The collective intelligence of evolution and development

Richard Watson
Electronics and Computer Science/Institute for Life Sciences, University of Southampton, Southampton, UK

Michael Levin [ORCID]
Allen Discovery Center at Tufts University, Medford, MA, USA

Wyss Institute for Biologically Inspired Engineering at Harvard University, Cambridge, MA, USA

Abstract
Collective intelligence and individual intelligence are usually considered to be fundamentally different. Individual intelligence is uncontroversial. It occurs in organisms with special neural machinery, evolved by natural selection to enable cognitive and learning functions that serve the fitness benefit of the organism, and then trained through lifetime experience to maximise individual rewards. Whilst the mechanisms of individual intelligence are not fully understood, good models exist for many aspects of individual cognition and learning. Collective intelligence, in contrast, is a much more ambiguous idea. What exactly constitutes collective intelligence is often vague, and the mechanisms that might enable it are frequently domain-specific. These cannot be mechanisms selected specifically for the purpose of collective intelligence because collectives are not (except in special circumstances) evolutionary units, and it is not clear that collectives can learn the way individual intelligences do since they are not a singular locus of rewards and benefits. Here, we use examples from evolution and developmental morphogenesis to argue that these apparent distinctions are not as categorical as they appear. Breaking down such distinctions enables us to borrow from and expand existing models of individual cognition and learning as a framework for collective intelligence, in particular connectionist models familiar in the context of neural networks. We discuss how specific features of these models inform the necessary and sufficient conditions for collective intelligence, and identify current knowledge gaps as opportunities for future research.

Keywords
Evolution, machine learning, networks, individuality, neural networks

Introduction

The identification of a suitable theoretical framework and appropriate engineering principles for collective intelligence are open problems. In this paper, we begin to address these gaps by developing a synthesis of perspectives usually considered to be quite distinct. To do this, we first dissolve a number of limiting misconceptions that cause collective intelligence and individual intelligence to be treated as separate topics; second, we introduce a speculative conceptual framework to unify them.

For an intelligence to belong properly to a collective, it must arise not from the cleverness of its members but from having the right kind of functional relationships between them. What kinds of functional relationships, and in what specific organisation, are required to turn a collective that is

Corresponding author:
Michael Levin, Tufts University, 200 Boston Ave., Medford, MA, USA. Email: Michael.Levin@tufts.edu

not intelligent into a collective that is? We use a specific understanding of cognition and learning that is already well-developed for individual intelligence to synthesise collective intelligence with aspects of development and evolution. In particular, we explore how connectionist models of cognition and learning, familiar in neural network models of *individual* intelligence, can address this question, and how they signpost directions for future research in *collective* intelligence. We especially emphasise the known emergent properties of cellular collectives as instructive examples of collective intelligence at a sub-organismal scale.

Individual and collective intelligence are distinct phenomena. Or are they?

At first glance, it might seem that models of individual intelligence are not relevant to collective intelligence. Individuals have brains that can cognise and learn, and although colonies and swarms might be composed of individuals with brains, the collective *as a whole* is not a brain and cannot cognise or learn. Moreover, it is easy to understand why the component parts of an individual work together so well because adaptive processes at the organismic level, such as evolution by natural selection and reward-based reinforcement, select or reward them for doing so. In contrast, collectives are composed of multiple evolutionary units or distributed multi-agent systems and thus present unique credit-assignment problems that complicate reinforcement of such adaptive processes. Such distinctions seem to justify the consideration of collective and organismic intelligences as different topics. We argue that these are false distinctions and there is a bigger, and much more interesting, picture. The basic tenets of this unified view are the following:
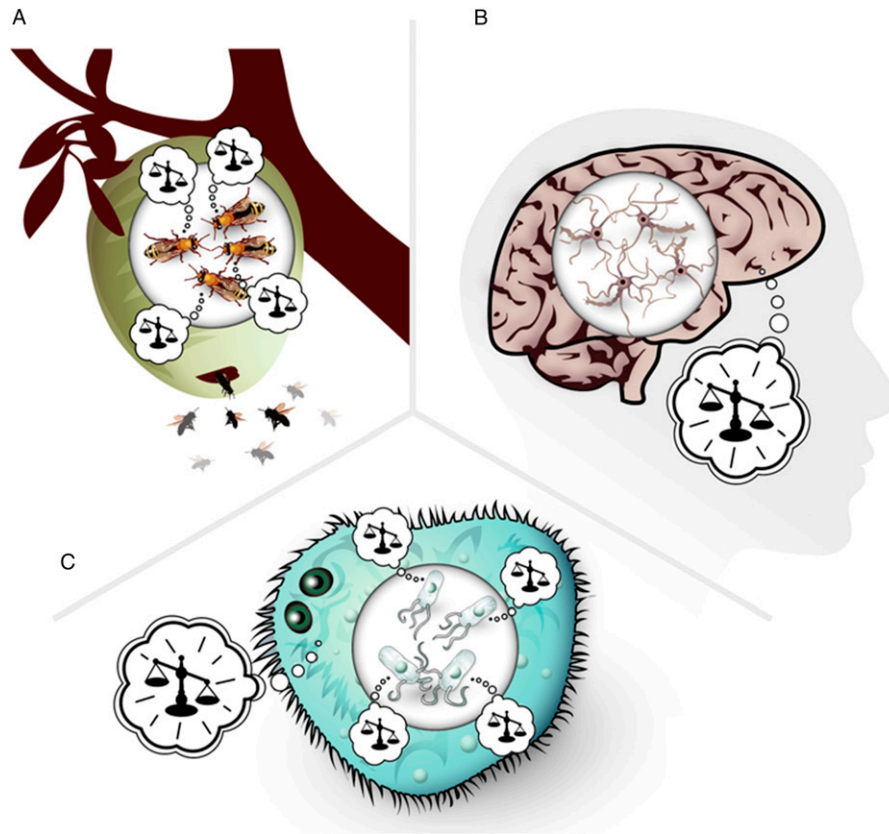
All individuals are collectives. All individuals are collectives, made of parts that used to be individuals themselves. This is true not only for multicellular organisms derived from unicellular ancestors but also for eukaryotic cells with multiple organelles arising from bacterial ancestors, and for simpler cells that contain the first chromosomes arising from the union of previously free-living self-replicating molecules (Godfrey-Smith, 2009; Maynard Smith and Szathmáry, 1997; Michod, 2000; Okasha, 2006; West et al., 2015). Moreover, the proper functioning of organisms – their robustness, adaptability and evolvability – depends on the continued autonomy of their component parts (Levin, 2019; 2021a). Multicellular organisms exhibit multi-scale autonomy, a dynamic interplay of competition and cooperation, and co-ordinated collective action inherent to their development, function and behaviour, while being a society of cells (Fields and Levin, 2022; Levin, 2019; 2022; 2023; Sonnenschein and Soto, 1999). Thus, individuals like you and I, and collectives

