involves the concept of a machine changing its internal organization." ([1]S. 5/16).

« 14 » It is apparent from this that in the first edition, the organism part of the organism–environment absolute system can be construed to be organizationally closed and, furthermore, engaged in a process of continual adaptation – homeostasis – to secure its survival as a living entity. The organism is not controlled by the environment (it is constitutively autonomous) but it is structurally coupled to it and it is structurally determined by it through this process of adaptation.

« 15 » It is also worth noting here that Varela himself explicitly included homeostasis in his definition of autopoiesis: "an autopoietic machine is a homeostatic (or rather a relations-static) system that has its own organization (defining network of relations) as the fundamental invariant" (Varela 1979: 13). Varela's qualification of homeostasis to refer to the relations that define the system is something that Ashby anticipated, as exemplified above in his reference to a machine changing its internal organization and also in the opening statements of this book:

> “… how can we specify the ‘correct’ properties for each part if the correctness depends not on the behaviour of each part but on its relations to the other parts? Our problem is to get the parts properly co-ordinated. The brain does this automatically. What sort of machine can be *self*-co-ordinating?” (¹S. 1/8)

Again, this reinforces the compatibility of Ashby’s position with that of Varela and Maturana on autopoietic self-organization.

« 16 » I leave the last word on this point to Ashby: “Adaptation demands independence as well as interaction” (¹S. 11/8).

Ultrastabilty vs. multistability

« 17 » The target article focuses mainly on ultrastability and on the homeostat as Ashby’s particular instantiation of ultrastability. However, both the homeostat and ultrastability have a number of limitations, which is why Ashby spends much of both editions of his book developing the more powerful concept of multistability (which utilizes ultrastability). The importance of this cannot be overstated.

« 18 » Ashby says “The simple ultrastable system, as represented by, say, the homeostat, is by no means infallible in its attempts at adaptation” (¹S. 11/1) and later clearly states:

> “the thesis that the nervous system is approximately multistable … and we ask to what extent the thesis can explain not only elementary adaptation of the type considered earlier but also the more complex adaptations of higher animals, found earlier to be beyond the power of a simple system like the homeostat.” (¹S. 17/3)

« 19 » Contrast this with the statement in the target article §5:

> “Ashby’s major work, *Design for a Brain* (1952), contains a detailed description of an electro-mechanical device, the *homeostat*, which, he claims, can concretely demonstrate some essential features of the nervous system.”

This statement is potentially misleading as a casual reader may get the incorrect impression that Ashby proposed the homeostat as a viable model of the behaviour of the nervous system.

« 20 » Furthermore, the reader may be misled by the manner in which the target article leaves these issues hanging in §7 where it summarizes a “further set of claims.” These claims conflate one of three shortcomings of a homeostat (as an instantiation of a simple ultrastable machine) with the solution Ashby arrives at in dealing with this and other shortcomings. In fact, in the first edition, Ashby identifies six shortcomings of the homeostat, three of which are consistent with biological capabilities: inability to take corrective action, inability to adapt to an environment with sudden discontinuities, and dependence on a suitable period of delay between each trial (¹S. 11/1), and three of which he stated were not “features in which the simple ultrastable system, as represented by the homeostat, differs markedly form the brain of the living organism” (¹S. 11/2). These are: an inability to adapt gradually (¹S. 11/2), the inability to conserve a previous adaptation (¹S. 11/3), and excessive time required to adapt (¹S. 11/6); this last shortcoming is the one identified in §7, point number 5. The second edition does not explicitly highlight these second three shortcomings; Ashby refers to them as “inadequacies” (²S. 8/12). Nonetheless, having introduced the homeostat and ultrastability, the remainder of both editions then builds on ultrastability to arrive at a more sophisticated model of the behaviour of the nervous systems: multistability.

« 21 » Multistability is defined in the first edition as follows.

> “A multistable system consists of many ultrastable systems joined main variable to main variable, all the main variables being part-functions.” (¹S. 16/1).

However, he describes it in a subtly but significantly different way in the second edition, defining it as a collection of ultrastable systems (²S. 16/6), each of which is adaptive through second-order feedbacks to a polystable environment. Heteronomy clearly applies. Ashby states:

> “such a system is essentially similar to the multistable system defined in the first edition.” (²S. 16/6).

Significantly, he adds:

> “The system defined there allowed more freedom in the connexions between main variables, e.g., from reacting part to reacting part, and between reacting part and an environmental subsystem other than that chiefly joined to it; these minor variations are a nuisance and of little importance.” (²S. 16/6).

However, they may not be minor. By redefining multistability in this way, Ashby has effectively sacrificed the possibility of autonomous homeostasis through self-organization in favour of heteronomous homeostasis through second-order feedback from the environment. As we saw in §13 (¹S. 5/16) and §15 (¹S. 1/8) above, Ashby apparently anticipated the former possibility in the first edition of *Design for a Brain*.

Conclusion

« 22 » As stated in the abstract, the target article aims to trigger “a philosophical and technical reevaluation of the traditional distinction between heteronomous and autonomous behavior.” Such a reevaluation need not lead to the conclusion that homeostasis entails that life is heteronomous. On the contrary, one can also conclude that homeostasis is an essential aspect of autonomous behaviour, depending on how you interpret Ashby and which edition of his book *Design for a Brain* you take as the basis for your argument. If you take the first edition, a natural conclusion would be to see Ashby’s work as supporting autonomy, with the autonomous organism adapting in the context of the absolute organism-environment system.

David Vernon is a Professor of Informatics at the University of Skövde, Sweden. He works on cognitive systems and is particularly interested in modeling autonomy. Over the past 34 years, he has held positions at Westinghouse Electric, Trinity College Dublin, the European Commission, the National University of Ireland, Maynooth, Science Foundation Ireland, Khalifa University, UAE, and the Technical University of Munich.

Received: 19 October 2013
Accepted: 28 October 2013

The Self-moving Oil Droplet as a Homeostat

Takashi Ikegami
University of Tokyo, Japan
ikeg/at/sacral.c.u-tokyo.ac.jp

> Upshot • Using the example of chemical oil droplets, the paper discusses the idea of a homeostat in terms of a default mode network.

« 1 » Major approaches in the study of robotics in complex systems and artificial life use autonomous (Brooks 1991) and evolutionary robotics (Nolfi & Floreano 2000). By overlaying a hierarchy of simple reflexive processes, Rodney Brooks produced many "intelligent" autonomous robots. A representation of the environment is not predefined in the robots, but evolves through the robots' interaction with their environment; thus, "the world is its own best model."

« 2 » Evolutionary robotics is the development of intelligent robots through the "Darwinian" evolution in computers. Action patterns of the robot, which are coded as binary strings, become more intelligent in response to selections made and by mutation. Evolutionary robotics does not use a set of conditional statements, but, through a neuronal network of continuous neural states and over time, will produce intelligent behaviors.

« 3 » However, these approaches assume that a robot already has a basic architecture. The origin of robots, i.e., the origin of sensors, motors, body boundaries, and so on, has been studied by only a few. In this target article, re-evaluation of the idea of Ross Ashby's homeostat was to tackle the difficult problem of a robot's "being there." That is, a homeostatic state comes to exist before its functional form emerges. Based on my own studies in chemical oil droplets, I wonder when a mere chemical reaction becomes a life system. Neither Brooks' robots nor various evolutionary robotics can provide an answer to this question, but perhaps Ashby's homeostat machine does.

« 4 » Stefano Franchi's paper also challenges the concept of autonomy (§13). A robot is determined to be an autonomous robot when it masters all sensorimotor loops by itself. The concept of autonomy has been roughly defined as a system's spontaneous behavior independent of human control (Pfeifer & Scheier 1999). But the concept of the homeostat also tackles this issue and suggests considering heteronomy (originally introduced by Immanuel Kant) as well as autonomy. The concept of the homeostat can connect heteronomy and autonomy as a continuous spectrum (as discussed by the target article, §13), which I also think is an important point. I will discuss how heteronomy and autonomy are linked by referring to my own work on self-moving oil droplets.

