Spontaneous Neural Activity and the Self: A Neuroscience Perspective

Alison Hanson* and Rafael Yuste

*To whom correspondence should be addressed: alison.hanson@nyspi.columbia.edu

Originally written March 2018

Introduction

The study of the self has historically been under the purview of philosophy where it has been a prominent subject for over two thousand years. Only recently, with the advent of human neuroimaging techniques, has the subject of self entered the realm of neuroscience (Northoff et al., 2006; Qin and Northoff, 2011; Sui and Humphreys, 2017). To date, most work on the self in neuroscience has focused on human neuroimaging studies resulting in the prevailing view that the self is an emergent property of only the most evolutionarily advanced higher cortical brain regions, thought to be unique to humans (Alcaro et al., 2017). However, this view conflates "self" with "self-awareness" and ignores the possibility of there being a sense of "self," or of having a subjective experience, without the need to be aware that one is having such a subjective experience. Thus, rather than the self appearing late in evolution and being unique to humans, there is much evidence to suggest a "primitive" sense of self was likely present in the very first life forms, and is, in fact, a fundamental aspect of all living systems.

Given the diverse definitions of "self" in the literature, let us be clear about what we mean when discussing "self" here. Here, we define "self" as any living system that draws a boundary between its "self" and "not-self," or, more simply, between "inside" and "outside" (such as in the case of a cell membrane or body) with the "self" being located on the "inside" of the boundary. This separation of the "self" from the rest of the "outside" world via a boundary construction allows the "self" to both be influenced by the external environment (to generate internal "experience" of the external world) and to influence or act on its environment across its boundary. Key to this process of "self-construction" is the generation of an internal model of the "self" in relation to the "outside" world. It is this "self-model" which we will be using as our definition of "self" here.

A minimal self at the origin of life?

Before delving into the origins of the self, we must first delve into the origins of life by addressing the question: what distinguishes living from non-living systems? This question was both eloquently asked and answered by Schrödinger in his classic book, *What is Life?*, in which he suggested the major difference between non-living and living systems is non-living systems tend toward disordered, high entropy states with low energy at or near equilibrium, whereas living systems have a surprising tendency to remain in ordered, low entropy states with high energy far from equilibrium (Schrödinger, 1944). How living systems maintain such "order from disorder" and avoid dissipation and decay, as mandated by the second law of thermodynamics, was a central question of this work.

Maturana and Varela put forth an additional answer to Schrödinger's question with their theory of autopoiesis (from the Greek *auto*, meaning "self" and *poiesis*, meaning "creation" or "production") in which they proposed that the major defining feature of all living systems is their ability to continually self-reproduce themselves (Maturana and Varela, 1980; Varela et al., 1974). The quintessential example of a living, or autopoietic, system, upon which they based their original theory, is the biological cell, which has two key components: (i) a boundary (in the form of a semi-permeable membrane) capable of both spatially and organizationally distinguishing the system from its environment by separating inside from outside, and (ii) an internal metabolic network capable of producing its own components, including the components necessary to maintain the system's boundary (i.e. its semi-permeable membrane). Stated differently, the living system of the biological cell distinguishes itself both spatially and organizationally from its surrounding environment by constructing a boundary (its cell membrane), which requires its internal metabolic network to continually repair and regenerate itself, while its internal metabolic network simultaneously requires the presence of the cell's boundary (its membrane) to maintain itself and prevent the diffusion of its components back into the environment; a process that would lead to disintegration of the system, i.e. cell death. Thus, the cell emerges as a result of an ongoing circular process in which these two components (its membrane and its internal metabolic network) co-emerge, each dependent on the other, to form a unified whole, and in which the cell continuously re-creates the boundary between itself and everything else. Hence, the evolution of life entails the emergence of the first autonomous, self-organizing individual capable of distinguishing its "self" from its surroundings, or, more broadly, from that which is "other" (Thompson, 2007; Varela, 1975, 1979, 1991).

Accordingly, a major distinction between living systems, like a biological cell, and non-living physical systems, like a machine, is the autonomous, internal, self-determining nature of living systems as opposed to the heteronomous, externally-determined nature of physical systems (Thompson, 2007; Varela et al., 1974). This is not to suggest that living organisms are entirely autonomous and independent from their environments, however. While they are operationally closed and independent organizational systems with their own internal identities, they are simultaneously thermodynamically open and intimately coupled with their environments on which they depend for constant exchange of metabolic components and energy (Maturana and Varela, 1980; Varela, 1979, 1991).

Let us take the example of one of the first autonomous, living systems to emerge: the bacterial cell. Despite their seeming simplicity, bacteria exhibit complex, flexible responses to their environment, including migrating towards rewarding and away from noxious stimuli in the process of chemotaxis (Bi and Sourjik, 2018). For example, bacteria are able to measure sucrose concentration gradients, which requires a basic form of memory, and swim toward the area of highest concentration, while alternatively measuring the concentration gradients of and swimming away from toxins. Hence, with the emergence of a simple bacterial cell emerges both: (i) an internal identity, which establishes a perspective or a "self-pole," which then entails (ii) the emergence of a world perceived through the perspective of the "self-pole" of the bacterium. In this way, the arrival of the bacterium brings forth the arrival of meaning in the world in which the previously indifferent physico-chemical milieu in which it is embedded is instantly

transformed into a world of biological significance in which molecules become "good," as in the case of sucrose that serves to preserve its life, or "bad," as in the case of acid, which would hasten its demise. Prior to the existence of the bacterium, all the elements of the physico-chemical soup were equally insignificant. With the evolution of this particular life form, however, suddenly sucrose and acid (and all other physico-chemical elements) become imbued with meaning according to the viewpoint of the bacterium (Merleau-Ponty, 2003; Thompson, 2007; Varela, 1991, 1997). Thus, at the outset, life entails a minimal autonomous "self" with an internal identity and perspective, which in turn entails a world with a "surplus of significance" from the individual organism's point of view (Weber and Varela, 2002).

