Bistability of Somatic Pattern Memories: Stochastic Outcomes in Bioelectric Circuits Underlying Regeneration

Giovanni Pezzulo¹, Joshua LaPalme², Fallon Durant²,³, and Michael Levin²*

¹ Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
² Allen Discovery Center, Tufts University
³ Current affiliation: Harvard University

* Author for correspondence:
  200 Boston Ave.
  Suite 4600
  Medford, MA 02155
  Email: michael.levin@tufts.edu

Keywords: regeneration, bioelectricity, planaria, bistability, memory, cognitive science

Short title: Bistability in pattern memories
Abstract

Nervous systems and brains’ computational abilities are an evolutionary innovation, specializing and speed-optimizing ancient biophysical dynamics. Bioelectric signaling originated in cells’ communication with the outside world and with each other, in order to cooperate toward adaptive construction and repair of multicellular bodies. Here we review the emerging field of developmental bioelectricity, which links the field of basal cognition to state-of-the-art questions in regenerative medicine, synthetic bioengineering, cognitive science, and even machine learning and artificial intelligence. One of the predictions of this view is that regeneration and regulative development are able to restore correct large-scale anatomies from diverse starting states because, like the brain, they exploit bioelectric encoding of distributed goal states - in this case, pattern memories. Based on this idea, we propose a new interpretation of recent stochastic regenerative phenotypes in planaria, by appealing to computational models of memory representation and processing in the brain. Moreover, we discuss novel findings showing that bioelectric changes induced in planaria can be stored in tissue for over a week, thus revealing that somatic bioelectric circuits in vivo can implement a long-term, re-writable memory medium. A consideration of the mechanisms, evolution, and functionality of basal cognition makes novel predictions and provides an integrative perspective on the evolution, physiology, and biomedicine of information processing in vivo.
Introduction

Living bodies exhibit remarkable anatomical plasticity: the ability to reach specific, highly complex anatomical outcomes from a range of starting configurations and despite often drastic perturbations. Regulative development in many species can produce normal individuals when embryos are cut into pieces or multiple embryos are combined [1,2]. During metamorphosis, tadpoles with facial organs in abnormal positions nevertheless become normal frogs by moving these organs in un-natural paths until a correct frog face is made [3]. Many species, such as salamanders, regenerate whole organ systems, such as appendages, jaws, or eyes, when these are amputated in adulthood [4]. Indeed, classical studies showed that transplanting tails to a limb location in salamanders causes the organ to remodel into a limb [5,6], highlighting the amazing ability of tissues to make decisions about large-scale anatomical layout and re-adjust the anatomy to spec as needed.

Despite the recent progress in the molecular biology of genetic components necessary for regeneration or embryonic morphogenesis, important gaps exist in our understanding of the algorithms that are sufficient to implement such robust pattern homeostasis [7]. From an engineering perspective, biological systems can execute closed loop controls that modulate cell activity toward minimizing the error (anatomical distance in morphospace) between the current, possibly injured, anatomy and the correct species-specific target morphology. How do living systems measure their organ-level geometric states, and store the geometric setpoints toward which cell activity must work? How do cells know when the “correct” anatomy has been restored, so that proliferation and remodelling can cease?

These are critical open questions, likely requiring not only molecular genetics but also concepts from control theory, cybernetics, and computer science. At stake are numerous important endpoints. Regenerative medicine is stalled by the complexity explosion of genetic data: radical repair therapies require an understanding of how to trigger organ-level rebuilding and exert rational control over anatomical outcomes from tweaking molecular-level components. Similarly, repair of birth defects toward a correct anatomy, or the creation of entirely novel bodies for bioengineered synthetic living machines, require the ability to induce cells to build to a specific pattern. Even the cancer problem could be solved if we knew what signals normally keep cells harnessed toward a specific structure (and away from unicellular-like behaviour of metastasis) and cause them to stop proliferating once the target morphology has been reached.