« 5 » After pouring oleic anhydrous acid into a high pH aqueous solution, we discovered the emergence of "self-moving oil droplets" about several hundred micro meters in size (Hanczyc et al. 2007; Hanczyc & Ikegami 2010). An oil droplet senses the chemical gradient and avoids low pH (< 10), preferring high pH (> 11) regions. Such a self-moving droplet can be interpreted as the origin of sensorimotor coupling, so that it provides an example of a chemical robot. I claim that this self-moving oil droplet is a natural version of Ashby's homeostat. A thermal equilibrium chemical system tends to restore the same equilibrium state; this tendency is referred to as Le Chartlelier's principle. Yet in our example, the droplet is not in its thermal equilibrium state. Also, it tends to restore the self-moving state before long. What should be restored by a homeostat is a survival state, which is translated as "activity of the self-movement" in the oil droplet. Therefore, the homoestatic nature of a robot and the continuity of autonomy and heteromy are also clearly observed in the chemical system in different contexts.

« 6 » My concern after reading this article (and Ashby's original book), is the origin of "dice" (randomness) being used as the main engine by a homeostatic system. It is necessary to escape from a dangerous state and search for a different equilibrium state. However, I wonder whether the dice has any bias or not. I think the answer can be found in the idea of a default mode network (DMN) of a brain and embodied chaotic itinerancy (Ikegami 2007). The DMN are activity patterns of the brain when in the resting state. Across the global brain area, there exists about 0.1 Hz of noisy but synchronized BOLD signals found between the medial prefrontal cortex and posterior cingulate cortex subsystems. Yet DMN is actually the default mode of a brain. It consumes 20 times more energy than when in the non-DMN state. It seems that a default mode may exist to maintain memorization as well as reaction to and even predictions of environmental images (Raichle & Synder 2007). This notion of the default mode must be generalized for any living system, with or without brain systems, which I believe will result in a system with ultra-stability.

« 7 » In the case of the oil droplet experiment, when we insert a high pH region as the external perturbation, the droplet reacts to the pH and changes its behavior. We observed that a droplet is usually attracted to the higher pH region, but sometimes it avoids such a region. Actually, a droplet spontaneously changes its direction of movement in some irregular ways when there is no external pH gradient. This autonomous behavior can be explained by convection flow inside the droplet, the internal pH gradient and the on-going chemical reaction on the surface of the droplet. Based on this example, I would argue that DMN is a realization of ultra-stability and the oil droplet is an example of a homeostat. It is a complex dynamic resulting from chemical dynamics and convection flow. Analogous to the DMN in a brain system, we thus expect cognitive behavior in the chemical oil droplets, or in artificial systems installed with DMN (e.g., the "mind time machine" that I constructed with three visual screens, 15 video cameras and artificial neural nets, running in the open air for three months, Ikegami 2013).

« 8 » To conclude, I agree with Franchi that Ashby's idea should be re-discovered and extended. On the other hand, we are already in the phase of out-and-out construction of artificial systems (chemical and mechanical). I think conceptual insights come only after realizing a new artificial system and not the other way around.

Takashi Ikegami is a professor at the Department of General System Studies at the University of Tokyo. He attends the International Conference on Artificial Life very frequently, and gave the keynote address at the ALife conference in 2008. He is on the editorial boards of *Artificial Life*, *BioSystems* and *Adaptive Behaviors*.

Received: 16 October 2013
Accepted: 22 October 2013

Author's Response: Is the Homeostat a Passive Machine? Is Life a Passive Phenomenon?

Stefano Franchi

> Upshot • The target article suggested that Ashby's device, the homeostat, embodies and illustrates a conception of life as a passive-contingent phenomenon. It advocated renewed experiments with updated and extended versions of his device that would allow us to understand better what passive-contingent life "would be like." In assessing the proposal, we should be particularly careful when dealing with the concept of "passivity," and we should not mistake the proposed theoretical exploration for a substantial metaphysical thesis about life in general.

Introduction

« 1 » The target article drew a number of varied responses from the commentators. I have been positively surprised by the many different ways in which my main claims have been understood and I am glad to have a chance to articulate their content better.

« 2 » The different challenges the commentators raised can be seen as different views on the truth values of two basic claims that they take to be at the core of the target article. The claims are:

1 | *Ashby's homeostat* (through his reliance on generalized homeostasis) supports a view of life as passive, heteronomous, and contingent.

2 | *Life* is passive, heteronomous, and contingent.

« 3 » Most commentators see the target article as an argument in favor of both (1) and (2). Its conclusion would therefore be:

3 | We should adopt and extend Ross Ashby's generalized homeostasis and build better, updated, and more complex homeostats as models of passive, heteronomous, and contingent biological organisms.

« 4 » The challenges against this argument take (mostly) two forms. Some commentators (e.g., **Inman Harvey**, **Matthew Egbert**, **Takashi Ikegami**) deny that (1) is true. Ashby's homeostat, they contend, is not a passive machine. They also take (2) to be false, for life, they state, is fundamentally active and autonomous. Their conclusion is thus similar to mine, but for the opposite reasons: we should keep working in Ashby's tradition, and update and extend his intuition precisely because his work supports a view of life as a *non-passive* phenomenon.

« 5 » Other commentators (e.g., **Tom Froese** and, to a lesser extent, **Robert Lowe**) accept (1) as true: the homeostat is indeed a passive-contingent machine. However, they reject (2) and take life as essentially active and autonomous, thereby rejecting (3). Their conclusion is that we should abandon or at least seriously temper Ashby's theses when modeling biological life. Thus, the majority of commentators agree that life is active and autonomous and they only differ on whether Ashby's homeostat supports or undermines such claim. **David Vernon** supports both truth and falsity for claim (1), but he ascribes them to a substantial philosophical difference between the two editions of Ashby's major work. The first, 1952 edition would present a non-passive, autonomous homeostat, whereas the second, 1960 edition would indeed switch to a passive device.

« 6 » **Helge Malmgren**'s contribution is the only exception to the consensus. He agrees that the homeostat is a passive machine and even argues that Ashby viewed not only life in general but his own life as well as essentially passive. He then connects Ashby's view of passivity with the views of other non-canonical cognitive scientists (as we would say today): Gustav Theodor Fechner and Sigmund Freud.

« 7 » The main challenge and major source of disagreements between the commentators is then the truth value of (1) – "Is the homeostat a passive, heteronomous, and contingent machine?" – and only to a minor extent about the truth of (2) – "Is life a passive, heteronomous, and contingent phenomenon?" – where an almost unanimous consensus exists. In what follows, I will address these two issues in order, before addressing **Malmgren**'s point.

Is the homeostat a passive machine?

« 8 » The original claim I put forward in the target article was that the homeostat's relationship to its environment is characterized by three non-equivalent although interrelated dimensions: passivity, heteronomy, and contingency. The statement under discussion would thus have to be decomposed into three separate sub-claims supplemented by an additional discussion of their relationship. I cannot carry out a task of this magnitude here: each of the three concepts under discussion would require an extensive analysis drawing its evidence from current uses in everyday language, from the concepts' history in the philosophical literature, and from their current and past deployments in scientific research. However, the full three-fold treatment, while necessary in a full-blown discussion of the homeostat, is not required in the present context, since the commentators' attention is mainly focused on the first of the three dimensions: passivity. This is not surprising, as the association between passivity and life in ordinary as well as in technical language tends to draw immediate negative connotations. Passivity, in other words, always comes with an associated value-judgment about its bearer. Unsurprisingly, the commentators who accept the homeostat's passivity (**Froese**, **Lowe**, **Vernon$_2$**) immediately cast doubts on its usefulness, while those who defend Ashby's view reject passivity altogether (**Harvey**, **Egbert**, **Ikegami**, **Vernon$_1$**). Who is right, then? Is the homeostat a passive device?

