Modular cognition

Powerful tricks from computer science and cybernetics show how evolution 'hacked' its way to intelligence from the bottom up

by Michael Levin & Rafael Yuste

Michael Levin is the Vannevar Bush Chair and Distinguished Professor of Biology at Tufts University in Massachusetts, where he directs the Allen Discovery Center and the Tufts Center for Regenerative and Developmental Biology.

Rafael Yuste is professor of biological sciences and neuroscience, co-director of the Kavil Institute for Brain Science, director of the NeuroTechnology Center and chair of the NeuroRights Foundation at Columbia University.

Intelligent decision-making doesn’t require a brain. You were capable of it before you even had one. Beginning life as a single fertilised egg, you divided and became a mass of genetically identical cells. They chattered among themselves to fashion a complex anatomical structure – your body. Even more remarkably, if you had split in two as an embryo, each half would have been able to replace what was missing, leaving you as one of two identical (monozygotic) twins. Likewise, if two mouse embryos are mushed together like a snowball, a single, normal mouse results. Just how do these embryos know what to do? We have no technology yet that has this degree of plasticity – recognising a deviation from the normal course of events and responding to achieve the same outcome overall.

This is intelligence in action: the ability to reach a particular goal or solve a problem by undertaking new steps in the face of changing circumstances. It’s evident not just in intelligent people and mammals and birds and cephalopods, but also cells and tissues, individual neurons and networks of neurons, viruses, ribosomes and RNA fragments, down to motor proteins and molecular networks. Across all these scales, living things solve problems and achieve goals by flexibly navigating different spaces – metabolic, physiological, genetic, cognitive, behavioural.

But how did intelligence emerge in biology? The question has preoccupied scientists since Charles Darwin, but it remains unanswered. The processes of intelligence are so intricate, so multilayered and baroque, no wonder some people might be tempted by stories about a top-down Creator. But we know evolution must have been able to come up with intelligence on its own, from the bottom up.

Darwin’s best shot at an explanation was that random mutations changed and rearranged genes, altered the structure and function of bodies, and so produced adaptations that allowed certain organisms to thrive and reproduce in their environment. (In technical terms, they are *selected* for by the environment.) In the end, somehow, intelligence was the result. But there’s plenty of natural and experimental evidence to suggest that evolution doesn’t just select hardwired solutions that are engineered for a specific setting. For example, lab studies have shown that perfectly normal frog skin cells, when liberated from the instructive influence of the rest of the embryo, can reboot their cooperative activity to produce a novel proto-organism, called a ‘xenobot’. Evolution, it seems, doesn’t come up with answers so much as generate flexible problem-solving agents that can rise to new challenges and figure things out on their own.

The urgency of understanding intelligence in biological terms has become more acute with the 'omics' revolution, where new techniques are amassing enormous amounts of fresh data on the genes, proteins and connections within each cell. Yet the deluge of information about cellular hardware isn't yielding a better explanation of the intelligent flexibility we observe in living systems. Nor is it yielding sufficient practical insights, for example, in the realm of regenerative medicine. We think the real problem is not one of data, but of perspective. Intelligence is not something that happened at the tail end of evolution, but was discovered towards the beginning, long before brains came on the scene.

From the earliest metabolic cycles that kept microbes' chemical parameters within the right ranges, biology has been capable of achieving aims. Yet generation after generation of biologists have been trained to avoid questions about the ultimate purpose of things. Biologists are told to focus on the 'how', not the 'why', or risk falling prey to theology. Students must reduce events to their simplest components and causes, and study these mechanisms in piecemeal fashion. Talk of 'goals', we are told, skirts perilously close to abandoning naturalism; the result is a kind of 'teleophobia', a fear of purpose, based on the idea that attributing too much intelligence to a system is the worst mistake you can make.