Tissue Decision-Making: an evolutionary perspective on anatomical control

One approach to this problem focuses on the decision-making of cells and tissues, seeking to control their behaviour via signals and inputs, not by rewiring or micro-managing their molecular networks. The field of basal cognition seeks to understand the evolutionary history of learning, and problem-solving in animals by studying the mechanisms of computation, memory, and decision-making in ancient, pre-neural life forms [8–12,12–14]. This is relevant to problems of regeneration and development because it suggests the use of techniques from cognitive science and computational neuroscience to understand the information flow that leads to cell group behaviors [15]. We conjectured [16–18] that anatomical homeostasis is a process that relies on pattern memory – biophysical properties stored in tissues that encode, to some rough level of detail, important aspects of the anatomy towards which cells will build (and which, once achieved, causes further activity to stop).

Cognitive neuroscience studies neural circuits that implement memory and goal-directed behaviour. However, nerves evolved from more primitive cell types that made use of ion channels, neurotransmitters, and synaptic plasticity long before brains evolved [10,19]. The ability to process information via electrical signalling is as old as bacterial biofilms.
Therefore, we have been pursuing the idea that sophisticated cognitive memory may be an evolutionary elaboration of a much more primitive and ancient system, whose job was to remember body configurations to guide cell behaviour in morphogenesis before the arrival of nervous systems (which then further specialized to, e.g., remember configurations of the outside world to guide whole animal behavior). This perspective makes the prediction that by manipulating the dynamics of non-neural bioelectric networks, it should be possible to induce coherent, predictable changes to large-scale anatomy.

Indeed, it is now clear that slow bioelectric communication among all cells in the body, from the earliest stages of embryogenesis, is critical for patterning outcomes [22,23]. Endogenous bioelectric patterns are required for the normal morphogenesis of the brain [24,25], wing [26,27], face [28,29], and eye [30]. Mechanistic experiments in model systems [31][31] are complemented by studies of human channelopathies—mutations in ion channel genes that cause birth defect syndromes in human patients [32–34]. Furthermore, it is now possible to induce regenerative repair of complex appendages in non-regenerative conditions [35,36], and even reprogram tumors into normal tissue [37–40], by specific manipulation of ion channels that implement bioelectric state transitions in vivo. Beyond recreating the genome-default anatomical features, it was recently found that genomically wild-type animals (with no DNA editing) could be made to build heads belonging to other living species 150 million years distant, simply by manipulating the bioelectric network during regeneration [41,42].

Taken together, this body of work is suggesting that evolution exploited the powerful, convenient computational capabilities of electric circuits to implement anatomical repair control loops, just like it did for (closed-loop) behavioral circuits in the brain [43–45]. Cells communicate electrically to sense anatomical states at long range [46], and use electric circuits to organize into networks that produce voltage prepatterns guiding gene expression and morphogenesis [47]. Research is now on-going to refine computational models of information processing by non-neural bioelectric networks [48–50], which are becoming accurate enough that they can suggest interventions that repair complex birth defects [51].

The frontier in this field is to understand the bioelectric code: the mapping of bioelectric states in tissues to the patterns they encode. This is analogous to the research program in neuroscience called neural decoding, which is attempting to decode the cognitive content (e.g., a specific visual image or scene representation) implemented by brain states and dynamics, on a moment-by-moment basis [52–54]. This is usually done by exploiting machine learning algorithms that are trained on specific instances of the problem (e.g., a person's brain activations when he or she is attending to specific visual images); but the question of how exactly brain states and dynamics carry information and content remains open (and could be informed by parallel investigation of evolutionarily ancient and perhaps simpler information storage systems in non-neural tissues). Below, we discuss a recent finding in the field of developmental biology, which is unique because it focuses on the stochastic, dynamic nature of pattern memory processes.

**Planarian regeneration: stable and stochastic pattern memories**

Planaria are flatworms with complex anatomy and amazing regenerative abilities: when cut into pieces, each fragment knows exactly what to build and where to regenerate a complete, normal worm [55,56]. Recent studies reveal that this process is partly mediated by bioelectric signalling [57]. Specifically, the decisions of where to place heads on a fragment (and how many heads there should be) are mediated by a bioelectric circuit [58,59]. Critically, this circuit implements a long-term stable and re-writable pattern memory (Figure 1A): once it is altered (by a brief pharmacological exposure that targets electrical connections between cells) to encode a bi-polar head-head state (instead of a head-tail configuration), subsequent fragments continue to form 2-headed worms with no further
manipulation [60]. The axial polarity and head number are thus set by a physiological memory system that stably holds the information needed to guide regeneration – a transient (48 hour) stimulus is able to permanently alter the pattern to which fragments regenerate in the future. It is an example of epigenetics in Waddington’s original sense of the term: a non-genomic, re-writable medium for information that controls growth and form. Recent work has identified functional and molecular parallels of epigenetic pattern control between planaria and the classic example of non-genomic (cortical) inheritance seen in protozoa [61,62]. This suggesting ancient and conserved physical media across which evolution diversified its patterning control systems [63]. Here, we focus not on the permanent and stable re-writing of target morphology by the persistent 2-head phenotype but on a fascinating new phenomenon: stochastic target morphology.