« 9 » The answer would be simple if we had a straightforward and widely-shared understanding of passivity. This is far from being the case. There is a somewhat minimal conception of passivity, which is best expressed by the grammatical passive voice. A subject-verb-direct object (English) construction – "I kicked the rock" – can always be turned into its corresponding passive voice version – "The rock was kicked by me" – because the verb-direct object pair implies the existence of an action whose effects are transferred upon the object. "Minimal" passivity would thus be equivalent to being on the receiving end of someone or something else's action. Call this notion p_1. The problem is that p_1 falls short of capturing the meaning(s) of our everyday uses of the term "passive." When we identify a person as being "passive," we usually mean more. Typically, we intend to convey a more complex meaning that involves at least the following components: (p_1) the person is on the receiving end of some action or event; (p_2) the person is capable of being affected by that action, in the sense that some important aspects of her being or behavior will be modified as a consequence; (p_3) those changes will be detrimental to the person's well-being and change her situation for the worse. Consider, for instance, the common expression "X is passively accepting her fate." To a greater or lesser extent, all aspects of passivity, from p_1 to p_3, must be present to make sense of it: some external event is acting upon the person (p_1), who is consequently and substantially affected by it (p_2), while she would be better off otherwise (p_3). Thus passivity implies at least a "capacity to be affected" or a capacity to suffer, often against the sufferer's optimal welfare state. This is why, while we could apply the minimal conception of passivity (p_1) when speaking of inanimate objects – as I did with my first example – we seldom speak of rocks and sticks as being "passive." We are straining the English language if we say that "The ball was passively kicked around." We are not when we say that the "The patient was passively following orders." In short, passivity requires more than being on the receiving end of someone else's action. It may even require, as in the previous examples, the capacity to carry out actions that are instigated by an external event (or a series of commands, in this case).

« 10 » There is more. Sometimes, we apply the term to inanimate objects, especially in technical vocabularies. For instance, we often speak of "passive receptors" or of "passive sensors." Here again, we mean something more than being merely on the receiving end: a standard light sensor is a passive receptor because it is capable of being affected by the photons impinging on its surface while not expending any energy. This is consistent with one of the possible origins of the word "passivity," which is usually traced back to the Latin form *passiuitas*, a word possibly linked to the verb *pati*, "to suffer," "to be affected," and possibly a cognate of ancient Greek *paskhein*, which had a similar meaning. In short, passivity is essentially related to the capacity to be affected and be pushed to act – often against the patient's best interests. Let us call this semantic aspect of passivity p_4 and provisionally conclude that a truly minimal conception of passivity capable of accounting for the most common uses of the term in ordinary language, and especially when referring to human beings, includes all four mentioned aspects, p_1–p_4.

« 11 » But passivity is more complex than that. Consider the common expressions "passive resistance" and "passive-aggressive." Even this revised minimal concept of passivity cannot provide a satisfactory account of these occurrences. The semantic content of "passive resistance" is actually much closer to "impassive" than to "passive": a person actively engaged in an act of passive resistance would achieve ultimate success by being truly impassive to the solicitation and teasing exerted on her from the outside: she would be "unaffectable." Call this new aspect of passivity p_5. Similarly, the passive-aggressive patient is unwilling to accept those modifications to her being or to her behavior that would normally be produced by external actions (I am referring here to the actual pathological state, whose description we can find, for instance, in the official diagnostic manuals of the American Psychiatric Association 2000). By closing up the normal receptivity to outside influence, she avoids following orders. She is apathetic, in the technical Stoic sense of the word: unable to be affected. Passive-aggressiveness may be a new semantic trait of passivity – call it p_6 – or perhaps it is an exaggerated manifestation of the previous component. It may even be an improperly named concept, in spite of some vague relationship to the "ultra-minimal" p_1 conception of passivity.

« 12 » Finally, let me add one more etymological piece of evidence to this brief sketch of the complex semantic landscape engulfing a seemingly ordinary concept. The first recorded uses of the Latin term *passiuitas* come from Tertullian, the early 3rd century Christian apologist. His use of the term is seemingly so different from our current linguistic habits that some scholars have been forced to posit the existence of two homonym words, only one of which survived. For Tertullian, *passiuitas* means "sexual promiscuity," "promiscuous intercourse," or, more generally "turbulence," "excessive opening to the outside."[4] Tertullian takes to an extreme the semantic trait p_2. He turns the subject's capacity to be affected into an essential and indeed excessive opening to the outside, both literally and, by extension, metaphysically. This is clearly the reading that allows Tertullian to use *passiuitas* in the sense of promiscuity, with explicit reference to sexual commerce. The more "passive" an organism is, the less "proper" it becomes ("at the mercy of everyone" [*pro arbitrio cuiusque*], in Tertullian's words). This is apparent in the sexual domain, but the expression should be intended in the most general sense, as indeed does Tertullian. When passivity comes to mean openness, its meaning shifts toward ex-propriation and loss of property (i.e., loss of what is "proper to" or "appropriate for" the subject both in the metaphysical and in the ethical sense). The passive recipient does not simply receive someone else's action (or worse, for Tertullian): it is so open to such reception that it enters into all kinds of improper, excessive, and ultimately immoral intercourse with the outside. The truly passive subject, in this reading, is essentially "promiscuous" and that means "*always ready* to being acted upon." This form of passivity, for which I will expend my last label, p_7, has some interesting characteristics. In the present context, the most important one is the explicit association between the excessive openness to the outside and the "turbulent" activity typical of "lust," to use Tertullian's terms. In his anti-pagan invective, he implies that the truly promiscuous subject is constantly seeking intercourse and therefore constantly engaged in *proximate* actions that make it ready for such. Their *ultimate* (rather, *distal*) goal, however, is to receive someone else, hence their passive character.

« 13 » There are even more aspects I should bring in to characterize passivity

---

[4] In the *Apologeticus*, Tertullian says: "Age iam recogitate, quam liceat erroribus ad incesta miscenda, suppeditante materias *passivitate luxuriae*," which could be translated as: "Now just reflect how many opportunities for errors toward incestuous mixings are made possible by the *passivity of your lust*" (1954: 104); similarly, in a metaphysical treatise, he asserts "Haec inquies non est, – haec *turbulentia et passivitas* non est; sed moderatio et modestia" (literally: "This is not disquiet, this is not *turbulence and passivity*, but rather moderation and modesty," 1954: 431, my emphases throughout).

properly, but I think this brief sketch should suffice to give an idea of the complexity of the concept while providing elements for an informed answer to the original question: "Is the homeostat a passive machine?" **Harvey** and **Egbert** argue that it is not, because Ashby's device is constantly engaged in some action in order to keep its relationship to the environment stable (i.e., to achieve homeostasis). **Harvey** offers a couple of examples: the (possibly mechanical and homeostatically driven) bicyclist who wobbles down the road and constantly changes the bicycle's relationship to the environment in order to keep its angle from the vertical constant; the starving 19th century Irish peasant who emigrates to America to escape the potato famine. **Egbert** adds a structurally similar third example: the continuous actions and subtle adjustments that are required to keep a pole vertically balanced on one's fingertips. In all cases, they claim, the agent is not passive, because there are actions being carried out (operating the handlebars, emigrating, moving one's hand and arm). These would be effective counterexamples only if we restricted passivity to its first, p_1 component. In that case, I would agree that the homeostatically controlled bicycle, the Irish farmer, and the pole-balancing hand are not passive. They are different from rocks being kicked around. As I argued, though, p_1 (i.e., "being acted upon") is a severely incomplete characterization of passivity. Though a necessary condition, it is far from being sufficient, to the point that we never use it for rocks and sticks, the only beings to which it could be applied. If we extend passivity to all four conditions p_1–p_4, therefore including the capacity to be affected and the related capacity to act in response to actions initiated from an external agency, the counterexamples are no longer such. In all three instances, the original source of action that prompted the recipient's adjustments is external to it: the original lack of equilibrium between the bicycle and the road, the potato's disease, the oscillation of the pole around its center of gravity. In all instances the original, external sources of action disturb the equilibrium between external agent and recipient and, due to the recipient's receptivity to that disturbance, force it to undertake restorative actions. The important point is that *the recipient's actions stop as soon as equilibrium is restored*. Such a complete stop is not always possible. It certainly is for the hypothetical Irish farmer envisioned by **Harvey**. We can reasonably speculate he will stop changing places of residence as soon as he can farm again in his new dwelling. Sometimes restorative actions on the recipient's part cannot stop because the external actions do not stop either. This is the case of the pole on the fingertips example, whose essentially unstable physical equilibrium compels the recipient to undertake a never-ending series of restorative counter-actions which *would* stop if the equilibrium *were* restored.