But the converse is just as bad: failing to recognise intelligence when it's right under our noses, and could be useful. Not only is 'why' always present in biological systems – it is exactly what drives the 'how'. Once we open ourselves up to that idea, we can identify two powerful tricks, inspired by computer science and cybernetics, that allowed evolution to 'hack' its way to intelligence from the bottom up. No skyhooks needed.

Embryos aren't the only things capable of flexible self-repair. Many species can regenerate or replace lost body parts as adults. The Mexican salamander known as the axolotl can regrow lost limbs, eyes, jaws and ovaries, as well as the spinal cord and portions of the heart and brain. The body recognises deviations from its correct anatomy as errors, and the cells work rapidly to get back to normal. Similarly, when Picasso-style tadpoles are made in the lab, with eyes and other organs randomly placed in different starting positions, they undergo novel paths of movement that nevertheless end up creating largely normal frog faces.

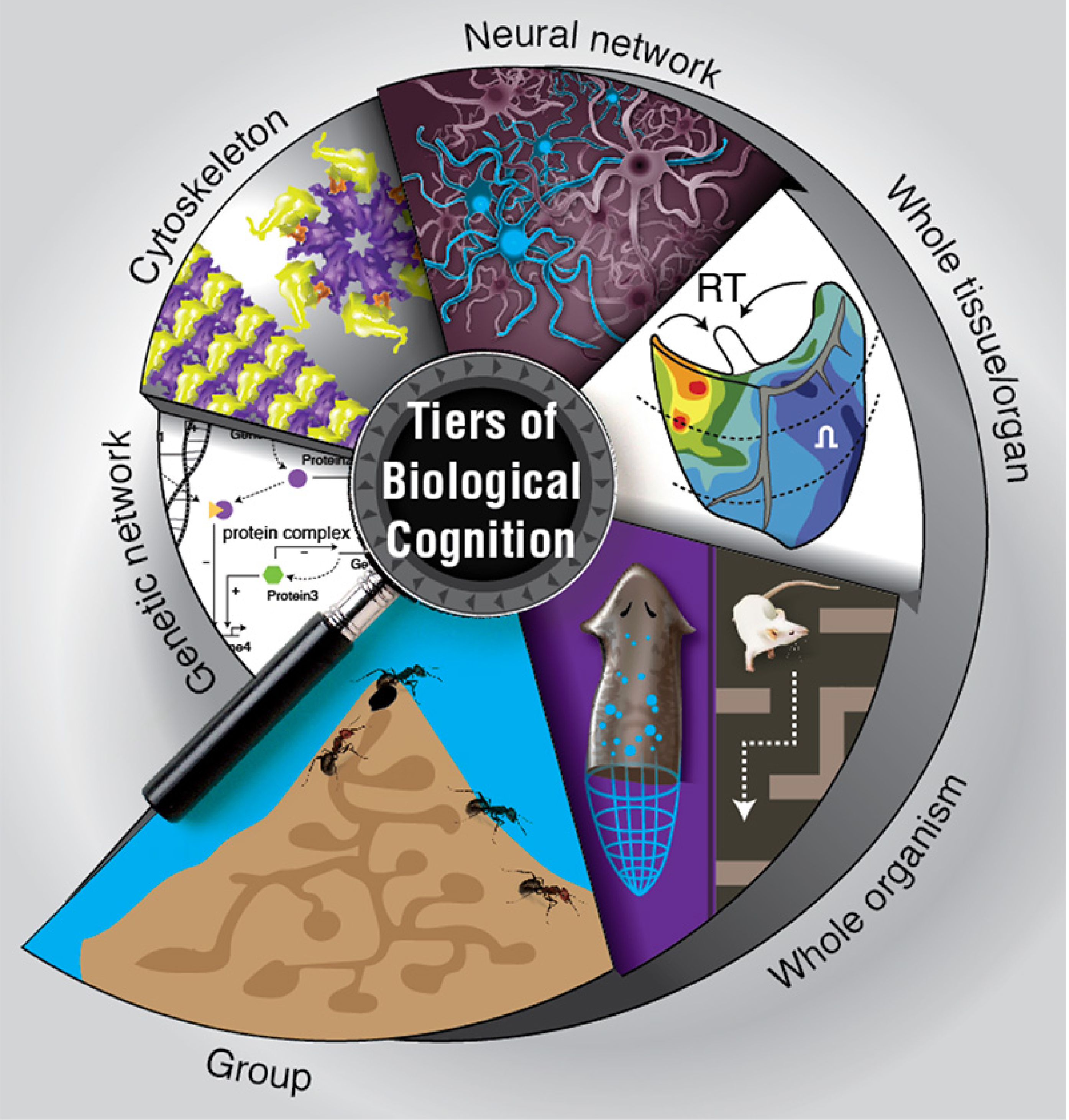
How do the tadpole's facial organs know when to stop moving? How does the salamander's tissue determine that a limb of the right size and shape has been produced, and that the remodelling can stop? It's clear that cell groups need to be able to 'work until goal X is satisfied', storing a memory of the goal (which might be an anatomical configuration much larger than any single cell), and course-correcting if they are perturbed along the way.