It is not enough for a living system to simply distinguish itself from its world, however. It must constantly interact with and adapt to its world in order to achieve its primary purpose of continuing its own being, i.e. its own survival (Thompson, 2007). Thus, in establishing its own physical "body," not only does the bacterium determine which elements in its environment are "good," "bad," or "neutral" in terms of its survival; it then generates adaptive behavior by either approaching, avoiding, or disregarding such elements in its surroundings (Varela, 1991). In this way, the bacterium is constantly interacting with and "making sense of" its world to remain viable by approaching things that sustain itself and avoiding things that threaten its existence in an ongoing primitive sensorimotor loop. This realization led Varela to claim that "all life is sense-making," which necessarily entails basic intentionality and cognition in its most primitive form (Weber and Varela, 2002).

Recent work by Friston on the "free-energy principle" takes these ideas one step further in an attempt to provide a mathematical formalization of living systems (Friston, 2010; Friston et al., 2006; Kirchhoff et al., 2018; Ramstead et al., 2018). In this work, Friston proposes that self-organizing (or autopoietic) systems maintain their structure and organization (and thus avoid disorder and decay) by generating statistical models of their environments which strive to minimize information theoretic free energy, or "surprise." He argues that all living systems are essentially "Bayesian inference machines" that generate internal predictive models of their environments that are continuously updated as the organism receives sensory input from its environment. The stated primary aim of the organism's model is to be as accurate a predictor of the environment as possible to avoid "surprising" or life-threatening conditions and thus maintain the organism in expected, self-sustaining environments. To do this, organisms must: (i) generate an internal predictive model of themselves in relation to their environment (in the form of a hierarchical predictive coding model), and (ii) minimize their model's long-term "surprise," or short-term "prediction error;" or, stated differently, they must maximize their model evidence (Friston, 2010; Ramstead et al., 2018). Organisms are thus involved in a continuous process in which they constantly compare the sensory input they receive from their environment with their model's prior predictions. If the sensory input they receive is consistent with their prediction (i.e. there is low "prediction error"), the organism has no impetus to act. If, however, the sensory input is not consistent with what the organism predicts (i.e. a bacterium finds itself in a highly acidic environment), it has one of two choices: (i) to act on its environment to change the sensory input to match its prediction (i.e. to swim toward an area of low acid

concentration which is consistent with its expectation), or (ii) to change its prediction (i.e. change its internal model) (Ramstead et al., 2018).

A key point here is that the organism does not simply *generate* a model of the world, it *is* a model of the world; a model which is encoded in its physical structure (Friston, K., 2011). That is, the physical structure of the bacterium (with its unique receptors and internal metabolic network) *embodies* the predictive model; a model which tunes its parameters by altering its physical structure. Consequently, if all models strive to maximize model evidence, and the organism *is* the model, the organism therefore strives to maximize evidence for its own existence. In other words, it is a "self-evidencing machine" that models itself as existing (Friston, K., 2011; Hohwy, 2014). Thus, all living systems possess a predictive "self-model" encoded in their physical structure, which is continually updated and shaped by ongoing interaction with the environment, and which is essential for maintaining their existence in the face of constantly changing conditions. Together, both the theory of autopoiesis and the free-energy principle imply the emergence of life thus entails the emergence of a minimal self with a minimal predictive self-model capable of integrating both internal and external input to generate coherent, adaptive behavior, which is essential for organism survival.

A bottom-up model of the self

How might such predictive self-models be implemented in simple organisms without nervous systems? Contrary to the widely held belief that neurons are special or unique in their computational abilities, computation is substrate-independent, requiring only a dynamic network of a sufficient number of linked components that send and integrate signals, along with the ability to alter their connectivity based on prior activity (Baluska and Mancuso, 2009; Baluška and Levin, 2016; Sterling and Laughlin, 2015). These basic properties are present in biological systems at all levels of scale from subcellular protein networks in single cells to groups of interacting cells in multi-cellular organisms (Baluška and Levin, 2016). In addition, the basic building blocks of neurons and neural networks (ion channels, electrical synapses mediated by gap junctions, and neurotransmitters) are all evolutionarily ancient and are already present in bacteria, suggesting that electrical coupling of cells is as old as life itself (Baluska and Mancuso, 2009; Baluška and Levin, 2016). In fact, it has recently been shown that single bacterial cells are electrically excitable with spontaneous fluctuations in their membrane voltage and that groups of bacteria in biofilms participate in long-range electrical signaling via propagation of synchronized potassium waves, appearing to act as a primitive "microbial brain" (Kralj et al., 2011; Prindle et al., 2015) . Moreover, recent work in a variety of plant and animal systems demonstrates that many non-neural cell types participate in bio-electric signaling networks throughout the organism to control body patterning during development, regeneration, and cancer suppression (Baluška and Levin, 2016; Levin and Martyniuk, 2018; Pezzulo and Levin, 2015). Thus, electrical signaling and computation are not new phenomena in biology and do not depend solely on neural tissue. Rather, it appears the nervous system likely co-opted and improved upon the much older electrical signaling mechanisms originally used by bacteria to coordinate biofilm formation and function (Baluška and Levin, 2016; Levin and Martyniuk, 2018).