« 14 » My conclusion is that an adequate reading of passivity forces us to conclude that the behaviors of the cyclist, the farmer, and the hands are indeed passive, for the goal of the (internal) actions is to eliminate the effects of the (external*)* actions. This is a point that Hans Jonas had seen very well. In the *Imperative of Responsibility*, he provided a detailed comment of the cybernetics explanation of a cat's behavior as following a sequence of steps such as:

> "physiological stress (homeostatic 'gradient'); internal secretion and nerve stimulation; readiness for selective triggering of a behavioral pattern; external (sensory) nerve stimulus triggering the appropriate behavior pattern; sensorimotor feedback guiding the course of behavior; subject of action eliminating the initial stress (homeostatic equilibrium), this itself being the 'goal'." (Jonas 1984: 62)

« 15 » Jonas was keen to point out that such a "homeostatic goal" is really no goal at all and concludes that the desired good would just be the "*subjective representation* of the state of indifference or quiet that waits at the end" (ibid., emphasis added). This is why the cybernetic explanation of purposive behavior, according to him, really turns all animals into passive machines whose only real goal is the ultimate end, or death.[5] I think Jonas's characterization of homeostatic equilibrium as an explanatory principle is ultimately correct, although I do not share the conclusions about cybernetics he draws from it. The homeostat is a passive machine because its ultimate goal – but none of its proximate ones – is a lack of action. All the actions it takes are efforts, no matter how futile, to eliminate the externally originated action that prompted it to act in the first place. This is also why the target article recalled William Grey Walter's joking characterization of Ashby's device as a *machina sopora*. I certainly did not mean to say that the homeostat *is* always asleep as **Harvey**, and perhaps Grey, thought. That would be obviously false, for only a rock would achieve such permanent slumber. Yet, the homeostat's search for equilibrium implies that "sleep" (lack of action, or, at least, minimal action) *would be* the ultimate, although often unattainable goal.

« 16 » **Ikegami**, I think, makes a somewhat similar point to **Harvey**'s and **Egbert**'s. I admit I am not sure I fully understood the relationship he suggests exists between the behavior of oil droplets in an aqueous pH gradient and the homeostat on the one hand, and DMN networks and ultrastability on the other. I believe his overall argument is meant to extend his previous suggestion (for instance in Ikegami & Suzuki 2008) about the necessity for a living system to change its viability parameters dynamically in order to cope with its necessary, constant, and autonomous movement. In 2008, **Ikegami** called such a conservation of homeostasis under constantly varying conditions "homeodynamics." The concept introduced the idea that a living homeostatic organism must constantly move around in order "to distinguish likes from dislikes and prey from predators" (Ikegami & Suzuki 2008: 398). It thus needs a baseline level of activity that is concurrent with and related to its equilibrium-seeking homeostatic processes. The "default mode network (DMN)" in the brain translates this homeodynamic architecture to the nervous system/external environment setting. When **Ikegami** asserts, referring to neuroscientific research as well as to his own work, that a brain is constantly expending

---

[5] It may be worth noting that this is the same conclusion that was reached a few decades earlier by Sigmund Freud when, in *Beyond the Pleasure Principle*, he took the principle of pleasure – which is essentially a homeostatic principle – to its ultimate consequences and posits the existence of a "death drive" – the urge to return to an inanimate state (1964). Of course, Freud did not share Jonas's negative assessment of the theoretical validity of this unpalatable consequence.

energy in order to "maintain memorization as well as reaction to and even prediction of environmental images," he is providing a brain-level equivalent of the autonomous exploratory movement of the organism he had previously described under the rubric of homeodynamics. The overall point is thus that the living organism needs a baseline level of activity that may be an integral and under-appreciated part of a generalized homeostatic system or, perhaps, an additional and to all extents logically prior subsystem that the homeostatic component needs to work with. In either case, the argument would go, the alleged passivity of the homeostat fails to account for this underlying activity level and must be considered a reductive characterization.

« 17 » On the contrary, I believe that the possible existence of such a base level is very consistent with the description of passivity I gave above if we only remember that a recipient's openness to the outside (its "capacity to be affected" or p_2), and, even more so, its passive promiscuity – with the inherent search for a "commerce with the environment" (my p_7 component) – are important traits of passive behavior. Briefly put, the activity baseline can be considered an operational translation of a subject's capacity to be affected. The distinctive criterion that allows this baseline activity to be reclassified as a form of passivity is, once again, the distinction between the proximate and distal goals of the actions themselves. As Jonas stated (and, subsequently, Francisco Varela confirmed, but previously, albeit in different forms, Aristotle and Immanuel Kant had argued) the supreme goal of an organism's autonomous action is *itself*. It is the imperative to preserve its own bodily integrity against the constant challenges mounted by internal needs and external threats that pushes an organism to posit its own identity as an ever renewable form that needs to be kept constant and indeed posited as the ultimate goal directing every proximate action. On the contrary, the ultimate goal of the baseline activity **Ikegami** describes could not be more different from Jonas's: it is the *other*, or rather the readiness to interact with an external agent and to accept the effects of the actions that may originate in it. The contrast could not be starker, in my opinion.

On the simulation of concepts

« 18 » On the basis of the previous discussion, I think we can affirm that the homeostat is indeed a passive device, provided we accept a notion of passivity that does not reduce it to a mere lack of action. Such a notion is forced upon us as soon as we examine its use. That being said, I do not mean to imply that we have a perfectly consistent notion of passivity at our disposal, nor that the homeostat's behavior represents a perfect instantiation of this ideal concept. In fact I think the opposite is the case: a close analysis of current linguistic uses (of which the above discussion presents only the barest sketch) and, even more importantly, a close reading of the philosophical works in which passivity is discussed (starting with Aristotle's works and especially *Categories*, *On Generation and Corruption*, and *De Anima*) shows that the concept is anything but a unified and coherent semantic whole.

« 19 » This is precisely why Ashby's work is particularly interesting. As the previous pages will undoubtedly have shown, my background and research is located within philosophy rather than in the sciences. I originally became interested in Ashby's work when trying to articulate a workable conception of passivity that would make sense of the many uses of this concept we find in 20th century European philosophy (starting from Husserl's conception of "passive synthesis" (2001) and continuing pretty much to the present day). Ashby's general homeostasis thesis became a theoretical lens for examining the concept of passivity in general, and the experimental instantiation Ashby devised (the homeostat) a practical tool to investigate its properties. The target article's introduction of the notion of "concept simulation" is meant to provide a general characterization of this procedure. Some commentators disputed the validity of the notion.

« 20 » **Malmgren** points out its linguistic faults and proposes an alternative formulation, equally based on Ashby's work. I will grant him that my term may lack stylistic elegance, and I am open to suggestions on that front. I do not think **Malmgren**'s suggested revision is different from my original proposal, however. He would like to replace "concept simulation" with a three step process (that he attributes to Ashby):

1 | revision of current concepts which are
2 | illustrated through
3 | the simulation of specific physical processes.

« 21 » The sequence is very similar to what I proposed, even though my discussion went in the reverse order. I claimed that Ashby used (3) the homeostat (2) to illustrate (1) a new view of life. Reversing the clauses' order, we obtain that: (1) a new view of life (**Malmgren**'s "revised concepts") (2) was illustrated (3) by a physical apparatus (the homeostat). My contention is that "concept simulation" is an accurate though possibly inelegant designation for clauses (1)–(3). I would like to stress again that (1)–(3) designate a concept simulation because the physical (or electro-mechanical, or, nowadays computational) processes used in step (3) bear no physical similarities whatsoever to the objects whose behavior is being investigated. Ashby never tired of stressing that the homeostat is not a physical analog of the brain or of the organism. Indeed, when Ashby designates a subset of the homeostat's units as the "agent" and the remainder as the "environment," he is not making any assumptions about what the "agent" and the "environment" may stand for. They may stand for the cat approaching the fire, or they may stand for one subsystem of a monkey's brain interacting with the muscular system (to quote two of Ashby's favorite examples in *Design for a Brain*). They may even refer to different collections of neurons interacting with each other, a hypothesis Ashby explicitly entertains when he comes to the discussion of multi-stable systems. The setup is very different in standard simulations used in physical contexts. A Boeing 777 flight-simulator does represent a real plane moving in physical space, and it is used to simulate the behavior of an actual object subject to real physical constraints (gravity, friction, torque, mass, and so on) when a human subject interacts with it. Similarly, a computer simulation of weather patterns simulates the temporal evolution of real spatio-temporal chunks – the fluids in the atmosphere. In contrast, "Agent" and "Environment" designate subsystems that may be individuated differently depending on the chosen perspective and have no built-in analog with the physical world. What do they designate then? And what does the homeostat's behavior simulate? My claim is that they designate a "concept." To be more precise (and even less elegant), I could say that the homeostat simulates the behavior of a system working according to a five-thesis conceptual framework about life, which I called "generalized homeostasis" (borrowing Ashby's own label). "Generalized homeostasis" is not a physical object, even less a psychological one. We certainly cannot pick it up and look under it, as we can do with planes and clouds. In my view, the homeostat is a physical object simulating the behavior of a theoretical one. As I said in the target article, this is not all the homeostat is, because the distinction between concept- and object-simulation individuates the ends of a spectrum of possible experimental options and each specific simulating device used in actual scientific research will present aspects of both. In my opinion, though, the conceptual component has been historically overlooked by Ashby's scholars and followers.