like swarms and colonies, are not as categorically different as they first appear.

All intelligences are collectives. Individual intelligence, in the familiar guise of a central nervous system or a brain, arises from the interaction of many unintelligent components (neurons) arranged in the right organisation with the right connections. This is the foundation of *connectionism*; that intelligence resides not in the individual parts but in the arrangement of the connections between them (LeCun et al., 2015; Watson et al., 2016; Watson and Szathmáry, 2016). The individual neuron is *not* where all the interesting cognition and learning occur. It is the distributed collective activity in the network that constitutes cognition and changes to the organisation of network connections that constitute learning. So brains are collectives, thus collectives *of the right kind* do cognise and learn. In fact, brains provide the archetypal example of an intelligent collective.

Cognition and learning are substrate-independent. The principles of distributed cognition familiar in artificial neural networks can be implemented by any network of signals and non-linear responses to suitably weighted inputs (Evans et al., 2022; Stern and Murugan, 2022; Watson et al., 2016). Gene-regulation networks, ecological networks and social networks can all compute in the same sense as neural networks if the connections are suitably arranged (Biswas et al., 2021; Davies et al., 2011; Herrera-Delgado et al., 2018; Power et al., 2015; Szabó et al., 2012; Tareen and Kinney, 2020; Watson et al., 2014). In development and organismic biology, many different levels of adaptive networks exist aside from neural networks, including gene-regulation networks, protein networks, metabolic networks, morphogen diffusion networks and endocrine systems. In addition, it is clear that morphogenesis, physiological function and the adaptive processes of robustness and repair all require information integration and collective action that constitute cognition – in many cases without neurons. Each of these phenomena exhibits the same learning behaviours, including the storage and retrieval of multiple associative memories, effecting classification and recognition with generalisation capabilities, and learning to solve combinatorial optimisation problems better with experience (Watson et al., 2011a; Watson et al., 2011b, 2011c).

The credit assignment problems inherent in collective intelligence are fundamental in all cognition and learning, and in all biological individuality. It is true that collective intelligence is fundamentally about *collectives* – meaning that we *cannot* presuppose the system as a whole to be a single selective or utility-maximising unit. However, when we take a larger perspective – for example, one concerned with their emergence over developmental or evolutionary timescales – neither can we presuppose that apparently

Figure 1. Perspectives on individual intelligence and collective intelligence. Complex systems are composed of many interacting components. But where is the agent – at the component level or the system level? (a) Swarms are often characterised as collectives, but the agency (reward feedback and decision making) is generally attributed to each component. These are obviously collectives but not obviously intelligent. (b) Animal intelligence is often characterised as a single, system-level agent (exhibiting information integration and collective action), but the components are generally considered to be 'parts' without agency. These systems are obviously intelligent but not obviously collectives. (c) In reality, all intelligences are made out of components that act on local information based on individual feedbacks. In a multicellular organism, for example, the individual cells exhibit agency based on local information and rewards, and the system (cellular swarm) as a whole does also, exhibiting information integration and anatomical decision making at the system-scale.

unambiguous individuals have always been (single) selective or utility-maximising units. Thus, the credit assignment issues of collective intelligence are not categorically distinct from related core issues in individual adaptation, evolution and intelligence.

Towards a unified theory of intelligence and cognition

In collectives, each component selects behaviours based on the rewards they receive for their own actions (Figure 1(a)). In intelligent systems, the reward feedback is effectively operating at a higher level – and the system as a whole selects behaviours based on the rewards received by the system as a whole (Figure 1(b)). Accordingly, it makes sense that the system selects behaviours that facilitate long-term collective reward. But operationally, each component within the intelligent system is still autonomous, selecting individual actions based on individual rewards given the relational context they find themselves in. The question is, *what kinds of interaction structures cause collectives to behave like intelligent agents, exhibiting information integration and coordinated action that effect reward feedback at the system level?* (Figure 1(c)). Here, we propose a formalism for thinking about these issues as a set of hypotheses to drive future research.

Establishing these commonalities has significant consequences for understanding: since some of these questions have well-developed answers in the context of individual intelligences, those answers can be transferred to provide a framework for approaching collective intelligences. While connectionist models of cognition and learning do not have all the answers, they do identify the kind of relationships that turn a collection of unintelligent components into a collective intelligence, with cognitive and learning abilities that belong to the whole and not the parts. Additionally, connectionist models identify conditions where collective intelligence can

arise bottom-up, using only distributed learning mechanisms without system-level or global feedback.