These findings imply the basic tools required for constructing a predictive self-model capable of basic computation were present from the beginning to allow even single cells to integrate their internal and external inputs and increase their chances of survival. Thus, rather than a model with the self located only at the top of the evolutionary tree of life, these data support a bottom-up organization of the "self," in line with complex adaptive systems theory (Holland, 1996; Kauffman, 1993). In this view, a single cell with its minimal self-model (likely implemented mostly by subcellular protein networks) can aggregate into larger groups, such as bacteria in biofilms, where single cells can communicate (via ancient gap junctions, ion channels, neurotransmitters, and other signaling mechanisms) to form larger "selves" with more complex internal self-models. In this way, the boundary between "self" and "not-self" and "inside" and "outside" gets re-defined and expands to include a group of previously individual, separate "selves" into a larger, more complex organism. This process can be iterated repeatedly by continuously re-defining and expanding the boundary between "self" and "not-self" to form increasingly complex "selves" with correspondingly increasingly complex internal self-models up to the level of humans (comprised of groups of groups of cells), at which point the self-model begins to model itself modeling the world (i.e. becomes "self-aware") (Kirchhoff et al., 2018; Palacios et al., 2017). Importantly, the original physical boundary around the minimal self (the single cell) does not disappear when it becomes aggregated into a larger self (a multi-cellular organism). Rather, the single cell remains an autonomous entity, but it becomes incorporated into a larger structure in which the theoretical boundary between "self" and "other" widens to include the entire structure.

Top-down self-regulation

This bottom-up conceptualization of the self then raises the question: how do the previously individual, autonomous "selves" (i.e. single cells) "know" their boundary has shifted and they have become part of a larger "self" (as in a multi-cellular organism), and how do all the previously independent internal self-models get integrated into one larger, coherent self-model? This leads us to the concept of emergence, which is a fundamental feature of self-organizing systems (Holland, 1998; Kauffman, 1993). Emergence in self-organizing systems works in two ways: both bottom-up and top-down in a reciprocal manner (Campbell, 1974; Ellis et al., 2012; Hoel et al., 2013) .That is, interactions among lower level components (i.e. individual units or cells) give rise to emergent properties at a higher level of scale (i.e. a multicellular organism), which then reciprocally constrains the dynamics of the lower level. Thus, prior to forming a multicellular organism, individual cells exhibit a wide range of behavior, but when incorporated into a larger structure, their behavior becomes constrained by this higher level of organization. This is commonly referred to as the "enslaving principle" in which the faster dynamics at the lower level of scale become "enslaved" by the slower dynamics at the higher level of scale (Haken, 1987).

One candidate top-down mechanism by which smaller "selves" (i.e. single cells) may become "enslaved" or constrained by the larger "self" (i.e. the multicellular organism) is via synchronous organism-wide electrical signaling that simultaneously communicates to all the individual units they have become part of a larger whole. As described above, this has been found in many organisms lacking nervous systems,

including bacterial biofilms where synchronous electrical signaling has been observed to coordinate nutrient sharing within and between bacterial communities (Liu et al., 2017; Prindle et al., 2015). In addition, it is becoming increasingly clear that organisms set up body-wide electrical circuits within somatic tissue that orchestrate the position of individual cells during development and regeneration and appear to be important in preventing individual cells from "going rogue," as occurs in tumorigenesis (Levin and Martyniuk, 2018; Pezzulo and Levin, 2015). Thus, it appears organism-wide electrical circuits are one top-down mechanism utilized by biological systems to coordinate organism unity out of otherwise autonomous components.

If living systems can form unified selves with integrated self-models in the absence of neural tissue, what advantage might a nervous system provide? In short: speed. Prior to the evolution of neurons, single cells relied mostly on diffusion for system-wide communication and computation, which is energetically very cheap and efficient over short distances, but becomes prohibitively slow as organisms increase in size (Sterling and Laughlin, 2015). Thus, the evolution of larger selves required the evolution of faster and more energetically expensive communication in the form of analogue, or bio-electrical signaling, as seen in bacterial biofilms and non-neural somatic tissues of multicellular organisms (Levin and Martyniuk, 2018; Prindle et al., 2015). This nearly 1,000-fold increase in speed allowed organism coherence over longer distances, as well as more complex self-models capable of integrating more internal and external information, thus allowing more complex behavior. With the advent of all-or-nothing action potentials, the speed of communication and computation increased another 100-fold allowing the emergence of even larger selves with ever-more complex self-models (Sterling and Laughlin, 2015).

Hence, it seems the main advantage of a nervous system is its increased speed of communication and computation which enables both: (i) fast long-range signaling to ensure organism unity across longer distances, thus allowing the evolution of larger selves, and (ii) the ability to efficiently integrate more internal and external inputs to generate more complex self-models, thus allowing more complex behavior. It is not known how the more recently evolved faster neuro-electrical signaling interfaces with the evolutionarily ancient bio-electrical signaling; however, just as bio-electrical signaling appears to coordinate and constrain the activity of individual bacteria in biofilms, it may be that neuro-electrical signaling in turn coordinates and constrains bio-electrical signaling networks among groups of cells in higher multi-cellular organisms. These remain interesting and open questions.