« 22 » **Ikegami** offers a different objection. He observes that "conceptual insights come only after realizing a new artificial system and not the other way around" (§8). This sparse remark could be read in two different ways, each one implying a well-defined hierarchy among disciplines and both eliminating philosophy from the ranks of intellectual undertakings. First, the remark could mean that philosophical results ("conceptual insights") provide a conceptual articulation of what exists (or of what has been built). This seems to me a misguided interpretation that attributes to the philosopher the task of the scientist, whose job is precisely the production of a theoretical understanding of what there is (actually or ideally, in the case of purely formal disciplines). But perhaps **Ikegami** implied that "conceptual insights" refer precisely to the results produced by the theoretical scientist, in which case philosophy exits the picture altogether or is reduced to an epistemological reflection upon science-produced concepts. Even though this neo-Positivism inspired conception of the discipline never ceases to be popular, I think it is mistaken, at least as it applies to the concretely and historically determined existing philosophical practice. Philosophers have historically been engaged in the invention of concepts that allow us to think differently from the existing frameworks. In other words, philosophy – if I may be allowed a dogmatic tone that space limitations render inevitable – fabricates the concepts that will allow us to think of the structure, limitations, and consequences of the intellectual and social situations we happen to live in. I am well aware that I am offering this sweeping statement as a dogmatic pronouncement that is bound to encounter some resistance in my scientific-oriented audience. Yet this is not the place for a systematic treatment of such a large issue.[6] I can perhaps partially remedy this unsatisfactory situation by pointing out that the concept that the target article introduced in §16 was offered precisely as an example of a philosophical concept. The somewhat provocative question it asked – "What is it like to live a passive, fully homeostatic life?" – implied that such a life may not exist at all. That is why, in the language I used above, "passive life" is a "fabricated concept" that is meant to explore a viewpoint on life normally excluded from scientific conversation, as the unanimous rejection of the thesis by all the commentators (**Malmgren** excluded) confirms. In fact, the question entailed an even stronger thesis. Namely, that we will not be able to assess the truth value of a statement such as "life is essentially passive" – nor, by implication, its contradictory one: "life is essentially active" – until we understand what passive life means. With this clarification in mind, let us now move to a discussion of the thesis itself.

Is life a passive phenomenon? Is it an active one?

« 23 » Recall the argument from which I started this response:

1 | *Ashby's homeostat* supports a view of life as passive, heteronomous, and contingent.

2 | *Life* is passive, heteronomous, and contingent.

3 | We should adopt and extend Ashby's generalized homeostasis and build better, updated, and more complex homeostats as models of passive, heteronomous, and contingent biological organisms.

« 24 » It should now be clear that (1)–(3) is not, in fact, the argument the target article intended to present for discussion and I can only be grateful to the commentators for offering me the chance to restate my point, and hopefully make it clearer. The difference lies in claims (2) and (3), which I would switch and rephrase as follows:

2' Simulations carried out on updated and extended homeostats would allow us a better grasp on what "passive life" would actually be.

3' Once we have a better grasp of "passive life," we could assess whether life actually is passive or active.

« 25 » It follows that I do not think we are presently in a position to determine the truth value of the contradictory propositions "Life is passive" and "Life is active," because we are not in possession of a solid concept of passivity, as I argued in the section "Is the homeostat a passive machine?" of this response.

« 26 » Almost all commentators disagree and offer different evidence in support of life's intrinsic activity. Indeed, **Harvey** takes the truth of the life/activity identity as obviously self-evident, to the point that he thinks my purpose is to "castigate" Ashby by attributing to him a patently false interpretation of life as passive. **Harvey** equates passivity with "stasis or quiescence in the organism itself" and denies that either the homeostat or life in general exhibit any such property. I have already explained above why I think we should reject this view of passivity as inadequate to the phenomenon, and why, therefore, the indubitable existence of proximate actions in living organisms' everyday behavior does not constitute, in my view, a valid objection. I addressed above **Ikegami**'s objection as well, and I will not elaborate further. Similarly to **Ikegami**, **Froese** states (§3) that life is essentially active since "organisms are intrinsically active and their behavior is non-contingent." The evidence in support of this claim is twofold. On the one hand, the difficulties encountered by evolutionary robotic models closely following the passive-contingent Ashbian paradigm (e.g., Iizuka & Di Paolo 2008) may be read as a sign of deeper theoretical issues. I would agree with **Froese** that

---

[6] See Franchi (2005) for a discussion of the relationship between philosophy and science in the specific context of artificial intelligence and cognitive science.

the research he mentions has been only partially successful. However, as I have argued (briefly) in the target article as well as elsewhere (Franchi 2011b), evolutionary robotics models inspired by Ashby's work have stayed away from a full implementation of his generalized homeostatic thesis, let alone his more ambitious multi-stable systems. This does not mean that a fully Ashbian approach would necessarily overcome those difficulties, of course. But I do think it would be fair to consider the problem as an open empirical issue that only further work may help us understand better.

« 27 » On the other hand, **Froese** claims that this deeper issue has to do with the "traditional fixation on equilibrium dynamics" (§11) and suggests that living organisms are always far-from-equilibrium since an organism falling into equilibrium would lose its ability "to do the work of self-producing its own material identity," which would therefore be tantamount to dying. I am not convinced by this claim, and for two reasons. First, it seems to me that **Froese** confuses Ashbian equilibrium with its physical counterpart. Keeping essential variables at constant levels (e.g., blood's sugar concentration, to use Ashby's favorite example) will necessary imply a far-from-equilibrium exchange of matter and energy with the environment. I would agree that only non-living matter can be in physical equilibrium with the environment, but I fail to see why keeping internal variables constant would imply physical equilibrium. Second, and more importantly, I think stating that life must be essentially active because a passive organism (in the sense discussed above) would not be able to self-produce its own identity begs the question about life's essential character. **Froese**'s mention of life's self-creation of identity epitomizes Jonas's view, which, as I mentioned in passing, assumes that life is based on self-positing activity, thereby reducing the evidence to the claim that life cannot be passive because otherwise it would not be active. **Froese**'s final remark in §11 that subjects – differently from rocks – can change their behavior even in absence of a change in external conditions brings this point home. It is trivially true that a rock cannot stand up and walk away from the bedrock it is resting upon, while I can do just that (most of the times). But what is under discussion is precisely *why* I can do that, *how* I do it, and under *what* conditions. Saying that subjects can and rocks cannot because the former are active and the latter are passive does not answer any of those questions.

« 28 » I take **Lowe**'s very detailed commentary as a direct answer to the series of questions I just raised and, partially, to the technical shortcomings **Froese** previously stressed. **Lowe** maintains that while Ashby's basic approach has shown to be very fruitful, it has turned out to be difficult to scale to multiple parallel tasks. His suggestion is that the key to further progress lies in relenting Ashby's requirement that ultrastability be achieved through random selection of control parameters. The resulting strategy, which he calls the "non-random self-configurable approach," would be able to incorporate a more sophisticated version of homeostasis (i.e., Sterling's 2004 allostasis) and allow more satisfactory modeling of those short-term non-adaptive behaviors that may result in long-term organismic gains. **Lowe** claims that this approach would account for the undeniable fact that biological agents cannot be passive-contingent in the sense stipulated by Ashby because they often need to operate outside their comfort zone. **Lowe**'s proposal could thus be construed as a first answer to the question my statement (2') above implicitly asked: "What would a passive-contingent life be like?" It would be a life structurally unable to leave short-term comfort zones, would be the reply. While such a form of life may prove adequate for certain classes of organisms (a possibility **Lowe** does not seem to consider), it would seem to be inadequate as a general requirement.