One explanation of embryos' amazing feats comes from control theory, which show us that dynamical systems (such as a thermostat) can pursue goals without magic, simply using feedback to correct for errors. Such 'homeostasis' in biology, as the process is known, keeps parameters such as pH within specific limits. But the same dynamic operates on a much larger scale, as long as cells are physiologically plugged into networks. Networks of cells are able to measure chemical, electrical and biomechanical properties of tissues, and make decisions about things that single cells cannot appreciate.

Single cells can process local information about their environment and their own states to pursue tiny, cell-level goals (such as staying pointed in a specific direction). But networks of cells can integrate signals from across distances, store memories of patterns, and compute the outcomes of large-scale questions (such as 'Is this finger the right length?' or 'Does this face look right?') Cell networks carry out computations that can assess larger quantities (such as anatomical shape) and direct the underlying cell activity to bring the system closer to a specific desired goal (called a 'setpoint').

> Modularity means that the stakes for testing out mutations are reasonably low

Achieving such intelligent, finely tuned calibration likely relies on *modularity* – the first step that we believe can explain the emergence of intelligent behaviour. Like a large organisation that deploys a number of specialised teams to make and sell a single product, modularity is about self-maintaining units that can cooperate or compete to achieve local outcomes, but end up collectively working towards some larger goal. Crucially, this structure avoids micromanagement – each level doesn't need to know how the lower levels do their job, but can simply motivate them (with things like reward molecules and stress pathways) to get it done.



Modularity - the presence of competent subunits, which solve problems in their own local problem space, that can cooperate and compete to achieve bigger goals - is part of what enables the emergence of intelligence in biology. The way these modules’ agendas are nested within one another in biological networks gives them the flexibility to meet goals at each level, even when conditions change at lower levels

When unicellular organisms joined up to make multicellular bodies, each module didn’t lose its individual competency. Rather, cells used specific proteins to merge into ever more complex networks that could implement larger objectives, possess longer memories and look further into the future. Networks of cells began to work as a society – measuring and pursuing goals defined at the level of the collective (such as 'organ size' and 'organ shape'). Stress about large-scale states (such as 'attention: finger too short') triggered change, which was shared across tissues to implement coordinated action. This multiscale architecture has many advantages. For one, it's easy for evolution to simply shift the modules around and let the cycle of error reduction take care of the rest – setting new conditions as the setpoint, or the source of stress (eg, 'wrong length of limbs' instead of 'misfolded proteins'). It's similar to how you can change the setpoint of your thermostat without having to rewire it, or even know how it works. Feedback loops within feedback loops, and a nested hierarchy of incentivised modules that can be reshuffled by evolution, offer immense problem-solving power.