**Hypothesis: bistability in memory systems**

When a cohort of planaria are cut into pieces and treated with a reagent that targets their bioelectric networks, some regenerate as 2-headed and some do not. For over a decade we called the resulting 1-head worms “escapees”, thinking that the vagaries of physiological diversity lowered the penetrance of the drug treatment and left some worms unaffected. However, re-cutting these seemingly-normal animals revealed a remarkable outcome [64,65]: their fragments likewise produce 1-head an 2-head animals in the same ratio. Thus, there are 3 kinds of worms: true wild-types that always produce 1-head worms, 2-head worms that always produce 2-head worms, and “cryptic” 1-head worms that have a de-stabilized target morphology. These cryptic worms have a bioelectric circuit that has been permanently shifted into a state in which it spontaneously settles into a 2-head or a cryptic state upon each cut (and this change did not require genomic manipulation). The state diagram summarizing these results is shown in Fig. 1B. Similar phenomena have been seen in bacterial “persisters” [66,67], but to our knowledge this is the first example of a permanently destabilized target morphology in metazoa.

Two interesting facts should be noted. First, the decision whether to make 1 head or 2 is made by each fragment independently: two pieces cut from the same worm do not always make the same choice – each piece undertakes its own “coin flip” between those outcomes (it is not set by the parent worm). Second, this decision is stochastic at the level of the fragment, not the individual cells. Thus, what we do not have mosaic, speckled worms in which 70% of the cells want to make one head and 30% want to make 2. Rather, each cell in the fragment agrees with the group decision to make 1 head or 2 (at a 70-30 frequency at the population level). This same phenomenon, where bioelectric networks enable group decisions within a large cell field, have also been described in left-right patterning (where targeting the decision by ion channel perturbations induces randomization of entire left or ride side identities, not individual cells [68], and in the conversion of melanocytes to melanoma (where bioelectric modulation cause a percentage of tadpoles, not of cells within one tadpole, to fully transform [69,70]). The stochasticity at the level of populations, but tight coordination within one individual, raises profound questions about noise, group behavior, and decision-making of multi-agent “swarms” that are not well understood but have many implications for not only biology and biomedicine but for dynamical systems and computational approaches to swarms of biological and robotic agents at many scales. Certainly, the problem of unifying the activity of numerous distributed, noisy micro-agents, such as bioelectrically-active neural cells, into a unified cognitive system with a coherent and functionally centralized decision-making capacity is a familiar problem in cognitive neuroscience.

Efforts are currently under way to reductively characterize the biophysical components of this stochastic process. The overall dynamics, in which a system can choose one of multiple global outcomes from the same starting conditions resembles the way visual
perceptual systems exhibits bistability and flip stochastically between two alternative interpretations of the same ambiguous data, e.g. either a duck or a rabbit for the famous duck-rabbit illusion, but never something in between a duck and a rabbit (Figure 2). Here, we seek a synthesis of conceptual aspects of these findings with state-of-the-art work in neuroscience, in an attempt to gain global level insight into this kind of process conducted by bioelectric networks that include non-neural cells. One possible way to understand this finding is that a planarian body can have two (or more) latent pattern memories, but only one gets (stochastically) expressed upon injury. An analogy with episodic memory mechanisms in the mammalian hippocampus - which also exhibits a form of multistability - may be helpful.

The hippocampal episodic code for spatial navigation: sequences of place cells

It is widely assumed that the hippocampus can encode multiples (episodic) memories [71]. Mechanistically, episodic memories may be "stored" as attractor states in (hippocampal) recurrent neural networks; this representation is convenient as it is high-capacity (because it minimizes interferences between different memories) and affords pattern-completion, thus permitting complete memories to be recalled from a subset of the cues that were present when they were firstly encoded; see [72,73] for computational analyses using Hopfield-style neural networks.