We do not attempt a comprehensive review of the many related topics involved. Rather, we have selected foundational points to clarify a vision of distributed cognition, bottom-up adaptation and, more generally, the 'more than the sum of the parts' conceptual territory.

A framework for interrogating collective intelligence

Our thinking builds on a core conjecture that the kind of relationships necessary to produce evolutionary individuality – the generation and heritability of fitness differences at the collective level (Watson et al., 2022) – are the same as those required to produce organismic individuality – the information integration and collective action characteristic of a self (Levin, 2019, 2022a). More specifically, we propose that these relationships are *cognitive architectures* regardless of the substrate in which they are implemented (Watson et al., 2022). That is, the causal structures necessary to create fitness that properly belongs to the whole rather than its parts (Watson and Thies, 2019) are the same as those required to carry out the integration of information and coordination of action characteristic of a 'self' (Levin, 2019; Manicka and Levin, 2019a). We propose that such functionality constitutes cognition in a formal sense, whether the causal structures are implemented by chemical, gene-regulatory, bio-electrical, neural, ecological or social interactions (Manicka and Levin, 2019a; Watson and Szathmáry, 2016), consistent with the emerging field of unconventional and basal cognition (Levin et al., 2021; Lyon, 2020; Lyon et al., 2021).

Understanding the parallels between individual and collective intelligence via a connectionist framework

The curious thing about collective intelligence is that the more intelligent something is, the less it looks like a collective. When component members act in an efficiently coordinated manner, with behaviours that can be diverted from their short-term self-interest to serve long-term collective interest, a collective looks more like an individual at a higher level of organisation. Recognising this, biology is full of collective intelligence – not just in the weak sense of swarms and colonies with emergent behaviours but because any organism is actually an intelligent collective. Here we look at (1) collective behaviours and functions within organisms, especially development and the idea of 'basal cognition'; (2) the complex relationship between organismal identity and evolutionary units, and how this has changed over evolutionary time; (3) the substrate-independence of intelligence and how cognition and learning can occur in various kinds of biological networks;

and briefly, (4) how the problem of credit assignment arises as a core theme in these issues.

Organisms as collective intelligences: Development and basal cognition

All organisms are collectives at multiple levels: from collections of active molecules in a cell, to collections of cells in a multicellular organism or a tissue, to collections of tissues in an individual organism. What makes collectives individuals (as opposed to merely populations in containers) is their intelligence – their degree of competency in solving novel problems (Fields and Levin, 2022; Levin, 2023). The processes of development are the substrate of this intelligence – the 'glue' that makes the whole more than the sum of the parts (Levin, 2019). A large body of work indicates that development is not well-characterised as the execution of a pre-programmed genetic script but rather as an active, dynamic and adaptive process. Although all cells in most multicellular organisms share the same genome, the remarkable protein machinery that genome encodes, along with the cytoskeletal and lipid structures each cell inherits from its ancestors (Fields and Levin, 2018b), enables a collective of embryonic cells to develop differentiated roles and self-organise into a large-scale, functional machine. Development thus involves a multi-scale hierarchy of co-operating and competing subunits (Fields and Levin, 2020), each with local computational and goal-directed capacity, that enables the whole to function as a singular subject of memories and preferences – a unified locus of learning and homeostatic loops that harness its subunits towards goal states.

Morphogenesis as an instantiation of collective intelligence. Anatomical homeostasis – the ability to adjust anatomy despite injury or drastic rearrangement (Harris, 2018; Levin et al., 2019) – requires the collective to have a degree of autonomous problem-solving activity in morphospace, defined as the space of possible anatomical configurations (Stone, 1997). For example, eyes developed ectopically in the tails of frog embryos still allow the animals to see (Blackiston and Levin, 2013) because the eye primordia cells succeed not only in forming an eye and optic nerve in an abnormal environment but also in connecting the optic nerve to the nervous system (in this case, via synapse onto the spinal cord, rather than the brain). Another example is the development of the newt kidney tubule (Fankhauser, 1945a; 1945b): normally cell–cell communication among ~8 cells produces the correct tubule diameter, but if the cells are made very large, they still produce the same diameter tubule by using fewer cells. Even when cell size gets very large, a single cell can achieve the same diameter tubule by bending around itself (this time using cytoskeletal

mechanisms). Thus, genetically wild-type cells can harness distinct molecular components, depending on the novel circumstances, to reach the same high-level anatomical goal.

This disrupts a straightforward reductionist or bottom-up account of organismal morphology and function. Whilst natural selection provides the genetic hardware, this hardware has a very particular kind of plasticity, which implements robustness to both external and internal novelty. This derives from an architecture of multi-scale competency (Fields and Levin, 2020; Gawne et al., 2020), where many subsystems are themselves goal-directed and can pursue specific endpoints despite changes in their tissue environment, greatly potentiating evolvability. The idea of organisms as pre-specified machines, assembled by genetic scripts, fails in the context of these and other examples of developmental robustness. We therefore seek to understand these capacities in the context of a different and more flexible conceptual space.

Basal cognition in development: Morphological problem-solving. 'Basal cognition' refers to information processing that occurs in an unconventional substrate and/or as a simpler evolutionary precursor to what we conventionally consider cognition (Baluška and Levin, 2016; Levin, 2019; Manicka and Levin, 2019a). This is not cognition that depends on neurons or necessarily involves second-order self-awareness (Levin, 2019). It refers to cognition in an algorithmic sense that is substrate-independent (Levin, 2019) and is observable as problem-solving across phylogenetic history (Keijzer et al., 2013; Levin et al., 2021; Lyon, 2015; Lyon et al., 2021). What is important in basal cognition is not the presence of neurons but the presence of functional and informational interactions that facilitate both information integration and the ability to orchestrate cued responses that coordinate action (Bechtel and Bich, 2021; Grossberg, 1978; Levin, 2019). This can be implemented by suitable interactions of any nature including gene regulatory networks, cell signalling, bio-electric networks and morphogenetic chemical feedbacks (Lyon et al., 2021).