One implication of this hierarchy of homeostatically stable, nested modules is that organisms became much more flexible while still maintaining a coherent 'self' in a hostile world. Evolution didn't have to tweak everything at once in response to a new threat, because biological subunits were primed to find novel ways of compensating for changes and functioning within altered systems. For example, in planarian flatworms, which reliably regenerate every part of the body, using drugs to shift the bioelectrically stored pattern memory results in two-headed worms. Remarkably, fragments of these worms continue to regenerate two heads in perpetuity, without editing the genome. Moreover, flatworms can be induced, by brief modulation of the bioelectric circuit, to regrow heads with shape (and brain structure) appropriate to other known species of flatworms (at about 100 million years of evolutionary distance), despite their wild-type genome.

Modularity means that the stakes for testing out mutations are reasonably low: competent subunits can be depended upon to meet their goals under a wide range of conditions, so evolution rarely needs to 'worry' that a single mutation could ruin the show. For example, if a new mutation results in an eye being in the wrong place, a hardwired organism would find it very hard to survive. However, modular systems can compensate for the change while moving the eye back to where it's supposed to be (or enabling it to work in its new location), thus having the opportunity to explore other, possibly useful, effects of the mutation. Tadpole eyes have been shown to do this, conferring vision even if asked to form on the tail, by finding connections to the spinal cord instead of the brain.

Modularity provides stability and robustness, and is the first part of the answer to how intelligence arose. When changes occur to one part of the body, its evolutionary history as a nested doll of competent, problem-solving cells means subunits can step up and modify their activity to keep the organism alive. This isn't a separate capacity that evolved from scratch in complex organisms, but instead an inevitable consequence of the ancient ability of cells to look after themselves and the networks of which they form a part.

But just how are these modules controlled? The second step on the road to the emergence of intelligence lies in knowing how modules can be manipulated. Encoding information in networks requires the ability to catalyse complex outcomes with simple signals. This is known as *pattern completion*: the capacity of one particular element in the module to activate the entire module. That special element, which serves as a 'trigger', starts the activity, kicking the other members of the module into action and completing the pattern. In this way, instead of activating the entire module, evolution needs only to activate that trigger.

Pattern completion is an essential aspect of modularity which we're just beginning to understand, thanks to work in developmental biology and neuroscience. For example, an entire eye can be created in the gut of a frog embryo by briefly altering the bioelectric state of some cells. These cells are triggered to complete the eye pattern by recruiting nearby neighbours (which were not themselves bioelectrically altered) to fill in the rest of the eye. Similar outcomes can be achieved by genetic or chemical 'master regulators', such as the *Hox* genes that specify the body plan of most bilaterally symmetrical animals. In fact, one could relabel these regulator genes as pattern completion genes, since they enable the coordinated expression of a suite of other genes from a simple signal. The key is that modules, by continuing to work until certain conditions are met, can fill in a complex pattern when given only a small part of the pattern. In doing so, they translate a simple command – the activation of the trigger – and amplify it into an entire program.

Crucially, pattern completion doesn't require defining all of the information needed to create an organ. Evolution does not have to rediscover how to specify all the cell types and get them arranged in the correct orientation – all it has to do is activate a simple trigger, and the modular organisation of development (where cells build to a specific pattern) does the rest. Pattern completion enables the emergence of developmental complexity and intelligence: simple triggers of complex cascades that make it possible for random changes to DNA to generate coherent, functional (and occasionally advantageous) bodies.

Modular pattern completion is also becoming evident in recent experiments in neuroscience. The pinnacle of intelligent behaviour in biology is the human brain. Nervous systems are built with large numbers of neurons, where every neuron is typically connected to very large numbers of other neurons. Over evolutionary time, there's been a steady progression to larger and more connected brains, reaching astronomical numbers – with close to 100 billion neurons and hundreds of thousands of connections per neuron in humans. This move towards higher numbers of neurons and connections cannot be a coincidence: a system with many interacting units is exactly what it takes to become more competent and complex.