Spontaneous, synchronous neural activity

Having considered the various ways in which aneural organisms likely embody minimal selves with minimal self-models, we will now turn our attention to how the nervous system might contribute to the development of larger and more complex selves in more detail. As discussed, it is likely that synchronous electrical signaling, which has been discovered in bacterial biofilms, plants, and non-neural somatic tissue, plays an important role in maintaining multi-cellular organism unity (i.e. in maintaining a coherent self) (Ciszak et al., 2016; Levin and Martyniuk, 2018; Prindle et al., 2015). Interestingly, spontaneous, synchronous neural activity was also discovered in human brains in 1929 when Hans Berger first used electroencephalography (EEG) to discover intrinsic

synchronous brain waves, which he termed "alpha waves" (Berger, 1929; Gloor, 1969). Despite Berger's early work showing the presence of ongoing intrinsic synchronous electrical brain activity, his work was largely ignored in favor of the prevailing view of the brain as merely a "reflex organ" that is only active in response to an external stimulus, as advanced by Sherrington (Sherrington, 1906). Consequently, the majority of neuroscience research following Berger's initial discoveries was focused on task-evoked responses in which spontaneous neural activity was seen as mere "noise" that most study designs attempted to suppress (Mitra and Raichle, 2016; Raichle, 2015a; Yuste et al., 2005).

It wasn't until the late 1990s when both Shulman and Raichle independently noticed a surprising result: in human neuroimaging experiments designed to study task-evoked neural activity, they observed a consistent decrease in neural activity in specific cortical areas while subjects were performing various goal-directed tasks as compared to a "resting state" in which the subject's eyes were closed or fixated on a cross (Raichle et al., 2001; Shulman et al., 1997). They showed that, contrary to the widely-held belief that the brain only increases its activity in response to external stimuli, there appeared to be ongoing, intrinsic brain activity "at rest" that was consistently inhibited during attention-demanding, non-self-referential tasks. This intrinsic, resting state network became known as the brain's "default mode network," which has become an area of intense study (Raichle, 2015b).

The discovery of the default mode network reignited interest in spontaneous brain activity that had mostly been ignored since Berger's initial discoveries. It is now known that the human brain produces a variety of spontaneous neural oscillations spanning a wide range of frequencies from the ultraslow (0.01-1.0 Hz) to the ultrafast (200-600 Hz); a frequency distribution which interestingly obeys a power-law, indicating a scale-invariant frequency structure (Buzsáki, 2006; Buzsáki et al., 2013). Accordingly, the same neural oscillation frequency distribution found in humans has now been identified in all mammalian brains studied to date; brains that vary widely in size (more than 17,000-fold). Such a robust frequency structure is one of the most highly conserved features of mammalian brains (Buzsáki et al., 2013) . This highly conserved structure, in addition to the fact that intrinsic brain activity consumes up to 20% of total body energy, suggests that, rather than being mere "noise," spontaneous brain activity is likely critical for brain function (Buzsáki et al., 2013; Zhang and Raichle, 2010). These findings led Buzsaki to propose that neural oscillations are likely the functional or "syntactical" units of the elusive "neural code" in which the faster, smaller, and more local oscillations become entrained, integrated, or "read" by the slower, larger, and more global oscillations, as in the "enslaving principle" discussed above (Haken, 1987). According to this theory, the highest frequency neural oscillations function as the "letters" of the code, which then get integrated or "read" by lower frequency oscillations at the level below thus forming "words," which in turn get integrated or "read" by the next lowest frequency level forming "sentences," and so on down the hierarchy (Buzsáki, 2010). Taken to its logical conclusion, this theory implies the presence of an ultimate downstream integrator or "reader" at the lowest frequency level, which integrates or "reads" all the higher frequency information in the brain. Despite this highly conserved and intriguing structure of intrinsic brain activity, its function remains largely unknown.

The default mode network and the self

The default mode network (DMN), introduced above, has become one of the most extensively studied intrinsic brain networks since its discovery nearly twenty years ago (Raichle, 2015b). While its function remains unclear, it is thought to be primarily involved in mental processes that take place during the "resting state," such as: spontaneous thought, episodic memory, mind-wandering, and, interestingly, self-related processing (Raichle, 2015b). Numerous studies have now been completed showing significant overlap between resting state activity in the cortical midline structures thought to comprise the DMN and those active during self-related processing (i.e., the perigenual anterior cingulate cortex (PACC), ventro- and dorsomedial prefrontal cortex (vmPFC, dmPFC), supragenual anterior cingulate cortex (SACC), posterior cingulate cortex (PCC), and the precuneus) (Northoff et al., 2006; Qin and Northoff, 2011). These findings have been replicated using a variety of self-specific versus non-self-specific stimuli in multiple domains, including facial, emotional, verbal, spatial, motor, and memory, in which subjects routinely respond more robustly to self-specific versus non-self-specific stimuli (Northoff et al., 2006; Qin and Northoff, 2011). Regardless of the domain studied, the same cortical midline structures active in the DMN at rest were also activated during self-specific stimulus processing, leading Northoff to propose that the DMN contains, or encodes, self-specific information (Northoff, 2016). Exactly how the DMN might encode self-specific information is not known.