The episodes may not be encoded as single states, but in terms of sequences of states. Rodent research has elucidated that episodic memories that relate to spatial navigation, such as for example spatial trajectories to specific reward locations, may be encoded as sequences of so-called place cells (i.e., hippocampal cells that code for specific spatial locations along the trajectory) - in such a way that each episode corresponds to a specific sequence of place cells.

The "stored" sequences can be endogenously (re)generated in a time-compressed manner, in at least two ways. First, so-called sharp-wave ripple (SWR) sequences can be decoded when the animal is sleeping or during wakeful rest [74]. SWR sequences are expressed as part the (very fast) neural SWR complex, can span several meters [75] and are non-local, i.e., they need not (and usually not) encode the animal's current location. For example, when the animal sleeps, SWR sequences usually "replay" experienced trajectories, in a forward or a reverse direction [76]. SWR sequences have been widely implied in memory consolidation and planning [77–80].

Second, so-called theta sequences are present during active spatial navigation, when the hippocampus expresses a theta rhythm (one theta sequence for each theta cycle), see [81]. Theta sequences are typically shorter than SWR sequences and local, i.e., they include an ordered set of place cells that code for a few previous positions, the current position, and a few future positions of the animal. Theta sequences have been mainly implied in memory encoding, short-term prediction [82] and - importantly for tour discussion - the estimation of the current navigation context or episode.

Theta flickering: temporal bistability in memory

Above we discussed how the hippocampus may allow an animal reconstructing "the current episode it is experiencing" based on one or a few cues, or one or a few place cells that form an experienced sequence. However, there are cases in which a rodent enters an environment whose cues (or place cells) are compatible with more than one (episodic) memory. In these cases, multiple episodic memories "compete" for being selected. Interestingly, in the hippocampus this competitive process is solved as a series of rapid switches between two (or perhaps more) spatial representations or theta sequences - a phenomenon called "theta flickering" [83].
Specifically, in a rodent experiment in which the environmental cues were instantaneously changed (as if the animal was "teleported"), the hippocampus expressed a sort of "bistability" of episodic memories, with repeated switches between spatial representations of the past and the present environment (expressed in different theta cycles); and then settled on the correct representation (i.e., of the present environment) after a few cycles [83]. In other words, within each theta cycle, it was possible to decode a spatial representation that referred to either the previous or the current environment, but not both simultaneously (Figure 3).

From a statistical (Bayesian) viewpoint, theta flickering can be cast in terms of contextual estimation, e.g., of the context (or environment) I am currently in. In this perspective, the hippocampus encodes different predictive models, each corresponding to a particular hypothesis (e.g., I am in environment 1; I am in environment 2). The competing predictive models generate different predictions (expressed as different sequences of hippocampal cells) and the competition between them is solved when one of them collects enough evidence or minimizes prediction errors [78,84]. In other words, if the animal (or the hippocampus) is uncertain about being in environment 1 or 2, it may stochastically sample from the spatial memories of environment 1 (in one theta cycle) and then of environment 2 (at the next theta cycle); and match these sampled representations with the incoming sensory cues, thus running an error-correction mechanism that eventually converges to the correct hypothesis (e.g., I am in environment 1) when prediction errors generated by one of the two competing memories are minimized.

In sum, theta flickering in the hippocampus illustrates the possibility of having two simultaneous but conflicting memories. This can be considered a kind of temporal bistability, as the two memories are expressed sequentially (i.e., in different theta cycles), without overlaps within single theta cycles. In principle, the two memories can be extracted stochastically from a repertoire or set of memories, e.g., a distribution of (latent) memories. To understand how this may possible, below we shortly discuss recent machine learning methods that permit to sample stochastically from an internal model (or memory) of experiences.

A machine learning perspective on stochastic memories

While computational neuroscience models of memory formation are plentiful, here we focus our attention on a class of models - called generative models - that support the development of (probabilistic) internal models or "memories" of external events, which can be successively stochastically resampled. A first illustration of this capability is provided by (deep) generative models in machine learning [85], which permit to learn data in unsupervised manner (i.e., without an external supervisor providing labelled examples or reinforcement signals) and to generate (old of novel) examples of data, by resampling from their learned memories.