> Analysing the brain by looking at a single neuron is like trying to understand a movie from an isolated pixel

But what exactly do all these vast neural circuits do? While many neuroscientists would agree that the function of the nervous system is to sense the environment and generate behaviour, it's less clear how that actually happens. The traditional view known as the 'neuron doctrine', proposed by Santiago Ramón y Cajal and Charles Scott Sherrington more than a century ago, is that each neuron has a specific function. This would make the brain analogous to an aeroplane, built with millions of components, each precisely designed for a particular task.

Within this framing, neuroscientists have teased apart the brain and studied it one neuron at a time, linking the activity of individual neurons to the behaviour of the animal or the mental state of a person. Yet if the true goals of biological systems derive from how their subunits or modules interact, then analysing the brain by looking at a single neuron is as futile as trying to understand a movie while fixating on an isolated pixel.

What type of properties might neural circuits generate? Because neurons can set each other off, neural circuits can generate internal states of activity that are independent of the outside world. A group of connected neurons could auto-excite each other and become active together for a period of time, even if nothing external is happening. This is how we might understand the existence of the concepts and abstractions that populate the human mind – as the endogenous activity of modules, made up of ensembles of neurons.

Using those intrinsic activity states as symbols, evolution could then build formal representations of reality. It could manipulate those states instead of manipulating the reality, just as we do by defining mathematical terms to explore relationships between objects. From this point of view – the gist of which was already proposed by Immanuel Kant in his *Critique of Pure Reason* (1781, 1787) – the evolution of the nervous system represents the appearance of a new formal world, a symbolic world, that greatly expands the possibilities of the physical world because it gives us a way to explore and manipulate it mentally.

Neuronal modules could also be organised in a hierarchy, where higher-level modules encode and symbolise increasingly more abstract entities. For example, lower-level groups of neurons in our spinal cord might activate muscle fibres, and be under the control of upper-level ensembles in the motor cortex that could encode the desired movement more abstractly ('to change the position of the leg'). In turn, these motor cortex ensembles could be controlled by higher-order neurons again ('to perform a pirouette'), which could be controlled by groups of neurons in the prefrontal cortex that represent the behavioural intention ('to perform in a ballet').

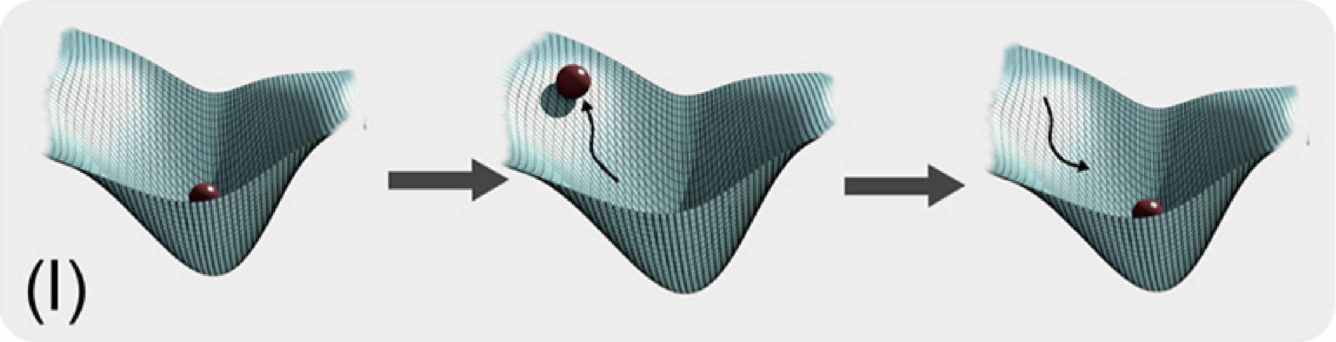
Using modules nested in a hierarchy provides a neat solution to a tough design challenge: instead of specifying and controlling every element, one at a time, nature uses neuronal ensembles as computational building blocks to perform different functions at different levels. This progression towards increasing abstraction could help explain how cognition and consciousness might arise, as emergent functional properties, from relatively simple neural hardware. This same powerful idea of hierarchical emergence is behind the layered neural network models in computer science, named 'neural' because they are inspired by neural circuits.