For example, the process of growing a limb constitutes basal cognition, as it requires both integration of multi-dimensional information (e.g. to 'decide' appendage type or handedness, from context) and collective action to put this 'basal decision' (Bechtel and Bich, 2021) into action (e.g. to coordinate the timing, abundance and positioning of cellular differentiation and growth (Dinet et al., 2021; Fields and Levin, 2020; Moczek, 2019). More broadly, regulative development, regeneration and remodelling (such as morphogenesis) require collective decision making and memory at two scales: on the part of cells (collectives of molecular networks) and of tissues (collectives of cells). Limb regeneration, for example, requires a memory of the correct pattern, the ability to compare current state with the target state and the ability to traverse anatomical morphospace in

different ways depending on context and perturbations (Pezzulo and Levin, 2016).

William James' definition of intelligence – the ability to achieve the same goal in multiple ways (James, 1890) – provides context for considering the basal intelligence of cell collectives in morphogenesis. It has become clear that the large-scale morphological goals of an organism override and harness the local competencies of individual cells to adaptively navigate morphospace (Levin, 2022a). That navigation capacity is not hardwired but shows considerable problem-solving plasticity (reviewed in (Levin, 2023)). Numerous examples indicate that morphogenesis meets James's definition of intelligence by achieving normal anatomy despite a wide range of serious perturbations. For example, developing Xenopus tadpoles can attain the same anatomical outcome despite starting with their craniofacial organs scrambled (Vandenberg et al., 2012) or with the wrong number of cells (Cooke, 1979, 1981). Even mammalian embryos can overcome drastic perturbations such as amputation; and early embryo splitting in humans results in normal monozygotic twins rather than partial bodies.

The ability of collectives of cells to pursue, with various degrees of competency, target states in anatomical morphospace (Levin, 2023; Stone, 1997) reveals an important aspect of being an individual: solving problems in a space different from that occupied by its parts (Fields and Levin, 2022; Levin, 2023). While individual cells cannot ascertain the right number, size or position of eyes or fingers, tissues do so routinely, that is, the tissue as a collective executes morphogenesis through differential cell reproduction and differentiation, stopping when the correct structure is complete (Birnbaum and Sánchez Alvarado, 2008). While cells navigate transcriptional and metabolic spaces, cellular collectives can navigate anatomical morphospaces and the conventional behavioural space (Fields and Levin, 2022).

Altered states: Basal cognition and manipulated target morphology. This framework makes a strong prediction: if intercellular signalling (not genes) is the cognitive medium of a morphogenetic individual, it should be possible to exploit the tools of behavioural and neuro-science and learn to read, interpret and re-write its information content in a way that allows predictive control over its behaviour (in this case, growth and form) without genetic changes. This prediction has been validated in several species. The bioelectric signatures that drive accurate regenerative reproduction/development in planaria have been identified ('reading and interpreting' anatomical target information, Durant et al., 2016; Durant et al., 2017; Pezzulo et al., 2021). Planaria normally have one head, but this is not genetically determined, merely a default: transient bioelectrical modulation of the body-wide pattern memory circuits can shift them to a persistent two-headed state, causing subsequent pieces of that planarian to regenerate into two-headed worms ('re-writing') (Durant et al., 2016). This induced

phenotype then persists through future rounds of amputation until set back to normal with a different bioelectrical manipulation (Durant et al., 2017); it even exhibits features of advanced individual cognition such as bi-stability (Pezzulo et al., 2021). These target morphology shifts occur despite the fact that all of the individual cells have unaltered normal genomes, showing that competent subunits can be pushed to implement diverse organism-scale goals by physiological signals (experiences) without modification of their essential hardware. In addition, this can happen rapidly – not requiring evolutionary timeframes. Other examples of reading, interpreting and rewriting the bioelectric information dictating morphogenesis have been described in a range of model systems (Levin, 2021b). Consistent with the idea that cellular swarms can act as a consolidated cognitive agent, morphogenesis is known to be altered by prior experiences (e.g. amphibian limbs ceasing to regenerate after repeated amputation (Bryant et al., 2017)) and confused by exposure to classic cognitive modifier drugs (Sullivan and Levin, 2016).

Bioelectricity: A 'cognitive glue' common to collective and individual intelligence. The many parallels between behavioural control by nervous systems, and the ancestral capacity of morphogenetic control by all cell networks (Fields et al., 2020), are reviewed elsewhere (Pezzulo and Levin, 2015). But it's crucial to note that the very same cognitive glue – bioelectrical networks implemented by ion channels and electrical synapses – operates to bind neurons into competent individuals in the 3D world of behaviour and to bind other cell types into competent individuals in the morphogenetic space of anatomical control. These insights are now driving computational models used to understand the tissue-level decision making that results in birth defects (Manicka and Levin, 2019b, 2022) and their repair (Pai et al., 2018; Pai et al., 2020; Pai and Levin, 2022), giving rise to promising therapeutics.

These capacities of morphogenetic cellular collectives are *basally cognitive* insomuch as they involve information integration and coordinated action (Fields and Levin, 2020; Grossberg, 1978; Levin, 2019; Manicka and Levin, 2019a; Newman and Bhat, 2008), characteristic of a self (Levin, 2019). More radically, perhaps this kind of cognition is actually what *constitutes* organismal individuality – that is, the processes of basal cognition essential for achieving specific system-level goals in anatomical morphospace are exactly what make the whole different from a collection of parts (Watson et al., 2022).