« 29 » As a philosopher, I cannot avoid being struck by the similarity between the short-term perspective that would seem to follow from Ashby's required random selection of internal parameters and the traditional characterizations of vegetable (or, rather, vegetative) and animal life within the Western tradition (for example in Aristotle's descriptions of plant-like and animal-like forms of life in *On Generation and Corruption*, a description that became enshrined in the philosophical canon). **Lowe**'s proposal about life *in general* would thus be a reformulation of the classic view of *human* life. Aristotle argued that human beings must have a plant-like soul responsive to the organism's internal drives (growth and reproduction); an animal-like soul responsible for the short-term emotion-based interactions with the outside world (fear being the basic emotion that determines an animal's relationship to its environment); and a rational soul that *directs* those lower levels by superimposing long(er)-term planning and ratiocination on life's basic functions. **Lowe**'s hierarchy repeats this classic setup by translating the Aristotelian necessity of *logos* in human life into the necessity of a "mental autonomy [involving] a top-down regulatory influence on [plant-and animal-like] homeostasis" (§10). The analogy with classic philosophical conceptions raises a question that, I believe, remains open in **Lowe**'s description. Namely: the Aristotelian distinction of three different kinds of life principle (or "souls" in the traditional translation) had also the goal of demarcating different regions within the biological universe. Is such a demarcation implicitly entailed by **Lowe**'s non-random self-configurable approach to life? In other words, are all the modules the approach requires necessary for *all* forms of life, or are we confronted with a possibly fundamental biological distinction among different forms? Perhaps only further work in this direction will answer these questions.

« 30 » However, there is a possible alternative route toward a regained ontological uniformity in the biological domain. **Lowe**'s commentary refers to my brief mention of Ashby's more sophisticated version of the homeostat, the Dispersive Adaptive Multistable Systems (DAMS). While it is a fact that Ashby failed to produce a satisfactorily working DAMS, let alone conduct biologically meaningful experiments with it, I do not think its overall architectural principles are necessarily doomed, as several scholars seem to imply (e.g., Pickering 2010, but also Husbands & Holland 2008). As **Vernon** rightfully stresses, Ashby's core interest was the concept of multistability, which is in fact the topic of the second half of *Design for a Brain* (in both editions). A multistable system is a collection of suitably connected ultrastable sub-systems (the details of this connection vary from the first to the second edition, the latter giving a more prominent role to the environment). In the target article, I pointed out Ashby's explicit claims, on purely theoretical grounds, that the shortcomings of simple generalized homeostatic systems **Lowe** and **Froese** underlined and that Ashby had to a certain extent anticipated (as **Vernon** correctly indicates in §20) would be avoided by a DAMS-like multistable system. This ambitious and very specific claim has not, as far as I know, been empirically tested. The perhaps excessively elliptic remark on which I concluded the target article referred precisely to this possibility: the construction of a (simulated) DAMS-like device would be the next step toward an exploration of passive-contingent life along Ashbian lines and is indeed one of the longer-term goals of the research program the target article introduced.

On Gustav Theodor Fechner and Sigmund Freud

« 31 » **Malmgren** points out that even if Ashby's general homeostasis thesis – and its related connections to passivity, contingency, and heteronomy – may put him at odds with a substantial portion of the Western intellectual tradition, as I claimed in the target article, his work is very much consistent with the psychological tradition that starts with Gustav Fechner and goes through Sigmund Freud. I completely agree with him. Indeed, I have written on the topic myself, even though, admittedly, I did not dwell at length on the relationship between Ashby and Fechner and focused rather on that with Freud.[7] I do indeed agree with **Malmgren** that the connections between Ashby's theory and psychoanalysis are quite robust, and especially so if we look at Freud's early work (such as the posthumous *Project for a Scientific Psychology* 1964). It is true that, as far as I know, there are few documented instances of an intellectual exchange between the two. Nonetheless, at the personal level, it is similarly true that psychoanalysis attracted strong interest in the late 1940s and 1950s, and especially so in England, due to the influence of Anna Freud's London-based school. In my view, the theoretical connections between Ashby's theory of general homeostasis and Sigmund Freud's conceptualization of an organism's reaction to external stimuli are evident, and indeed were often noted in the 1950s.[8] An important difference, which I stressed in the paper cited above, is that Freud sometimes speaks as if he conceived of an almost ontological distinction between animal homeostasis, which almost always succeeds, and human homeostasis, which almost always fails.[9] This distinction was picked up by the Lacanian school, for which, indeed, human homeostasis is *always* failing, with the consequence that the human being is conceived as a structurally maladapted animal (whose ultimate roots are to be found in humans' neotenic character).[10] Nonetheless, I think that there is much work to be done to re-uncover a possibly cybernetic (or, rather, Ashbian) reinterpretation of psychoanalysis. My 2011a paper was an early step meant to further this line of inquiry. I find Catherine Malabou's recent work (2012), although very critical of psychoanalytic orthodoxy, as very much proceeding in the same direction. In my opinion, her notion of "plasticity" (1996) has important similarities to Ashby's view of homeostasis through ultrastability. But this is not the place to explore this point further.

« 32 » One learning and memory, I cannot but agree with **Malmgren** as well. Perhaps I did not express myself clearly enough. When I stressed (in §35 cited by Malgrem in his §4) that the homeostat lacks "explicit or implicit forms of memory," I meant to say that Ashby's view on learning were in explicit contradiction with theories of explicit learning and memory that see the latter as the storing of representation-like objects or chains of representation-like objects (i.e., plans) as well as theories of implicit learning and memory that follows the orthodox Hebbian dictum ("neurons that fire together wire together"). Ashby was indeed very much interested in showing that *even in the absence of explicit representations or of modified neuronal structures*, an organism could achieve equilibrium in changing as well as in repeating environmental situations (assuming the environment is relatively stable). The whole second half of *Design of a Brain* is meant to bring this point home, in fact, as I believe I stressed in the target article. I thank **Malmgren** for giving me the opportunity to stress this aspect of my interpretation of Ashby. Even though I am personally not primarily interested in learning and memory as a main area of research (we all need to pick our horses, after all), I agree with **Malmgren** that Ashby was considerably invested in the issue. I would also draw the conclusion, which remains somewhat implicit in **Malmgren**'s discussion of learning in §§5–6, that Ashby's theory of "learning" (or, in Ashbian terms: "improved adaptive performances in repeating environmental conditions") is as radical as his overall view of life and strongly deserving of renewed attention.

Received: 5 November 2013
Accepted: 9 November 2013

---

[7] This is partly because Fechner's view of stability, while similar to Freud's, as the latter explicitly admits, differs from it in significant respects both at the technical level and in terms of the general framework. See Franchi (2011a) for a discussion.

[8] Among the few studies on the subject, see Walker (1956) and the brief notes in Alexander (1948), who takes for granted the identity between Freud's principle of constancy and Claude Bernard's and Walter Cannon's notions of physiological homeostasis. The behaviorist school, and especially Clark Hull's drive theory (1943), took the analogy even further.

[9] I have briefly dealt with this issue in Franchi (2010). Much work remains to be done, obviously.

[10] Lacan (1988: 51ff.) has an explicit discussion of cybernetics and of Grey Walter's turtles. Lacan's conception of human psychic apparatus as structurally failing was already present in his early pre-psychoanalytic work (2001: 30ff.) and is later intensified until it turns humans into, literally, retarded apes suffering from arrested development.

Combined References

Abelson H. & DiSessa A. (1982) Turtle geometry. The computer as a medium for exploring mathematics. MIT Press, Cambridge MA.

Alexander F. (1948) Fundamental of psychoanalysis. W. W. Norton, New York.

American Psychiatric Association (APA) (2000) Diagnostic and statistical manual of mental disorders (DSM-IV-TR). Fourth edition. American Psychiatric Association, Washington DC.

An G. & Wilensky U. (2009) From artificial life to in silico medicine: NetLogo as a means to translational knowledge representation in biomedical research. In: Komosinski M. & Adamatzky A. (eds.) Artificial life models in software. Springer, Frankfurt: 183–214.

Asaro P. (2011) Computers as models of the mind: on simulations, brains and the design of early computers. In: Franchi S. & Bianchini F. (eds.) The search for a theory of cognition: early mechanisms and new ideas. Rodopi, Amsterdam: 89–114.

Ashby W. R. (1933) The physiological basis of neurosis. Proceedings of the Royal Society of Medicine 26(11): 1454–1460.

Ashby W. R. (1947) The nervous system as a physical machine; with special reference to the origin of adaptive behavior. Mind 56: 44–59.