But back to Darwin's problem: if evolution is blind and acting solely on the individual units, one mutation at a time, how can the overall architecture and function of an organism be modified for the common good? Besides generating modules, neural networks have another interesting property we've already discussed: pattern completion.

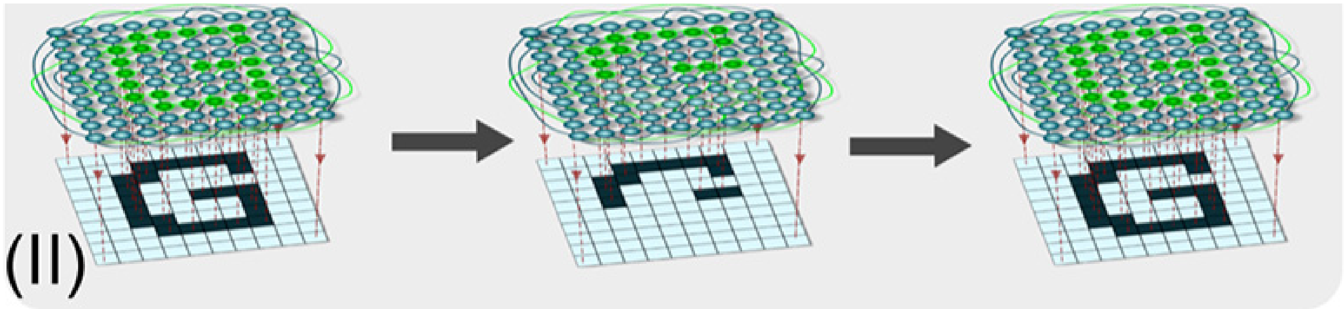
In recent experiments, mice were induced to have artificial perceptions or visual hallucinations by activating only two neurons in their visual cortex. How is this possible, given that the mouse brain has around 100 million neurons? The answer is that those neurons can trigger a neuronal ensemble, via pattern completion. Neurons' connectivity seems to amplify activity, so, like an avalanche, a change in one neuron ends up triggering the entire module. This means that you can activate an entire module of neurons by turning on only one key member of the group.

> Pattern completion shows us how a single event – say, a mutation – can change an army, or build an eye

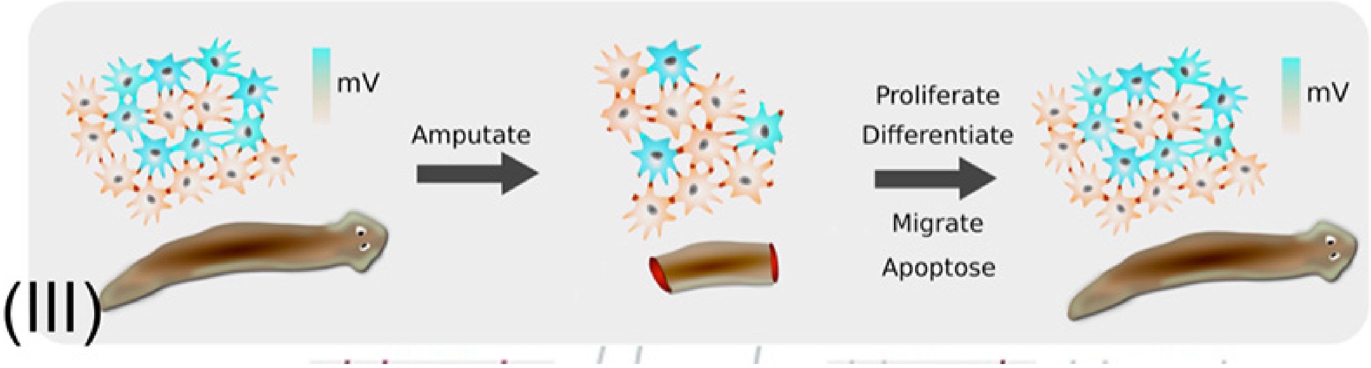
Pattern completion could be at the core of how the brain works internally – recruiting module after module, at different levels of the hierarchy, depending on the task at hand. But why wouldn't pattern completion end up tripping the entire brain into an epileptic seizure? By adding inhibitory connections to these neural circuits – small circuit breakers – one can restrict these avalanches to small groups of neurons, instead of catastrophically activating the entire brain. By harnessing pattern completion along with inhibitory circuits, the brain has the ability to select and manipulate modules at different levels as needed.



