sarily) vague, but clearly covers topics with multiple overlapping circularities at an abstract level.

« 26 » It is therefore highly appropriate that the article discusses many of the component links in such circularities, introduces Daisyworld models (and variants) that are relatively simple exemplars of circular causation, and introduces Goodwin's principle (§65).

« 27 » But it is regrettable that the authors then seek simplistic linear "*X* causes *Y*" summaries of their analysis, when circular causation requires so much more. Daisyworld is a minimal straightforward well-defined system of circular causation, yet typically causes so much misunderstanding. How much more room for confusion there is in the article's conclusion (§92) with less strictly defined terms such as autonomy, learning, evolution and life.

Inman Harvey has a mathematical, philosophical and anthropological background. He is a founder member of the Evolutionary and Adaptive Systems group at the University of Sussex, and his research interests include evolutionary robotics, artificial life and Gaia theory. http://users.sussex.ac.uk/~inmanh

Received: 17 June 2018
Accepted: 18 June 2018

Further Support for the Stabilization Thesis: Circular Causality, Ecosystems Growth & Development, and Allostasis

David Vernon
Carnegie Mellon University Africa, Rwanda • vernon/at/cmu.edu

> Upshot • I highlight certain aspects of the target article's arguments and identify additional support that reinforces both the foundations of the argument and its conclusions.

« 1 » By identifying the origin of *causal accountability* in the system itself as a key attribute of autonomy, Niall Palfreyman and Janice Miller-Young's target article presents an extended argument leading the authors to postulate the *Stabilization Thesis* (§92). This states that an autonomous system exhibits a bi-directional mutual conditioning of the system's local structural components and the system's global dynamical flows: upwards, by which structure conditions dynamical flow, and downwards, by which dynamical flow selects structural variation. This results in self-stabilization of the system and the emergence of a self-sustaining autonomous identity.

« 2 » This bi-directional mutual conditioning is sometimes referred to as *circular causality*.[1] Specifically, Scott Kelso (1995) uses the term to describe the situation in dynamical systems where the cooperation of the individual parts of the system determines the global system behaviour, which, in turn, governs the behaviour of these individual parts. In circularly causal systems, global system behaviour influences the local behaviour of the system components and yet it is the local interaction between the components that determines the global behaviour. Circular causality exists between levels of a hierarchy of system and sub-system. This influence of macroscopic levels on microscopic levels in a system is also captured in the term *downward causation* (Thompson & Varela 2001; Seth 2010), i.e., that global-to-local aspect of circular causality whereby the global system behaviour causally influences the individual system components. It has been noted that this phenomenon of circular causality may be one of the pivotal mechanisms in autonomous cognitive systems (Vernon 2014).

« 3 » In the same vein, the characterization of flows and agents (§20) and the agent-flow architecture in which "agents condition cumulative changes in diffusive flows, which in turn constrain stochastic variation in the states of those agents" (§25) is similar to the characterization by Robert Ulanowicz (1998, 2000, 2011) of the manner in which dissipative ecosystems achieve dynamic stability through sustainable growth and development in the face of precarious environmental conditions. Modelling the ecosystem as a flow network, Ulanowicz asserts that self-sustaining autonomous ecosystems tend to self-organize so as to optimize a function of the average mutual information in the entire network, referred to as *ascendancy*. This global system attribute forms the basis of the local adaptation of the individual agents in the ecosystem. He postulates a principal of *optimal* ascendancy whereby the system dynamically stabilizes in a state that has sufficient order and organization to facilitate prediction and withstand environmental perturbations but sufficient disorder (i.e., entropy) to allow alternative states to be self-selected if there is a failure in some part of the system. The greater the unpredictability of the environment, the higher the amount of entropy required. The balance between the two is dynamic and depends on the long-term variability of the environmental perturbations. Environments that are less precarious and less prone to change will result in a balance biased towards greater ascendancy since it is well matched by an internal organization with low surprisal; *vice versa*, an environment that is highly uncertain needs to be matched by a system that has a capability to deal with a greater degree of surprisal and, hence, has more entropy in its organization. Ulanowicz's model is also hierarchical and parallels the concept of niche-construction, whereby environmental fields can be shaped by the autonomous system "into a local niche that functions quasi-independently of the remaining environment" through a "recursively self-constituting organism-niche dynamic" (§42) as well as the concept of downward selection of constituent components (§60), dynamical stabilization (§61), and self-stabilization (§65).