Ashby W. R. (1951) Review of B. P. Babkin, Pavlov: A biography. The British Journal of Psychiatry 97(408): 596.

Ashby W. R. (1952a) Can a mechanical chess-player outplay its designer? The British Journal for the Philosophy of Science III(9): 44–57.

Ashby W. R. (1952b) Design for a brain. First edition. John Wiley & Sons, New York.

Ashby W. R. (1953) Mechanical chess player. In: Foerster H. von, Mead M. & Teuber H. L. (eds.) Cybernetics. Circular causal and feedback mechanisms in biological and social systems. Transactions of the Ninth Macy Conference, March 20–21, 1952, Josiah Macy, Jr. Foundation, New York NY: 151–154.

Ashby W. R. (1954a) Design for a brain. First edition. Reprinted with corrections. John Wiley & Sons, New York.

Ashby W. R. (1954b) The application of cybernetics to psychiatry. The British Journal of Psychiatry 100(418): 114–124.

Ashby W. R. (1956a) Design for an intelligence amplifier. In: Shannon C. & McCarthy J. (eds.) Automata studies. Annals of Mathematics studies no. 34. Princeton University Press, Princeton NJ: 215–234.

Ashby W. R. (1956b) Introduction to cybernetics. Chapman & Hall, London.

Ashby W. R. (1957) Pavlov reconditioned. Review of I. P. Pavlov, Experimental psychology and other essays. The British Journal for the Philosophy of Science 8(31): 249–252.

Ashby W. R. (1960) Design for a brain. Second edition. John Wiley & Sons, New York.

Ashby W. R. (1970) Connectance of large dynamic (cybernetic) systems: Critical values for stability. In: Conant R. (ed.) Mechanisms of intelligence: Ross Ashby's writings on cybernetics. Intersystems Publications, Seaside CA: 89–91. Originally published in: Nature 228(5273): 784.

Ashby W. R. (1971) Review of E. Maggio, Psychophysiology of lerning and memory. The British Journal of Psychiatry 119(552): 574.

Avila-Garcìa O. & Cañamero L. (2005) Hormonal modulation of perception in motivation-based action selection architectures. In: Proceedings of the symposium "Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action" at the AISB'05 Convention. University of Hertfordshire, Hatfield: 9–17.

Barandiaran X. E. & Egbert M. D. (2013) Norm-establishing and norm-following in autonomous agency. Artificial Life 1–24.

Barandiaran X. E. & Ruiz-Mirazo K. (2008) Modelling autonomy: Simulating the essence of life and cognition. Biosystems 91(2): 295–304.

Braitenberg V. (1984) Vehicles. Experiments in synthetic psychology. MIT Press, Cambridge MA.

Brooks R. A. (1991) Intelligence without representation. Artificial Intelligence 47: 139–159.

Clark A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences 36(3): 181–204.

Di Paolo E. A. (2000) Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. In: Meyer J.-A., Berthoz A., Floreano D., Roitblat H. L. & Wilson S. W. (eds.) From animals to animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior. MIT Press, Cambridge MA: 440–449.

Di Paolo E. A. (2003) Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In: Murase K. & Asakura T. (eds.) Dynamical systems approach to embodiment and sociality. Advanced Knowledge International, Adelaide: 19–42.

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.

Dreyfus H. L. (2007) Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. Philosophical Psychology 20(2): 247–268.

Egbert M. D. (2013) Bacterial chemotaxis: introverted or extroverted? A comparison of the advantages and disadvantages of basic forms of metabolism-based and metabolism-independent behavior using a computational model. PLoS ONE 8: e63617.

Egbert M. D., Barandiaran X. E. & Di Paolo E. A. (2010) A minimal model of metabolism-based chemotaxis. PLoS Computational Biology 6(12): e1001004.

Egbert M. D., Barandiaran X. E. & Di Paolo E. A. (2012) Behavioral metabolution: the adaptive and evolutionary potential of metabolism-based chemotaxis. Artificial Life, 18 (1): 1–25.

Egbert M. D., Di Paolo E. A. & Barandiaran X. E. (2009) Chemo-ethology of an adaptive protocell: Sensorless sensitivity to implicit viability conditions. In: Advances in artificial life. Proceedings of the 10th European Conference on Artificial Life. Springer, Berlin: 242–250.

Ellenberger H. (1970) The discovery of the unconscious. Basic Books, New York.

Foerster H. von, Mead M. & Teuber H. L. (eds.) (1953) Cybernetics. Circular causal and feedback mechanisms in biological and social systems. Transactions of the Ninth Macy Conference, March 20–21, 1952, Josiah Macy, Jr. Foundation, New York NY

Franchi S. (2005) Hunters, Cooks, and Nooks. Diacritics 33(2) 98–109.

Franchi S. (2006) Herbert Simon, anti-philosopher. In: Magnani L. (ed.) Computing and philosophy. Associated International Academic Publishers, Pavia: 27–40.

Franchi S. (2010) Can machines have an unconscious? Would they have to? In: Mainzer

K. (ed.) Proceedings of the VIII European Conference on Philosophy and Computing (ECAP10). Verlag Dr. Hut, Munich: 506–513.

Franchi S. (2011a) Jammed machines and contingently fit animals: Psychoanalysis's biological paradox. French Literature Series 38: 213–256.

Franchi S. (2011b) Life, death, and resurrection of W. Ross Ashby's homeostat. In: Franchi S. & Bianchini F. (eds.) The search for a theory of cognition: Early mechanisms and new ideas. Rodopi, Amsterdam: 3–52.

Freud S. (1964a) Beyond the pleasure principle. In: The standard edition of the psychological works of Sigmund Freud. Volume XVIII. Hogarth Press, London. Original published in 1920.

Freud S. (1964b) Project for a scientific psychology. In: The standard edition of the psychological works of Sigmund Freud. Volume I. Hogarth Press, London: 283–387.

Froese T. (2009) Hume and the enactive approach to mind. Phenomenology and the Cognitive Sciences 8(1): 95–133.

Froese T. (2010) From cybernetics to second-order cybernetics: A comparative analysis of their central ideas. Constructivist Foundations 5(2): 75–85. Available at http://www.univie.ac.at/constructivism/journal/5/2/075.froese

Froese T. (2011) From second-order cybernetics to enactive cognitive science: Varela's turn from epistemology to phenomenology. Systems Research and Behavioral Science 28: 631–645.

Froese T. & Ikegami T. (2013) The brain is not an isolated "black box," nor is its goal to become one. Behavioral and Brain Sciences 36(3): 213–214.

Froese T. & Stewart J. (2010) Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. Cybernetics & Human Knowing 17(4): 83–106.

Froese T. & Stewart J. (2013) Enactive cognitive science and biology of cognition: A response to Humberto Maturana. Cybernetics & Human Knowing 19(4): 61–74.

Froese T. & Ziemke T. (2009) Enactive artificial intelligence: Investigating the systemic organization of life and mind. Artificial Intelligence 173(3–4): 466–500.

Froese T., Virgo N. & Ikegami T. (in press). Motility at the origin of life: Its characterization and a model. Artificial Life.

Froese T., Virgo N. & Izquierdo E. (2007) Autonomy: A review and a reappraisal. In: Almeida e Costa F., Rocha L. M., Costa E., Harvey I. & Coutinho A. (eds.) Advances in artificial life. Proceedings of the Ninth European Conference on Artificial Life. Springer, Berlin: 455–465.

Gánti T. (2003) The principles of life. Oxford University Press, New York.

Gaussier P., Lepretre S., Quoy M., Revel A., Joulain C. & Banquet J. P. (2000) Experiments and models about cognitive map learning for motivated navigation. In: Demiris J. & Birk A. (eds.) Interdisciplinary approaches to robot learning. World Scientific, Singapore: 53–94.

Gomila A. & Müller V. C. (2012) Challenges for artificial cognitive systems. Journal of Cognitive Science 13(4): 453–69.

Hanczyc M. M. & Ikegami T. (2010) Chemical basis for minimal cognition. Artificial Life 16(3): 233–243.

Hanczyc M. M., Toyota T., Ikegami T., Packard N. & Sugawara T. (2007) Chemistry at the oil-water interface: Self-propelled oil droplets. Journal of the American Chemical Society 129(30): 9386–9391.