A clue to how the DMN might contribute to the human self may lie in its structure. In accordance with the overall small world network architecture of the brain in which there are many short, local connections and few long-range connections between nodes, the DMN appears to serve as one of the brain's main integrators that connects major "rich hubs" via long-range thickly myelinated axons (Cabral et al., 2014; Deco et al., 2011; Lord et al., 2017; Park and Friston, 2013). This network architecture puts the DMN in a central position in the brain in which it both receives and sends information rapidly among otherwise segregated local brain regions. It is thought the DMN receives exteroceptive input from all of the primary sensory areas as well as interoceptive input from the insula, thalamus, hypothalamus, midbrain, and brainstem, and, in turn, can rapidly send information back to and between these same areas (Bär et al., 2016; Northoff et al., 2006; Panksepp and Northoff, 2009; Raichle, 2015b) . Thus, the DMN seems to be in a unique position to act as the ultimate downstream "reader" or integrator in the brain, as predicted by Buzsaki's theory discussed above. Interestingly, the DMN has been found to oscillate in the ultraslow range (0.01-0.1 Hz), consistent with this hypothesis (Fox and Raichle, 2007; Thompson and Fransson, 2015; Van Dijk et al., 2010).

Another way to think about the potential role of the DMN in self-construction is as the top layer of the hierarchical predictive coding "self-model" as put forth by Friston and discussed above (Friston, 2010; Kirchhoff et al., 2018). Like Buszaki's theory which predicts the need for an ultimate downstream brain integrator or "reader" (i.e. a "self"), a hierarchical predictive coding model also implies the need for an ultimate brain integrator or "predictor" (also a "self") at the top of the hierarchy. According to predictive coding brain models, prediction error is passed up the hierarchy from the low-level primary, unimodal sensory areas to the ultimate, multi-modal "predictor" at the top of the hierarchy that contains a high-level abstract representation (of the "self") that then

passes predictions back down to the lower levels (Friston, 2010; Kirchhoff et al., 2018). In this way, the DMN, oscillating at the lowest frequency in the brain, might act as the brain's ultimate information integrator, receiving input from all the lower-level, otherwise isolated units (oscillating at higher frequencies), and passing on one unified "self" prediction back down to generate coherent, adaptive behavior. Importantly, if the DMN is the ultimate integrator or "self-model" predictor, one would expect it to be continuously active to maintain whole-brain and organism unity. Accordingly, the DMN appears to be incessantly active with high baseline activity that is never completely turned off even during sleep or under anesthesia (Raichle, 2015b). In addition, there is mounting evidence that disruption of DMN activity via psychedelics or meditation correlates with "ego dissolution," or the loss of a sense of self, consistent with the idea that the DMN plays an important role in the formation of the self in humans (Brewer et al., 2011; Carhart-Harris and Friston, 2010; Carhart-Harris et al., 2014; Garrison et al., 2015; Lebedev Alexander V. et al., 2015; Tomasino et al., 2013).

      Taken together, these findings indicate the DMN may be implementing a top-down control mechanism in the human brain as it receives bottom-up information from all brain areas (which oscillate at higher frequencies) and may, in turn, constrain these lower levels via its slow-wave oscillations, while also rapidly communicating its unified output to all brain regions via its synchronous electrical activity to maintain organism unity (i.e. a coherent self). Hence, the human "self" may be constructed bottom up with the DMN emerging as the ultimate neural integrator and top-down "enslaver" of all the lower levels of organization in the entire body, including likely bio-electrical signaling networks in the somatic tissues, which in turn enslave smaller groups of cells, on down. Importantly, this view of the human self does not imply that the DMN *is* the self or that the self is a *thing* located *in* the DMN. Rather, it suggests that the self is an ongoing *process* in which the DMN continuously receives internal and external sensory information and adaptively updates its predictive model of itself and the world. Thus, while the DMN is likely critical for self-construction, the self does not reduce to the DMN as the self requires the entire body in constant communication with its environment and is therefore both embodied and embedded, in accord with the enactive and embodied cognition view (Clark, 2013; Thompson, 2007; Varela et al., 2016).

A sub-cortical self
      In line with the idea that the self is not solely a product of higher-order cortical structures, there is evidence to suggest that the DMN contains subcortical nodes in the midbrain and brainstem that are highly conserved among mammals and co-active with cortical DMN nodes, thus forming a cortical-subcortical DMN (Bär et al., 2016; Northoff and Panksepp, 2008; Panksepp and Northoff, 2009). These findings led Northoff and Panksepp to propose a "basis model" for the self in which an unconscious, pre-reflexive or "proto-self" is present in all mammals as a product of their subcortical DMN. They argue that the sub-cortical DMN provides lower animals with a minimal subjective experience and orientation towards the world, which is critical for basic emotions and survival. They propose that a conscious, or reflexive self, is likely built upon this evolutionarily ancient "proto-self," and mediated by the more recently evolved cortical DMN structures (Northoff and Panksepp, 2008; Panksepp and Northoff, 2009). While this view supports a bottom-up self-model, they propose that a basic or minimal self is

only present in "mammals with a particular neural structure." In contrast, as outlined above, there is much evidence to suggest that a minimal self does not require a particular neural structure, but, rather, only requires an autopoietic organization, which is present at the origin of life (Maturana and Varela, 1980; Varela, 1979; Varela et al., 1974). In addition, a key aspect of a coherent self seems to be a mechanism to ensure organism unity, which, like computation, is likely implemented in various substrate-independent ways in living systems (i.e. via synchronous bioelectrical signaling networks and/or synchronous neuro-electrical signaling networks with potentially diverse structures distinct from the mammalian central nervous system).

Along these lines, recent work done in multiple "lower" organisms, including zebrafish, fruit flies, and even the fresh water polyp, *Hydra vulgaris*, which possesses one of the first nervous systems in evolution, demonstrates they all generate intrinsic synchronous neural activity in the ultraslow frequency range (0.01-0.1 Hz) (Ahrens et al., 2013; Dupre and Yuste, 2017; Mann et al., 2017). These findings indicate that ultraslow frequency spontaneous neural oscillations, as first discovered in the human DMN, are conserved throughout all nervous systems in evolution, further underscoring their likely importance for brain function. Furthermore, if the DMN in humans is the ultimate brain integrator (or top-layer of the hierarchical predictive self-model), it may be that the homologous resting state networks in these simpler systems also serve as the ultimate integrators of their simpler self-models, making them critical for coordinated organism behavior (i.e. a unified self).