One of the most successful variants of this idea is the variational autoencoder or VAE [86], which essentially learns a low-dimensional internal representation or “memory” of its training data (e.g., images or videos) that affords a good reconstruction of the same data as well as the generation of novel examples of unseen data (e.g., novel images that combine parts of learned images).

The VAE is composed of two components, both usually implemented as (recurrent) neural networks, trained together using backpropagation. The fist component, the encoder, learns to compress its inputs in a low-dimensional feature vector Z. The second component, the decoder, learns to reconstruct the input using the feature vector Z, produced by the encoder – which implies that the VAE must learn (low-dimensional) Z vectors that capture essential components of its inputs. What makes the VAE particularly efficient compared to other similar methods is that during learning, it enforces both an accurate reconstruction of
the input and the acquisition of internal Z codes that are not sparse (or too spread from a prior distribution) and afford good interpolation.

The VAE permits to stochastically sample from its internal representation Z examples of data of the kind it was trained with (e.g., face images), including examples that are to some extent novel (e.g., novel faces that combine features from trained examples). This is possible because the elements of Z are (Gaussians) probability distributions (more specifically, the VAE learns two vectors, one for Z means and one for Z standard deviations), which can be stochastically sampled after learning - thus affording generative capabilities. The VAE and similar generative models are increasingly used in machine learning to generate high-quality, synthetic examples of images, videos, etc. and they provide a mechanistic illustration of (unsupervised) learning of probabilistic internal "memories" of data, which can maintain competing hypotheses (in probabilistic form) and can be stochastically resampled.

There are several other examples of computational models that incorporate similar principles (of generative modelling) as VAE, but are more concerned with biological realism. These include for example predictive coding models of visual perception [87,88], active inference models of action dynamics [89–91], and networks of spiking neurons [92–94], which may potentially shed light on biological implementations of stochastic resampling from memory, as in the case of hippocampal theta-paced flickering [78].

Memory consolidation and construction

The above computational and machine learning methods illustrate viable methods to learn (probabilistic) memories that can be stochastically resampled. Yet the memory systems of biological organisms are much more sophisticated than our current machine learning models. In biological brains, memory is plausibly a systems-level function, to which several brain areas may contribute differentially.

According to the influential “dual memories” theory, novel memories are firstly acquired by the hippocampus (whose connectivity affords rapid, one-shot learning) and successively transferred to and consolidated in cortex (where learning is slower) - with SWR sequences possibly playing a key role in such memory consolidation process [95]. The general idea is that (episodic) hippocampal memories are “replayed” multiple times, especially overnight; and the replays essentially “train” the cortex to form its (semantic) memories. Supporting this view is the fact that interrupting replays interferes with memory consolidation [96,97]. Note that a similar “experience replay” method is widely used in machine learning, to improve the robustness of neural networks during training [98]. Recent studies are increasingly showing that the interplay between hippocampal and cortical networks may be more sophisticated and bidirectional than previously believed, hence permitting the reciprocal shaping of episodic and semantic memories during consolidation [78,99].

This also implies that during consolidation, not only memories become stronger (in the sense that they are protected from interferences) but they can also change, at least partially. It has long been known that memory consolidation and retrieval – but of course also memory-based imagination [100–102] – are highly constructive processes [103–105]. One example of constructive memory is the fact that some memories can be (unconsciously) completed with information that is coherent with it, or just imagined – thus giving origins to “false memories”, e.g., the memory that Tom was at my last birthday’s party. Similarly, specific elements not initially present in an episodic memory can “intrude” it when later on, when the memory is consolidated or retrieved; perhaps to make the memory (retrospectively) coherent with novel information (e.g., if I am told that Tom was at the party); with some “schematic” information (e.g., Tom attends all parties); or with one’s self-image. These examples suggest that memory consolidation and retrieval may be better seen as
constructive, inferential processes rather than verbatim storages of events. Remarkably, this kind of consolidation and embellishment occurs in morphogenesis, too: cryptic planaria, with an altered bioelectric memory but normal 1-headed anatomy, have a ~30% chance in each recutting and regeneration round (a memory recall event) of becoming permanently 2-headed – a stronger phenotype (with a more robustly altered pattern memory, since middle fragments of 2-head worms never give rise to anything but 2-head worms). In the next section we discuss how brief changes in bioelectric state result in novel anatomical modifications which are long-term stable and in fact can get stronger over time.