« 4 » In §66, Palfreyman and Miller-Young note that stabilization goes beyond the notion of homeostasis, in that stabilization requires a link, as we have seen already, between the local structural variation (agents) and the dynamic variables (flows). This global-to-local link is also reflected in the process of *allostasis* (Sterling 2004; Muntean & Wright 2007; Sterling 2012), which differs from homeostasis in its predictive character and in its ability to anticipate and adapt to change rather than

1| Circular causality in a single system should be distinguished from Andy Clark's concept of continuous reciprocal causation (CRC) between two systems (Clark 1997) which "occurs when some system S is both continuously affecting and simultaneously being affected by, activity in some other system O" (Clark 1998: 356).

resist it, echoing the predictive capacity of the Stableworld example (§78). Allostasis provides a *global* mechanism for overriding normal homeostasis, serving the organism as a whole with the resources previously learned to be necessary to meet predicted environmental pressures. Significantly, allostasis, which also appears in social systems (Schulkin 2011), is effected at a higher level of organization, involving greater number of sub-systems acting together in a coordinated manner with global processes modulating local ones, again reflecting the character of circular causality and the dynamics of the Stableworld example (§78). Conversely, mechanisms for homeostasis operate at a simpler level of negative feedback control. Anil Seth extends the concept of allostasis, viewing it as "the process of achieving homeostasis" (Seth 2015: 7), emphasizing its roots in cybernetics, in general, and the ultrastability of Ross Ashby's homeostat (Ashby 1960; Vernon 2013), in particular. He notes that the fundamental cybernetic principle is for systems to ensure their continued existence by successfully responding to environmental perturbations in order to maintain their internal organization. He goes further, stating that "the purpose of cognition […] is to maintain the homeostasis of essential variables *and of internal organization* (ultrastability)" (Seth 2015: 8, emphasis added). Francisco Varela makes the same point, stating that "an autopoietic machine is a homeostatic […] system that has its own organization […] as the fundamental invariant" (Varela 1979: 13). Both these statements echo the target article's assertion that "A system is autonomous if it self-stabilizes, i.e., it stabilizes its own ability to stabilize" (§62). This assertion is further supported by Humberto Maturana's and Francisco Varela's definition of autonomy as "the condition of subordinating all changes to the maintenance of the organization" (Maturana & Varela 1980: 135) and Mark Bickhard's concept of *recursive self-maintenance*, whereby systems adjust their processes of self-maintenance as the environment shifts so as to maintain self-maintenance (Bickhard 2000).

« 5 » Having highlighted additional support for the stabilization thesis, I close the commentary by making one complementary point. Palfreyman and Miller-Young make clear that their focus is not on an operational account (§5), in the sense of aiming to create a theory, computational or otherwise, that can be implemented and validated. However, such an operational theory would nevertheless be desirable: It could ground the argument and give it a substance that would allow their stabilization thesis to be embraced by a wider range of communities in all domains concerned with the issues of autonomy, cognition, and development. Furthermore, since the thesis is based on dynamical systems, it merits some attempt at a formal explication if the meaning of crucial aspects such as flows (§20), are to be fully understood. In §6, the authors state that their theory should be capable of justifying a decision to conclude that artificial robots are inherently non-autonomous. Conversely, if one concluded that robots could be autonomous, the theory, or an operational version of it, should ideally be capable of justifying that position and stating the architectural entailments in concrete terms that would allow scientists and engineers to implement such an autonomous robot.

David Vernon is a professor at Carnegie Mellon University Africa in Rwanda. His main interest is cognitive robotics, focusing on cognitive architectures, development, and models of autonomy. Over the past 38 years, he has worked for Westinghouse Electric in Ireland and the USA, for the European Commission and Science Foundation Ireland, and for leading universities in Europe, Russia, and the Middle East. He is a past Fellow of Trinity College Dublin, co-chair of the IEEE Robotics and Automation Technical Committee for Cognitive Robotics, and an associate editor of *Cognitive Systems Research*. http://www.vernon.eu

Received: 18 June 2018
Accepted: 21 June 2018

A Few Pending Challenges from the Perspective of a Theory of Organisms

Maël Montévil
Institut de Recherche et d'Innovation, France
mael.montevil/at/gmail.com

> Upshot • I discuss convergences between the approach of the authors and my work aiming for a theory of organisms. I also discuss some pitfalls and challenges pertaining to biological randomness, which, I argue, require original developments.

« 1 » Biology faces a big challenge: It lacks an encompassing theoretical framework for studying organisms, their physiology, development and behaviors. This challenge is often overlooked because empirical analyses at the molecular level dominate the biological field. However, this reductionism is incomplete. In general, reductionism proceeds by decomposing the object of study followed by its theoretical recomposition to ensure that it is properly understood. In biology, however, there are no reliable methods or theoretical framework to provide guidance for such recompositions when studying organisms. Overcoming this limitation is especially important if we want to harness the opportunities provided by Big Data and ensure that they provide biologically meaningful results. Moreover, a suitable theoretical framework for biological organisms should help us overcome the shortcomings of current medical and pharmacological methods, and provide insights into the many changes that technological developments bring about and which characterize the anthropocene.

« 2 » This challenge has led a group of biologists, philosophers and mathematicians, including myself, to work together and propose several principles for a theory of organisms by building on existing theoretical traditions (Soto et al. 2016a). Here, I will use the perspective developed in this work to discuss several points from the target article of Niall Palfreyman and Janice Miller-Young. While my perspective differs from theirs, since my starting point is more