Conclusions and future directions

This then brings us to an overall model of the self which likely arose with the emergence of the first life form and has been built upon and elaborated ever since. Early on were minimal selves (single cells) with minimal self-models likely implemented by intracellular protein networks. As selves grew in size and complexity they required faster long-range signaling via bio-electrical signaling to maintain organism unity out of otherwise autonomous components. As selves further evolved, they required yet faster communication over larger distances to ensure coherence of their smaller "selves," leading to the birth of the nervous system with its rapid action potentials. This long evolutionary line leads to the present-day human; the most complex self with the most complex self-model known—a model that has become aware of itself. Exactly how the nervous system contributes to the self remains a mystery, but spontaneous, synchronous electrical activity likely plays an important role; an activity present in all living systems studied to date from bacteria to humans (Buzsáki et al., 2013; Ciszak et al., 2016; Levin and Martyniuk, 2018; Prindle et al., 2015).

The picture thus painted is one that is much less anthropocentric than the prevailing view in which the self is presumed to be limited to humans with higher order cortical brain structures. If, instead, the self is not unique to humans, but, rather, permeates life, the door opens to studying fundamental aspects of the self in simpler systems. This is good news as the human brain is one of the most complex systems known and deciphering its function remains an immense challenge (Yuste and Church, 2014). Thus, studying simpler nervous systems may allow the discovery of the basic structural and functional design principles of neural circuits, including those of resting state networks, which are likely critical for self-construction and function. Hence, rather

than depending solely on neuroimaging techniques with poor temporal and spatial resolution to study the DMN in humans, simpler model systems may be used allowing more precise imaging and manipulation of intrinsic brain activity (Bosch et al., 2017).

An ideal model system for studying fundamental aspects of the self is the fresh water polyp, *Hydra vulgaris*, with one of the first and simplest nervous systems in evolution, as introduced above (Bosch et al., 2017). The nervous system of *Hydra* is arranged in a diffuse nerve net containing a few hundred to a few thousand neurons depending on the size of the animal (David, 1973). Its small size and diffuse neuronal arrangement allows the simultaneous imaging of each of its neurons with single cell resolution; a feat recently accomplished, revealing the emergence of three major behavioral networks and one resting state network out of a seemingly "simple" nerve net (Dupre and Yuste, 2017). The ability to observe the entire *Hydra* brain at single cell resolution makes it possible to characterize each of its behavior-generating neural circuits and their relationship to its resting state network in great detail. In addition to a simple nervous system, *Hydra* also exhibits a dozen simple and robust behaviors that have recently been categorized and quantified using machine learning (Han et al., 2018). Thus, using this simple system, one can begin to rigorously empirically address questions such as: What is the relationship between resting state synchronous neural activity and the generation of coherent animal behavior (i.e. a coherent "self"), and what happens to its behavior when its resting state network is disrupted (via optogenetic manipulation, pharmacologic manipulation, or physical manipulation, such as in split-head *Hydra*)? Moreover, *Hydra* reproduce by budding (Clarkson and Wolpert, 1967), and exhibit ongoing synchronous neural activity between mother and daughter, allowing the study of how such shared synchronous activity might produce coordinated behavior while they are physically attached, and how their physical separation might lead to desynchronized resting state networks with resultant uncoordinated, independent behavior. Lastly, given that *Hydra* is likely on the cusp of the transition from self-models generated by bio-electrical signaling to models with neuro-electrical signaling at the helm, it would be interesting, and possible, to test how its resting state neural network is related to its bio-electrical signaling networks in its non-neural tissue, which would provide further insight into the brain-body connection and the organism's overall self-construction mechanisms.

In sum, we propose a bottom-up concept of the self that arises with the origin of life and is likely constrained by organism-wide electrical signaling from the top-down. The higher cortical brain structures, arriving late on the evolutionary scene, are therefore likely built upon much evolutionarily older selves, and only add awareness of a self that was there in an increasingly complex manner all along. Thus, it is likely that fundamental aspects of the self can be studied in more experimentally tractable simple systems, such as *Hydra*, to gain further insight into how resting state neural activity, and its underlying bio-electrical activity, might be related to self-construction and coherent animal behavior. Understanding the basic mechanisms of self-construction and maintenance in simple systems will likely be applicable to living systems at all levels of scale from bacteria to humans and have important implications for neurology, psychiatry, and, potentially, tumorigenesis.

Bibliography

Ahrens, M.B., Orger, M.B., Robson, D.N., Li, J.M., and Keller, P.J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat. Methods *10*, 413–420.

Alcaro, A., Carta, S., and Panksepp, J. (2017). The Affective Core of the Self: A Neuro-Archetypical Perspective on the Foundations of Human (and Animal) Subjectivity. Front. Psychol. *8*, 1424.

Baluška, F., and Levin, M. (2016). On Having No Head: Cognition throughout Biological Systems. Front. Psychol. *7*, 902.

Baluska, F., and Mancuso, S. (2009). Deep evolutionary origins of neurobiology: Turning the essence of "neural" upside-down. Commun. Integr. Biol. *2*, 60–65.