Long-term bioelectric memory in planarian tissues

One of the novel aspects of our hypothesis is the claim that tissue bioelectric states can underlie long-term memory. This is unexpected, as standard models of tissue physiology treat bioelectric patterns as house-keeping parameters that quickly revert to normal levels. Thus, we tested this counter-intuitive prediction in planaria by depolarizing the somatic tissues, and examining their bioelectric state weeks later (Figure 4). Ionophore treatment is a reliable method to manipulate resting membrane voltage ($V_{\text{mem}}$) in planaria because it collapses concentration gradients of specific ions and thus modifies resting potential. We utilized the ionophore Nigericin, a specific ionophore of potassium as we've previously found transient soaks in nigericin to consistently result in significant depolarization of somatic tissue during drug treatment [64]. To assess whether induced bioelectric states could persist over longer periods of time, long after the ionophore is removed, intact *D. japonica* were treated with a 0.27 µM Nigericin with a 15mM K-Gluconate (sigma Aldrich G4500) solution for 3 days, after which the solution was washed out and samples were cultured in Poland Spring water (at 13°C to prevent fissioning). After 3 weeks in water, membrane polarization was quantified using the voltage-sensitive dye DiBAC$_4$(3) [64,65,106]. Relative membrane polarization was tracked at 2 different time points: worms which had been removed from Nigericin and were put in water for 1 week (1 week washout), worms which had been removed from Nigericin 3 weeks prior to DiBAC imaging (“3 week washout”) and a wild-type control group which had never been drug-treated. Both of the Nigericin treated group were significantly depolarized relative to controls, (**P<0.01 Students T-test), while there were no significant differences between the treated samples. Note also that the altered pattern gets stronger between 1 week and 3 weeks (with no further treatment), as discussed above in the context of consolidation and construction of memories in the brain. These data reveal the remarkable fact that very brief alterations of bioelectric mechanics are remembered by tissue as long-term stable alterations in resting membrane potential in planaria.

Conclusion

Regeneration highlights numerous aspects of robust control that are challenging to explain via molecular-genetic pathway paradigms. These include: the ability of cellular systems to coordinate activity across large distances, to process spatial information at multiple scales, including those much bigger than single cells, to integrate stochastic and noisy behaviour of micro-agents into organ-level decisions, to store "potential", future, or "counterfactual" pattern memories which may be latent (not expressed as anatomy now, but guiding future anatomical events if recalled by injury) in addition to ones that are expressed (as control of morphogenetic mechanisms in the current anatomy), and to organize activity that works toward specific invariant outcomes (anatomical specifications) despite perturbations. Decades of work in neuroscience provide a rich background from which to seek novel perspectives on this problem. Importantly, the similarity between decision-making mediated by neuronal tissue and during active morphogenesis is only a functional metaphor, but also reflects profound molecular conservation of mechanisms, as neural
networks evolved from ancient pre-neural bioelectric mechanisms that were guiding morphogenesis long before they were sped up and harnessed to the control of motile organism-level behaviour.

Thus, we suggest that this deep symmetry between control of cell behaviour by pre-neural networks and control of animal behaviour by the nervous system can drive a rich experimental research program. Just some of the future opportunities include: the use of experiences and sensory data to train tissues and organs to specific morphogenetic outcomes via behaviour-shaping paradigms, the use of pharmacological tools targeting brain memory pathways to modify stored pattern memories (and thus anatomy) in vivo, the extension of neural network modelling tools and content extraction pipelines (used in neural decoding) to understand the anatomical states encoded in developmental bioelectric prepatterns and develop strategies for reading and writing pattern memories (as well as constructing synthetic tissues capable of holding and acting on such memories), and the use of psychopharmacological agents (neurotransmitter drugs, hallucinogens, etc.) to modify the perception and processing of structural “sensory” data as cells and tissues ascertain the current anatomical layout of the body. It is likely that concepts from computational psychiatry [107–109] may have a lot of bearing on the systems-level causes in complex disease states like cancer [110,111] and birth defects [112,113]. The consilience of tools and conceptual approaches between the study of cognitive memory and the efforts to control growth and form is not only likely to transform not only regenerative medicine and synthetic bioengineering, but also perhaps will help cognitive neuroscience understand how computational agents (proto-minds) arise from cellular structure.