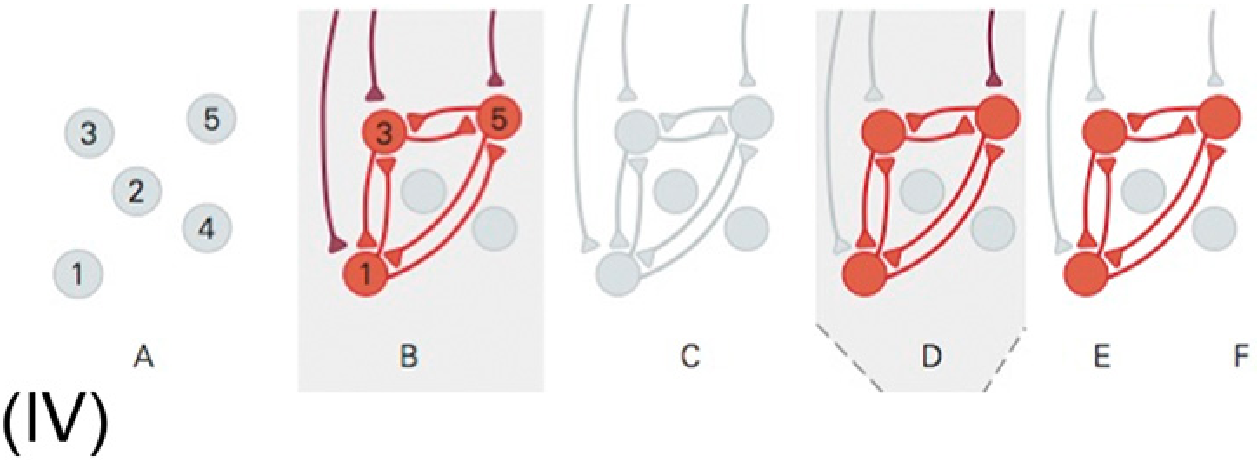
(I) Trigger stimuli allow for pattern completion, because networks can be organised such that they tend to settle in specific states (memories) from diverse positions (like the way a ball rolls down into a well from many different starting points). The topology of this landscape means that the system automatically returns to the same state when disturbed - symbolised by the ball falling to the bottom of a valley, when it is placed near the edge. This metaphor for pattern completion captures the idea of how the 'landscape' of the system 'makes you do it': it allows subunits to pursue local homeostatic goals (such as minimising a variable to move down a gradient), but enables the end result to be a higher-level pattern



(II) The ability for networks to reach the same outcome from a range of partial inputs - the generalisation of pattern completion - is also exploited in computer science. For example, computational neural networks can recover an entire remembered image based on only a partial example, from which some aspect has been removed



(III) The flatworm is a master of pattern completion, generating its entire anatomy from a small piece of its pattern. The bioelectric network of cells stores a pattern memory that controls individual cells in order to restore the whole



(IV) Pattern completion is also present in neural circuits, where a small group of connected neurons can store associative memory. Here, you see: a) independent neurons are present without synaptic connections; b) neurons 1, 3 and 5 are activated simultaneously by external inputs, which forms and strengthens the synaptic connections between them; c) when the input ceases, neuronal activity also stops (however, the synaptic connections between the three neurons remain; these neurons have formed a module, and their interconnectivity determines how the module activates); d) an input activates just one of the original three neurons, but the connections activate all three neurons and complete the entire pattern; e) even after the input current has ended, the neurons can remain persistently active, in effect, storing a memory of the input