Bär, K.-J., de la Cruz, F., Schumann, A., Koehler, S., Sauer, H., Critchley, H., and Wagner, G. (2016). Functional connectivity and network analysis of midbrain and brainstem nuclei. NeuroImage *134*, 53–63.

Berger, H. (1929). über das elektrenkephalogramm des menschen. Arch. Psychiatr Nervenkr 527–570.

Bi, S., and Sourjik, V. (2018). Stimulus sensing and signal processing in bacterial chemotaxis. Curr. Opin. Microbiol. *45*, 22–29.

Bosch, T.C.G., Klimovich, A., Domazet-Lošo, T., Gründer, S., Holstein, T.W., Jékely, G., Miller, D.J., Murillo-Rincon, A.P., Rentzsch, F., Richards, G.S., et al. (2017). Back to the Basics: Cnidarians Start to Fire. Trends Neurosci. *40*, 92–105.

Brewer, J.A., Worhunsky, P.D., Gray, J.R., Tang, Y.-Y., Weber, J., and Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. Proc. Natl. Acad. Sci. U. S. A. *108*, 20254–20259.

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

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. Neuron *68*, 362–385.

Buzsáki, G., Logothetis, N., and Singer, W. (2013). Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. Neuron *80*, 751–764.

Cabral, J., Kringelbach, M.L., and Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. Prog. Neurobiol. *114*, 102–131.

Campbell, D.T. (1974). 'Downward Causation' in Hierarchically Organised Biological Systems. In Studies in the Philosophy of Biology, (Palgrave, London), pp. 179–186.

Carhart-Harris, R.L., and Friston, K.J. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. Brain J. Neurol. *133*, 1265–1283.

Carhart-Harris, R.L., Leech, R., Hellyer, P.J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D.R., and Nutt, D. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Front. Hum. Neurosci. *8*, 20.

Ciszak, M., Masi, E., Baluška, F., and Mancuso, S. (2016). Plant shoots exhibit synchronized oscillatory motions. Commun. Integr. Biol. *9*, e1238117.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. *36*, 181–204.

Clarkson, S.G., and Wolpert, L. (1967). Bud morphogenesis in hydra. Nature *214*, 780–783.

David, C.N. (1973). A quantitative method for maceration of hydra tissue. Wilhelm Roux Arch. Für Entwicklungsmechanik Org. *171*, 259–268.

Deco, G., Jirsa, V.K., and McIntosh, A.R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. Nat. Rev. Neurosci. *12*, 43–56.

Dupre, C., and Yuste, R. (2017). Non-overlapping Neural Networks in Hydra vulgaris. Curr. Biol. CB *27*, 1085–1097.

Ellis, G.F.R., Noble, D., and O'Connor, T. (2012). Top-down causation: an integrating theme within and across the sciences? Interface Focus *2*, 1–3.

Fox, M.D., and Raichle, M.E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. *8*, 700–711.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. *11*, 127–138.

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. J. Physiol. Paris *100*, 70–87.

Friston, K. (2011). Embodied inference: Or "I Think Therefore I Am, If I Am What I think". In W. Tschacher & C. Bergomi (Eds.), The Implications of Embodiment (cognition and communication) (pp. 89-125). (Exeter, UK: Andrews UK Ltd).

Garrison, K.A., Zeffiro, T.A., Scheinost, D., Constable, R.T., and Brewer, J.A. (2015). Meditation leads to reduced default mode network activity beyond an active task. Cogn. Affect. Behav. Neurosci. *15*, 712–720.

Gloor, P. (1969). Hans Berger and the discovery of the electroencephalogram. Electroencephalogr. Clin. Neurophysiol. Suppl 28:1-36.

Haken, H. (1987). Synergetics. In Self-Organizing Systems, (Springer, Boston, MA), pp. 417–434.

Han, S., Taralova, E., Dupre, C., and Yuste, R. (2018). Comprehensive machine learning analysis of Hydra behavior reveals a stable basal behavioral repertoire. ELife *7*.

Hoel, E.P., Albantakis, L., and Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. Proc. Natl. Acad. Sci. U. S. A. *110*, 19790–19795.

Hohwy, J. (2014). The Self-Evidencing Brain. Noûs *50*, 259–285.

Holland, J.H. (1996). Hidden order: how adaptation builds complexity (Reading, Mass.: Addison-Wesley).

Holland, J.H. (John H. (1998). Emergence : from chaos to order (Reading, Mass: Addison-Wesley).

Kauffman, S.A. (1993). The origins of order : self organization and selection in evolution (New York: Oxford University Press).

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. J. R. Soc. Interface *15*.

Kralj, J.M., Hochbaum, D.R., Douglass, A.D., and Cohen, A.E. (2011). Electrical spiking in Escherichia coli probed with a fluorescent voltage-indicating protein. Science *333*, 345–348.

Lebedev Alexander V., Lövdén Martin, Rosenthal Gidon, Feilding Amanda, Nutt David J., and Carhart-Harris Robin L. (2015). Finding the self by losing the self: Neural correlates of ego-dissolution under psilocybin. Hum. Brain Mapp. *36*, 3137–3153.

Levin, M., and Martyniuk, C.J. (2018). The bioelectric code: An ancient computational medium for dynamic control of growth and form. Biosystems *164*, 76–93.

Liu, J., Martinez-Corral, R., Prindle, A., Lee, D.-Y.D., Larkin, J., Gabalda-Sagarra, M., Garcia-Ojalvo, J., and Süel, G.M. (2017). Coupling between distant biofilms and emergence of nutrient time-sharing. Science *356*, 638–642.