Collective intelligence as a product of evolutionary selection, or evolutionary selection as a product of collective intelligence?

Biological individuality has traditionally been associated with the scope of an evolutionary unit (Clarke, 2016) – the unit that is subject to differential survival and reproduction.

Within this orthodox view, whilst the processes of developmental basal cognition are certainly complicated and might have the appearance of collective behaviour, they are merely complex parts of a single individual. However, this view turns out to be wholly inadequate to understand and manipulate the multi-scale nature of life.

Genetic identity and biological individuality. The idea that biological individuality can be defined by genetic identity is clearly insufficient: the structural and functional demarcations of coherent individuals often diverge from their genetic information. Note that a colony of bacteria may be genetically homogeneous but not an individual, while planaria are biological individuals by any reasonable sense of the word but not genetically homogeneous (Fields and Levin, 2018a). Even though genetically identical, the tissues and cells within a classical organism (body) often compete with each other (Gawne et al., 2020); conversely, cells from distant species cooperate well within chimeric organisms (Nanos and Levin, 2022). In addition, genetic information does not always predict the structure and function of bioelectrically modified organisms (Levin, 2014, 2021a) or of self-organising synthetic living machines (Blackiston et al., 2021; Kriegman et al., 2020). Likewise, often it is the degree of bioelectrical coupling, not genetic differences, that determines whether cellular optimisation occurs at the single-cell level (cancer) vs. at the organ-level (normal morphogenesis) (Chernet and Levin, 2013).

Evolutionary units and biological individuality. Can a notion of evolutionary units beyond genetic relatedness rescue a meaningful concept of biological individuality? That is, the ability to exhibit heritable variation in reproductive success might obtain for a complex or composite whose components are not genetically related. For example, despite being of separate ancestral origins, the nuclear and mitochondrial DNA of eukaryotes can be considered a single evolutionary unit (under most conditions) by the virtue of their common vertical transmission. However, identifying what exactly constitutes an evolutionary unit in general is also non-trivial – especially because they change over evolutionary time and new units arise at new levels of organisation (Okasha, 2006).

To be a bona fide evolutionary unit, a collective must exhibit heritable variation in reproductive success that belongs properly to the collective level – over and above the sum of that exhibited by its component parts (Okasha, 2006; Watson et al., 2022; Watson and Thies, 2019). This requires organised functional relationships that cause short-sighted self-interested entities to behave in a manner that serves the long-term collective interest of the whole. In this light, the complex nature of functional relationships between component parts begins to look less like the product of selection at the system level, and more like the source of evolutionary individuality.

Practical implications: Beyond philosophy. Such considerations matter fundamentally to our understanding of the organismic, evolutionary and developmental biology (i.e., emergent functionality) and thus to our ability to predict, control, manage and manipulate multi-scale biological systems. Understanding what kind of relationships instantiate biological individuality is thus of great importance to synthetic bioengineering, regenerative medicine, exobiology, robotics and artificial intelligence.

For example, to intervene in the processes that coordinate component parts to create or regenerate an organ or a limb – or produce an entirely novel construct such as a self-assembling biobot (Ebrahimkhani and Levin, 2021) – we must be able to manipulate the very relationships that define individuality (Levin, 2021c). Such bioengineering goals therefore depend intimately on our knowledge of collective intelligence at multiple levels of biological organisation (Beane et al., 2013; Herrera-Rincon et al., 2018; Pezzulo and Levin, 2015).

Recent work has begun to apply the tools of collective intelligence and cognitive neuroscience to morphogenesis and its disorders, including cancer, a disease of dysregulated morphogenesis (Deisboeck and Couzin, 2009; Doursat et al., 2013; Friston et al., 2015; Pezzulo et al., 2021; Pezzulo and Levin, 2015, 2016; Rubenstein et al., 2014; Slavkov et al., 2018). Disconnection from the bioelectric network of tissues often gives rise to fragmenting of coherent anatomical individuals into invasive single cells and tumors; their release from higher level collective goals is readily apparent because they pursue anatomical, histological and physiological states quite different from those that the organism tries to maintain (Egeblad et al., 2010; Levin, 2021c; Radisky et al., 2001; Soto et al., 2008). This fragmentation can be reversed: despite strong oncogenic mutations, cancer phenotypes can be suppressed by forcing bioelectrical connections among cells, thus overriding single-cell level goals with large-scale morphogenetic ones (Chernet and Levin, 2013).

Cognition, learning and problem-solving in biological networks: Generalised principles of connectionism

The link between evolution and simple types of learning has often been noted (Skinner, 1981; Watson and Szathmáry, 2016) but sometimes interpreted in an uninteresting way: *learning is simply a form of random variation and selection* (Campbell, 1956; Skinner, 1981; Watson and Szathmáry, 2016). However, the formal equivalence between evolution and learning (Campbell, 2016; Frank, 2009; Harper, 2009; Shalizi, 2009) also has a much more interesting implication, namely: *Evolution is more intelligent than we realised* (Chastain et al., 2014; Parter et al., 2008; Valiant, 2013; Watson and Szathmáry, 2016). *Connectionist* models of conventional learning, familiar in artificial neural networks, greatly expand this perspective (Watson et al., 2016, 2022;

Watson and Szathmáry, 2016). Connectionist models inherently implement the fact that intelligence resides not in the parts but in the organisation of the relationships between them (Watson et al., 2016; Watson and Szathmáry, 2016). Such models demonstrate how networks of organised functional relationships between simple reactive (stateless) components are sufficient to exhibit information integration and coordinated responses. Moreover, these relationships can be organised by simple distributed, incremental processes, that is, learning (Watson et al., 2011a; 2011b; 2011c; 2016; Watson and Szathmáry, 2016).