In this way, pattern completion enables connections between modules at the same and different levels of the hierarchy, knitting them together as a single system. A key neuron in a lower-level module can be activated by an upper-level one, and vice versa. Like changing the march of an army, you don't need to convince every soldier to do so – just convince the general, who makes the others fall into line. Consistent with the many parallels between neurons and non-neural signals, pattern completion shows us how a single event – say, a mutation – can change an army, or build an eye.

From microbe cells solving problems in metabolic space, to tissues solving problems in anatomical space, to groups of people navigating the world as we know it, life has ratcheted towards intelligent designs by exploiting the ability of modules to get things done, in their own way. Homeostatic loops provide flexible responses, working until setpoints are achieved, even when things change. Modularity means that evolution can readily explore what the collective considers to be a 'correct' condition, and what actions it takes to get there. Hierarchies of modules mean that simple signals can trigger complex actions that don't need to be rediscovered or micromanaged, yet can adapt when only a small piece of the puzzle triggers them to do so.

We have sketched a set of approaches to biology that rely heavily on concepts from cybernetics, computer science, and engineering. But there's still a lot of work to do in reconciling these approaches. Despite recent advances in molecular genetics, our understanding of the mapping between the genome on the one hand, and the (changeable) anatomy and physiology of the body on the other, is still at a very early stage. Much like computer science, which moved from rewiring hardware in the 1940s to a focus on algorithms and software that could control the device's behaviour, biological sciences now need to change tack.

We call on biologists to embrace the intentional stance: treating circuits and cells as problem-solving agents

The impact of understanding nested intelligence across multiple scales cuts across numerous fields, from fundamental questions about our evolutionary origins to practical roadmaps for AI, regenerative medicine and biorobotics. Understanding the control systems implemented in living tissue could lead to major advances in biomedicine. If we truly grasp how to control the setpoints of bodies, we might be able to repair birth defects, induce regeneration of organs, and perhaps even defeat ageing (some cnidarians and planarian flatworms are essentially immortal, demonstrating that complex organisms without a lifespan limit are possible, using the same types of cells of which we are made). Perhaps cancer can also be addressed as a disease of modularity: the mechanisms by which body cells cooperate can occasionally break down, leading to a reversion of cells to their unicellular past – a more selfish mode in which they treat the rest of the body as an environment within which they reproduce maximally.

In the field of engineering, designers have traditionally built robots from dumb but reliable parts. By contrast, biology exploits the unreliability of components, making the most of the competency of each level (molecular, cellular, tissue, organ, organism, and colony) to look after itself. This enables an incredible range of adaptive plasticity. If we break the neuronal code in biology, we could begin to program behaviour into synthetic nervous systems, and build self-repairing, flexible robots. Recent work shows that agential cells with their own local agendas can already be guided to create entirely new, autonomous biorobots. And beyond robots’ bodies, these ideas also open up new approaches to machine learning and AI: raising prospects for architectures based on ancient and diverse problem-solving ensembles beyond brains, such as those of bacteria and metazoa.

This emerging confluence of developmental biology, neuroscience, biophysics, computer science and cognitive science could have profound and potentially transformative applications. Top-down strategies exploiting – in effect, collaborating with – biology’s native intelligence could enable transformative progress in areas oppressed by a narrow focus on molecular and genetic detail. With that in mind, we call on biologists to embrace the intentional stance: treating circuits, cells and cellular processes as competent problem-solving agents with agendas, and the capacity to detect and store information – no longer a metaphor, but a serious hypothesis made plausible by the emergence of intelligent behaviour in phylogenetic history. If we can come to recognise intelligence in its most unfamiliar guises, it might just revolutionise our understanding of the natural world and our very nature as cognitive beings.

aeon.co 8 March 2022