Lord, L.-D., Stevner, A.B., Deco, G., and Kringelbach, M.L. (2017). Understanding principles of integration and segregation using whole-brain computational connectomics: implications for neuropsychiatric disorders. Philos. Transact. A Math. Phys. Eng. Sci. *375*.

Mann, K., Gallen, C.L., and Clandinin, T.R. (2017). Whole-Brain Calcium Imaging Reveals an Intrinsic Functional Network in Drosophila. Curr. Biol. CB *27*, 2389-2396.e4.

Maturana, H.R., and Varela, F.J. (1980). Autopoiesis and Cognition: The Realization of the Living (Springer Netherlands).

Merleau-Ponty, M. (2003). Nature : course notes from the Collège de France / (Evanston, Ill. : Northwestern University Press,).

Mitra, A., and Raichle, M.E. (2016). How networks communicate: propagation patterns in spontaneous brain activity. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *371*.

Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. Cogn. Neurosci. *7*, 203–222.

Northoff, G., and Panksepp, J. (2008). The trans-species concept of self and the subcortical-cortical midline system. Trends Cogn. Sci. *12*, 259–264.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., and Panksepp, J. (2006). Self-referential processing in our brain--a meta-analysis of imaging studies on the self. NeuroImage *31*, 440–457.

Palacios, E.R., Razi, A., Parr, T., Kirchhoff, M., and Friston, K. (2017). Biological Self-organisation and Markov blankets. BioRxiv 227181.

Panksepp, J., and Northoff, G. (2009). The trans-species core SELF: the emergence of active cultural and neuro-ecological agents through self-related processing within subcortical-cortical midline networks. Conscious. Cogn. *18*, 193–215.

Park, H.-J., and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. Science *342*, 1238411.

Pezzulo, G., and Levin, M. (2015). Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. Integr. Biol. Quant. Biosci. Nano Macro *7*, 1487–1517.

Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J., and Süel, G.M. (2015). Ion channels enable electrical communication in bacterial communities. Nature *527*, 59–63.

Qin, P., and Northoff, G. (2011). How is our self related to midline regions and the default-mode network? NeuroImage *57*, 1221–1233.

Raichle, M.E. (2015a). The restless brain: how intrinsic activity organizes brain function. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *370*.

Raichle, M.E. (2015b). The brain's default mode network. Annu. Rev. Neurosci. *38*, 433–447.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., and Shulman, G.L. (2001). A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. *98*, 676–682.

Ramstead, M.J.D., Badcock, P.B., and Friston, K.J. (2018). Answering Schrödinger's question: A free-energy formulation. Phys. Life Rev. *24*, 1–16.

Schrödinger, E. (1944). What is Life? The Physical Aspect of the Living Cell (Cambridge University Press).

Sherrington, C.S. (1906). The integrative action of the nervous system (New Haven, CT: Yale University Press).

Shulman, G.L., Corbetta, M., Buckner, R.L., Fiez, J.A., Miezin, F.M., Raichle, M.E., and Petersen, S.E. (1997). Common Blood Flow Changes across Visual Tasks: I. Increases in Subcortical Structures and Cerebellum but Not in Nonvisual Cortex. J. Cogn. Neurosci. *9*, 624–647.

Sterling, P., and Laughlin, S. (2015). Principles of Neural Design (MIT Press).

Sui, J., and Humphreys, G.W. (2017). The ubiquitous self: what the properties of self-bias tell us about the self. Ann. N. Y. Acad. Sci. *1396*, 222–235.

Thompson, E. (2007). Mind in Life: Biology, Phenomenology, and the Sciences of Mind (Harvard University Press).

Thompson, W.H., and Fransson, P. (2015). The frequency dimension of fMRI dynamic connectivity: Network connectivity, functional hubs and integration in the resting brain. NeuroImage *121*, 227–242.

Tomasino, B., Fregona, S., Skrap, M., and Fabbro, F. (2013). Meditation-related activations are modulated by the practices needed to obtain it and by the expertise: an ALE meta-analysis study. Front. Hum. Neurosci. *6*.

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., and Buckner, R.L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. *103*, 297–321.

Varela, F.J. (1975). A calculus for self-reference. Int. J. Gen. Syst. *2*, 5–24.

Varela, F.J. (1979). Principles of Biological Autonomy (North Holland).

Varela, F.J. (1991). Organism: A meshwork of selfless selves. In: A. Tauber (ed), Organisms and the Origins of Self (Dordrecht ; Boston : Norwell, MA, U.S.A: Kluwer Academic).

Varela, F.J. (1997). Patterns of Life: Intertwining Identity and Cognition. Brain Cogn. *34*, 72–87.

Varela, F.J., Maturana, H.R., and Uribe, R. (1974). Autopoiesis: the organization of living systems, its characterization and a model. Curr. Mod. Biol. *5*, 187–196.

Varela, F.J., Thompson, E., and Rosch, E. (2016). The embodied mind: cognitive science and human experience (Cambridge, MA: MIT Press).

Weber, A., and Varela, F.J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. Phenomenol. Cogn. Sci. *1*, 97–125.

Yuste, R., and Church, G.M. (2014). The new century of the brain. Sci. Am. *310*, 38–45.

Yuste, R., MacLean, J.N., Smith, J., and Lansner, A. (2005). The cortex as a central pattern generator. Nat. Rev. Neurosci. *6*, 477–483.

Zhang, D., and Raichle, M.E. (2010). Disease and the brain's dark energy. Nat. Rev. Neurol. *6*, 15–28.