Hebbian learning in networks. A simple example of such a neural model, demonstrating distributed computation and learning, is the Hopfield network (Hopfield, 1982) (Appendix Box 1). Given that the Hopfield network is inspired by neural dynamics and learning in cognitive systems, its learning and problem-solving abilities are perhaps not so surprising, despite their decentralised operation. However, the underlying principles are extremely simple and general: the same computational algorithms also apply in systems that we don't normally expect to be capable of cognition or learning; gene-regulation networks, protein interaction networks and ecological community networks can all implement the same kinds of functions as neural networks if organised appropriately (Biswas et al., 2021; Herrera-Delgado et al., 2018; Szabó et al., 2012; Tareen and Kinney, 2020). However, cognition in different substrates may have very different spatio-temporal scales – from the cellular, to the familiar organismic scale, and perhaps to the ecological scale (Power et al., 2015; Watson et al., 2014). Can these kinds of networks also learn as neural networks do?

The answer is yes. Hebbian learning in a self-modelling dynamical system (Appendix Box 1) effects a positive feedback on correlations; the more things co-occur, the more the connection between them changes to make them more likely to co-occur in future. This positive feedback on correlations is quite natural. In some conditions, it does not require an active learning mechanism that strengthens connections, instead it is sufficient to differentially relax or weaken connections according to the frustration or stress experienced in that connection (Buckley et al., in prep). Thus, connectionist modes of cognitive learning can be instantiated in various kinds of non-neural networks (Davies et al., 2011; McCabe et al., 2011; Power et al., 2015; Watson et al., 2011b).

Importantly, the application of connectionist models also extends into the domain of evolutionary systems, where the connections of a network are changed by variation and selection, as seen in the evolution of interaction networks in development and ecology (Brun-Usan, Rago, et al., 2020; Brun-Usan, Thies, et al., 2020; Kouvaris et al., 2017; Rago et al., 2019; Watson et al., 2014; Watson et al., 2016; Watson and Szathmáry, 2016). In these 'evolutionary

connectionism' models, ordinary processes of random variation and selection act on the functional interactions between components, altering their organisation in a way that positively reinforces correlations – functionally equivalent to connectionist learning models (Watson and Szathmáry, 2016). The algorithmic principles well-understood in neural networks, are equally demonstrable in gene-regulation networks (Brun-Usan et al., 2020; Brun-Usan et al., 2020; Kounios et al., 2016; Kouvaris et al., 2017; Rago et al., 2019; Watson et al., 2014), and ecological community networks (Power et al., 2015) and social networks (Davies et al., 2011; Watson et al., 2011a). This algorithmic unification between connectionist learning and evolution (Watson et al., 2016; Watson and Szathmáry, 2016) opens up the transfer of an extensive, well-developed toolset from machine learning into evolutionary theory to naturalistically explain evolutionary 'intelligence' (Kounios et al., 2016; Watson et al., 2022; Watson and Szathmáry, 2016).

In particular, it is important to recognise that connectionist models can exhibit learning bottom-up, without centralised control or an external teacher, and without any performance feedback applied at the system level, via fully distributed and unsupervised learning principles (Watson et al., 2011a; Watson et al., 2011b, 2011c). This means that the same learning behaviours can be exhibited by an ecological community without selection at the community level (Power et al., 2015). This is potentially important to understanding the evolution of intelligent collectives (and evolutionary transitions in individuality (ETIs)) because it identifies conditions where relationships between evolving entities can be organised via natural selection acting at the lower level *before* selection at the higher level takes effect (Watson et al., 2022; Watson and Szathmáry, 2016).

So, what kind of cognition can such networks exhibit?. We find it useful to operationalise *cognition* in an algorithmic sense, namely: what kind of problem-solving can it do? Organisms solve problems in many different spaces including morphological, metabolic, transcriptional or behavioural (Fields and Levin, 2022). Limited forms of problem-solving can be demonstrated with simple networks like the Hopfield model (Hopfield and Tank, 1986). The problem-solving behaviour of such a system without learning can be taken as a base line, or null model, as it merely describes a local energy descent process with fixed points corresponding to locally optimal solutions (of the energy-minimisation problem implicit in the constraints between its components). To do better than that – to avoid being trapped in local minima – requires a system to learn an internal organisation that knows something about the solution space from past experience, either on agent timescales (the familiar scale of cognition) or on evolutionary timescales (Kounios et al., 2016; Kouvaris et al., 2017; Watson and Szathmáry, 2016).

The ability of distributed learning to improve problem-solving ability in this way is now well-developed (Kounios et al., 2016; Mills, 2010; Mills et al., 2014; Watson et al., 2011a; Watson et al., 2011b, 2011c; Watson et al., 2016). In some conditions, a learning neural network can enable a sort of 'chunking', rescaling the search process to a higher level of organisation (Caldwell et al., 2018; Mills, 2010; Watson et al., in review; Watson et al., 2011c; Watson et al., 2016). Elsewhere, we hypothesise that this rescaling of the problem-solving search process is intrinsic to transitions in individuality (Watson et al., 2016), suggesting that ETIs constitute a form of deep model induction (Czegel et al., 2019; Vanchurin et al., 2021; Watson et al., 2022).