Harvey I. (2008) Homeostasis via chaos: Implementing the uniselector as a dynamical system. In: Bullock S., Noble J., Watson R. A. & Bedau M. A. (eds.) Proceedings of the 11th International Conference on Artificial Life. MIT Press, Cambridge MA: 774.

Harvey I. (2011) Opening stable doors: Complexity and stability in nonlinear systems. In: Lenaerts T., Giacobini M., Bersini H., Bourgine P., Dorigo M. & Doursat R. (eds.) Advances in artificial life, (ECAL 2011). MIT Press, Cambridge MA: 318–325.

Harvey I., Husbands P., Cliff D., Thompson A. & Jakobi N. (1997) Evolutionary robotics: The Sussex approach. Robotics and Autonomous Systems 20: 205–224.

Hull C. L. (1943) Principles of behavior. An introduction to behavior Theory. D. Appleton-Century, New York.

Husbands P. & Holland O. (2008) The Ratio Club: A hub of British cybernetics. In: Husbands P., Hollands O. & Wheeler M. (eds.) The mechanical mind in history. MIT Press, Cambridge MA: 91–148.

Husserl E. (2001) Analyses concerning passive and active synthesis. Lectures on transcendental logic. Translated by A. J. Steinbeck. Dordrecht: Kluwer.

Iizuka H. & Di Paolo E. A. (2008) Extended homeostatic adaptation: Improving the link between internal and behavioural stability. In: Asada M., Hallam J. C. T., Meyer J.-A. & Tani J. (eds.) From animals to animats 10: Proceedings of the Tenth International Conference on Simulation of Adaptive Behavior (SAB 2008). Springer, Berlin: 1–11.

Iizuka H., Ando H. & Maeda T. (2013) Extended homeostatic adaptation model with metabolic causation in plasticity mechanism – Toward constructing a dynamic neural network model for mental imagery. Adaptive Behavior 21(4): 263–273.

Ikegami T. (2007) Simulating active perception and mental imagery with embodied chaotic itinerancy. Journal of Consciousness Studies 14(7): 111–125.

Ikegami T. (2013) A design for living technology: Experiments with the mind time machine. Artificial Life 19(3–4): 387–400.

Ikegami T. & Suzuki K. (2008) From a homeostatic to a homeodynamic self. Biosystems 91(2): 388–400.

Jonas H. (1984) The imperative of responsibility. Foundations of an ethics for the technological age. Chicago University Press, Chicago IL.

Jonas H. (2001) The phenomenon of life: Toward a philosophical biology. Northwestern University Press, Evanston IL. Originally published in 1966.

Lacan J. (1988) The Ego in Freud's theory and in the technique of psychoanalysis, 1954–1955. Seminar II. W. W. Norton, New York.

Lacan J. (2001) Les complexes familiaux dans la formation de l"individu. In: Autres écrits. Autres écrits. Le Seuil, Paris: 23–84. Originally published in 1936.

Lowe R. & Ziemke T. (2011) The feeling of action tendencies: On emotional regulation of goal-directed behaviour. Frontiers in Psychology 346(2): 1–24.

Lowe R. & Ziemke T. (2013) The role of reinforcement in affective computation: Triggers, action and feeling. In: Proceedings of the 2013 IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (CICAC). Research Publishing, Singapore: 17–24.

Lowe R., Montebelli A., Ieropoulos I., Greenman J., Melhuish C. & Ziemke T. (2010) Grounding motivation in energy autonomy: A study of artificial metabolism constrained robot dynamics. In: Fellermann H., Dörr M., Hanczyc M., Laursen L., Maurer S., Merkle D., Monnard P.-A., Sty K., & Rasmussen S. (eds.) Artificial life XII. MIT Press: Cambridge MA: 725–732.

Malabou C. (1996) L'avenir de Hegel: Plasticité, temporalité, dialectique. Vrin, Paris.

Malabou C. (2012) The new wounded: From neurosis to brain damage. Fordham University Press, New York.

Malmgren H. (2006) The essential connection between representation and learning. Philosophical Communications, Web Series 38. University of Gothenburg, Gothenburg.

Manicka S. & Di Paolo E. A. (2009) Local ultrastability in a real system based on programmable springs. In: Kampis G., Karsai I. & Szathmary E. (eds.) Advances in artificial life. Proceedings of the Tenth European Conference on Artificial Life (ECAL09). Berlin: Springer: 87–94.

Maturana H. R. (2011) Ultrastability … autopoiesis? Reflexive response to Tom Froese and John Stewart. Cybernetics & Human Knowing 18(1–2): 143–152.

Maturana H. R. & Varela F. J. (1980) Autopoiesis and cognition. The realization of the living. Reidel, Dordrecht.

Maturana H. R. & Varela F. J. (1987) The tree of knowledge. The biological roots of human understanding. New Science Library, Boston.

McFarland D. (2008) Guilty robots, happy dogs. Oxford University Press, New York.

McFarland D. & Bösser T. (1993) Intelligent behavior in animals and robots. MIT Press, Cambridge MA.

McFarland D. & Spier E. (1997) Basic cycles, utility and opportunism in self-sufficient robots. Robotics and Autonomous Systems 20: 179–190.

Melhuish C. R., Ieropoulos I., Greenman J. & Horsfield I. (2006) Energetically autonomous robots: Food for thought. Autonomous Robots 21: 187–198.

Montebelli A., Lowe R., Ieropoulos I., Melhuish C. R., Greenman J. & Ziemke T. (2010) Microbial fuel cell driven behavioral dynamics in robot simulations. In: Fellermann H., Dörr M., Hanczyc M., Laursen L., Maurer S., Merkle D., Monnard P.-A., Sty K., & Rasmussen S. (eds.) Artificial life XII. MIT Press: Cambridge MA: 749–756.

Muntean I. & Wright C. D. (2007) Autonomous agency, AI, and allostasis. Pragmatics and Cognition 15(3): 485–513.

Nolfi S. & Floreano D. (2000) Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines. MIT Press, Cambridge MA.

Pfeifer R. & Scheier C. (1999) Understanding intelligence. MIT Press, Cambridge MA.

Pickering A. (2010) The cybernetic brain: Sketches of another future. University of Chicago Press, Chicago IL.

Pitonakova L. (2013) Ultrastable neuroendocrine robot controller. Adaptive Behavior 21(1): 47–63.

Raichle M. E. & Snyder A. Z. (2007) A default mode of brain function: A brief history of an evolving idea. NeuroImage 37(4): 1083–1090.

Riegler A. (2002) When is a cognitive system embodied? Cognitive Systems Research 3(3): 339–348.

Ruiz-Mirazo K., Pereto J. & Moreno A. (2004) A universal definition of life: Autonomy and open-ended evolution. Origins of Life and Evolution of the Biosphere: 323–346.

Schrödinger E. (1944) What is life? The physical aspect of the living cell. Cambridge University Press, Cambridge MA.

Sterling P. (2004) Principles of allostasis: Optimal design, predictive regulation, pathophysiology and rational therapeutics. In: Schulkin J. (ed.) Allostasis, homeostasis, and the costs of adaptation. Cambridge University Press, Cambridge.

Tertullian (1954) Opera. Edited by E. Dekkers, J. G. P. Borleffs, R. Willems, R. F. Refoulé, G. F. Diercks & A. Kroynman, Volume 1. Opera catholica. Adversus Marcionem. Brepols, Turnhout.

Thompson E. (2007) Mind in life: Biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge MA.

Turing A. M. (1950) Computing machinery and intelligence. Mind 59: 433–460.

Varela F. J. (1979) Principles of biological autonomy. Elsevier North Holland, New York.

Villalobos M. (2013) Enactive cognitive science: revisionism or revolution? Adaptive Behavior 21(3): 159–167.

Walker N. (1956) Freud and homeostasis. British Journal for the Philosophy of Science 7: 61–72.

Walter W. G. (1961) The living brain. Penguin Books, Harmondsworth.

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.

Wiener N. (1948) Cybernetics. First edition. MIT Press, Cambridge MA.

Wiener N. (1976) Collected works with commentaries. Volume IV. Edited by P. Masani. MIT Press, Cambridge MA.

Wiener N. & Schadé J. P. (1963) Introduction to neurocybernetics. Progress in Brain Research 2: 1–7.

Ziemke T. & Lowe R. (2009) On the role of emotion in embodied cognitive architectures: From organisms to robots. Cognitive Computation 1: 104–117.