Acknowledgements:

We thank many people in the neuroscience, AI, and regenerative biology communities for many helpful discussions. M.L. is supported by the Paul G. Allen Frontiers Group (via an Allen Discovery Center Award 12171), as well as the Templeton World Charity Foundation (WCF0089/AB55 and TWCF0140). This research has received funding from the European Union’s Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2 to GP)
References


Figure Legends:

(A) Planarian mid-body fragments after head and tail amputation regenerate into 2-headed worms if exposed to octanol (8OH), a blocker of the electrical synapses known as gap junctions (GJ). Remarkably, such worms go on to regenerate as 2-headed worms in future rounds of amputation, in plain water, with no further treatment [54]. These animals can be re-set back to normal by exposure to the proton-potassium ion pump blocker SCH28080, because the pattern memory is stored via a stable state of the bioelectric circuit [59, 65]. Together these data show how the target morphology (the anatomy to which an animal will regenerate) can be permanently re-written without genomic editing [51].

(B) The penetrance of the gap junction blocker is about 30% - the rest of the animals regenerate as 1-headed. However, they are not normal – recutting them in water results in the same 30%/70% proportion of 2-headed worms and worms with a de-stabilized (stochastic) anatomy, as opposed to truly wild-type worms’ 100% normal (1-headed) regeneration [59]. This stochastic decision is made independently, as each piece of a worm can have an independent headedness fate. Panel (C) shows the transition diagram for the bioelectric pattern memory editing: a wild-type worm treated with GJ blocker can become Cryptic (destabilized) or 2-headed. The Cryptic state can perpetuate but once a worm becomes 2-headed, it always gives rise to 2-headed worms. Both Cryptic and 2-headed worms can be re-set to normal by SCH28080 treatment. Abbreviations: wt = wild-type; 8-OH = octanol; DH = double-headed; CRPT = cryptic (destabilized target morphology).

Figure 1. Permanent and stochastic editing of pattern memories in planaria

Figure 2. Bistability in perceptual systems
Bistability (or more generally, multistability) during perception refers to (apparently random) spontaneous subjective fluctuations between competing interpretations of ambiguous perceptual (e.g., visual) stimuli. Some examples are (A) the Necker cube illusion, which can be interpreted to have either the lower-left or the upper-right square as its front side; and (B) the duck-rabbit illusion, which can be interpreted either as a duck or a rabbit. Note that the two competing hypotheses are never considered (or perceived) simultaneously.

Figure 3. Schematic illustration of theta flickering
Rats are trained for several days in two boxes (A and B) having different cues and develop uncorrelated neural populations in hippocampal area CA3 that fire preferentially in either A or B. After training, rats are "teleported": they start in one environment (e.g., box A) but the cues are instantaneously switched to those of the other environment (e.g., box B). After "teleporting", distinct neural populations tend to be active at each theta cycle, which fire preferentially in A or B. Example from [83] showing that neural populations code for spatial locations coherent with the actual animal position (indicated by +) in either the correct box B or the incorrect box A, at different theta cycles. Extract of Figure 3 of [83], authorization pending. See [83] for details.

Figure 4: Bioelectrical long-term memory in planaria
Voltage-sensitive fluorescent dye reveals the spatial distribution of bioelectric states in intact planaria [66]. Our model, of the bioelectric pattern as a memory, predicts that after editing this pattern, it should remain altered for long periods of time (relative to the normal rate of change of bioelectric parameters, which is milliseconds for the CNS and minutes-hours in developmental bioelectricity). Control animals show depolarization at the anterior end (A'). After soaking in ionophore, which induces a shift to Cryptic phenotype, both ends become relatively depolarized and stay that way for 1 week (A''). Remarkably, even 3 weeks later, this altered pattern persists and indeed becomes stronger, as discussed above in the context of consolidation and construction of memories over time (A''').

Quantification of the altered bioelectric states showing significant differences persisting at 1 week and 3 weeks after exposure to ionophore.