Credit assignment in individuals and collectives

Conventional accounts of intelligence and behavioural protocols (Watson, 1967) assume a singular subject of intelligence and of the goals that it can pursue. However, this is a significant over-simplification that obscures important questions about how centralised intelligences arise out of cellular components (Levin, 2019, 2021c). For example, one trains a rat to press a lever and receive a delicious reward, in instrumental or associative learning paradigms (Abramson, 1994; Best, 1965; Rescorla and Solomon, 1967). The rat is understood to be an intelligent agent solving an instrumental learning problem; but it is also a collection of cells. Indeed, the cells that perform the action (muscle and skin cells that interact with the lever) and the ones involved in sensing the environment (seeing the lever, feeling the lever and tasting the reward) are not the cells that immediately receive the nutritional benefit of the reward (intestinal lining). No individual cell has the entire experience of performing an action and reaping its benefits – that relationship only exists in the 'group mind' of the collective agent. How do the parts discern which of their actions should be reinforced? Problems of distributed credit assignment are a key aspect of intelligence, even in conventional organisms.

It is imperative to understand the developmental algorithms and signals by which tissue-level agents incentivise lower-level subunits (e.g. cells and molecular pathways), distorting their option space so that simple, local descent down free-energy paths (short-sighted self-interest) result in higher order adaptive activity (long-term collective interest). The key to being an individual is to have a functional structure in which diverse experiences across its components are bound together in a way that generates causal relationships and composite memories that belong to the higher space of the individual and not its components (Fields and Levin, 2022).

How does scaling of reward dynamics bind subunits into intelligent collectives that better navigate novel problem spaces? Lessons from machine learning. It is no accident that the issue of credit assignment, and the application of credit to parts or wholes, is a central one in evolutionary selection, developmental and organismic biology and cognitive science. It is a feature of many difficult learning tasks that they require sequences of actions that are many steps away from ultimate goals – making it intrinsically difficult to incentivise the component parts involved. This is what makes difficult tasks difficult; conversely, having feedbacks that are additive and individual, is what makes easy tasks easy. It is no coincidence then, that these issues of credit assignment have well-developed formalisms in the domain of machine learning (Watson et al., 2022). In particular, one of the touchstones of machine learning – the ability to represent non-linearly separable functions (such as XOR - Exclusive OR logical operator) – is distinguished from linearly separable functions exactly because improvements in the output cannot be ascribed to the independent contribution of individual inputs (Watson et al., 2022). Nonetheless, simple connectionist models can learn such functions if they have a suitable architecture (see below).

Connectionist models thus identify some basic criteria about the kind of relationships that turn a collection of unintelligent components into a non-decomposable intelligence with cognitive and learning abilities that belong properly to the whole and not the parts. Moreover, the ability of *unsupervised* learning processes to exhibit collective problem-solving capabilities suggests conditions where this can arise bottom-up, using only distributed learning mechanisms without pre-supposing collective-level feedback. These principles do not require that the collective is already an evolutionary unit, nor do they require that the members of the collective are neurons.

Together, these observations show that the apparent distinction between individual intelligence and collective intelligence is not substantial: at a minimum, they exist on a continuum. Further, the connectionist models of cognition and learning developed for individual intelligence are not simply relevant to understanding what is required for a collective to be intelligent, it may be that it is precisely these cognitive capacities that are the fundamental difference-maker with respect to individuality itself; i.e. between 'many individuals' and 'one individual'.

What kinds of interaction structures are necessary for what kind of (collective) intelligence and how can these structures emerge?
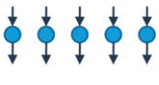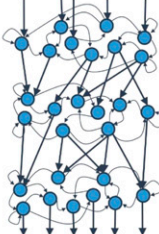
Some of the different cognitive behaviours we might be interested in for collective intelligence include information integration, holding state over time, storing and recalling multiple memories and recognising past states, generalising, problem-solving and multi-scale autonomy (Baluška and Levin, 2016; Levin, 2022a). Moreover, we are interested in how any of these behaviours can be understood to belong to the whole – or indeed, to multiple organisational scales – rather than the parts. Whilst some of these behaviours might not be very well defined in the context of collective behaviour, our approach is to describe how they relate to the different types of connectionist architectures, familiar in artificial neural networks, where these behaviours are better understood. This approach offers a speculative synthesis of machine learning concepts with basal cognition and evolutionary theory – and a roadmap of gaps and opportunities for future research in collective intelligence (Table 1). We first discuss what interaction structures are needed and then how such structures can emerge 'bottom-up' through distributed learning mechanisms.

The structure of interactions

Naturally, the ability to represent relationships (e.g. correlations or associations) among variables, rather than a system of independent variables, is essential for any connectionist model of cognition, and requires components to have connections of one kind or another. For example, in development, gap-junctions between non-neural cells are physiologically tunable 'synapses' that communicate cellular behaviours (Mathews and Levin, 2017; Palacios-Prado and Bukauskas, 2009) and indeed can drive genetically wild-type cells to build body organs belonging to diverse species (Emmons-Bell et al., 2015). There are many other levels of biological organisation with different 'signals and responses' between components, or sensitivity to one another's behaviours. Being connected is necessary but not sufficient for cognitive functions, however. Connectionist principles enable us to be more specific about what kinds of connection structure are important.

Instructive neural architectures from machine learning. This is by no means a survey of machine learning techniques or a comprehensive description of neural architectures; our aim is simply to highlight some of the key architectural issues and their significance with respect to different cognitive abilities. Three particular architectural issues have special significance:

Table 1. What kinds of relationships are necessary to turn a society into an individual? A central aspect of how intelligence arises from a collection of subunits is the specific communication and functional linkages between them, as well as the algorithms for updating those interactions in light of experience. It is thus essential to determine what kinds of architectures underlie different degrees of agency (which support memory, problem-solving, information integration and collective action, higher-level autonomy, etc.) across the continuum. Here, we leverage connectionist models of cognition and learning (top row) to specify known architectures that embody key waypoints along the collective intelligence spectrum (bottom row), as well as to identify knowledge gaps that highlight opportunity for next steps in this field. References indicate examples of potentially relevant models where available. Shading indicates speculative suggestions and opportunities for future research. In the final column, the biological examples are known but the relevant topology is not.

	Unconnected	Shallow feed-forward	Deep feed-forward	Shallow recurrent	Deep and recurrent	Beyond current ML
Interaction structure						Dynamic topology? Networks of networks? ??
Functional characteristics	Each variable is a function of its own input only (e.g. *univariate* model). A model of independent variables. Cannot represent any interactions or associations between variables. The reward that a particular input-ouput relation receives might depend on some other input, but this model cannot capture such an interaction.	Each output variable is a non-linear weighted sum of multiple inputs (e.g. *Perceptron*). Can capture interactions but represents outputs that are *linearly separable functions* only. Signal from one input can modify the *strength* of the response between another input-output pair but not the direction.	Each variable is a non-linear weighted sum of inputs from the layer above (e.g. *multi-layer Perceptron*). Output can represent a non-linearly separable function of inputs (from two, or more, layers above). One input can modify the strength *and direction* of the response between another input-output pair.	Each variable at time t is a function of outputs from several others at time t-1 (e.g. *Hopfield network*). Arbitrary connections including self-connections and recurrent loops. Outputs are no longer a static mapping of inputs - represents a dynamical system with potentially many attractors that can hold state over time; system behaviour depends on history.	Each variable is a function of any other in its own layer or the layer above (e.g. *deep belief network* or *restricted Boltzmann machine*). Recurrence restricted to variables in same layer. Each layer can hold state over time, and has behavior that depends on history. And lower layers can represent a non-linearly separable function of layers above.	Data-dependent and time dependent topology? Represents a strategy for searching topology space (rather than a particular topology)? Networks that reorganise to seek a goal, or target behavioural attractor, or to maintain a set-point (rather than computing a static function/dynamic pattern)?
Learning methods	Each weight can be learned independently.	Perceptron learning algorithm gaurantees convergence to set of weights that represent a linearly sepable function accuratley.	Gradient methods (e.g. back-propagation algorithm) are not guaranteed to converge but can work well in practice (for networks that are not too deep).	Gradient methods generally inapplicable. Weights can be stochastically adjusted incrementally but, again, this can be less effective for deep networks.	Very difficult to train in general. But one approach trains one layer then adds another, and so on to construct a deep network (Hinton et al., 2006). Each new layer represents higher-order associations and higher-level emergent control.	Evolutionary, learning and behavioural timescales overlapping? Learning to learn in real time? Learning and meta-learning at multiple scales, not just learning higher-order correlations?
Biological examples and significance (for interactions within and between units)						
Interactions *within* one evolutionary unit, or credit assignment at level of whole system	A model of a phenotype where each trait is controlled by one gene. Asexual reproduction enables heritable combinations of alleles (phenotypes are particulate units). (Harper, 2009; Shalizi, 2009)	Phenotypic traits are pleiotropic functions of multiple genes (a simple pleiotropic genotype-phenotype map). (Hansen, 2006; Jones et al., 2007; Jones et al., 2014; Kashtan et al., 2009; Pavlíček & Cheverud, 2015)	A 'deep' pleiotropic genotype-phenotype map (complex but feed-forward/non-dynamic) represented by higher-order pleiotropy (Brun-Usan, Thies, et al., 2020; Draghi & Wagner, 2008, 2009; Hansen et al., 2006; Nash et al., in prep; Parter et al., 2008)	Gene expression dynamics determined by a recurrent gene-regulation network. Selection acts on attractor states (Watson et al., 2014). Network can store and recall multiple phenotypes favoured by past selection and use generalisation to improve evolvability.	Multi-scale autonomy? Basal cognition/development with semi-autonomous control at multiple levels of organisation.	Dynamic regeneration of functional units under injury (Pezzulo & Levin, 2016). Reconfiguration to novel stresses/damage on-the-fly (Pezzulo & Levin, 2016).

Table 1. (continued)

Interactions *between* multiple evolutionary units, or credit assignment at level of individual components	Independent evolution of multiple species within an ecological setting.	Evolution of ecological interactions under individual selection limited to one trophic level?	Evolution of ecological interactions with multiple trophic levels?	Evolution of ecological interactions with arbitrary (non-trophic)	Evolutionary transitions in individuality?	Flexible individuality (e.g. slime molds), autonomous modules that goal-seek/ exhibit
	Also, selection on independent genes under free recombination (combinations of alleles are not heritable). (Chastain et al., 2014)	Ecological configurations could be linearly separable function of environmental conditions, thus ecosystem has evolved structure and can change strength, but not direction, of selection on other components	Ecological configurations could be non-linearly separable functions of environmental conditions (including other species), thus interactions can change the direction of selection on other components and the ecosystem responds in a non-decomposable manner.	species interactions (Power et al., 2015) Demonstrates ability to store and recall multiple 'ecological memories' at system level without system-level selection, and improve resource allocation problem-solving with experience (Power, 2019)	Associations between low-level evolutionary units become subsumed by higher-level units. Each layer integrates information and coordinates action of units at layer below.	multi-scale competency (Fields & Levin, 2020; Gawne et al., 2020). Enables evolutionary search in higher-level components that autonomously fill in the details on-the-fly.

1. *Feed-forward mappings and recurrent dynamics:* Artificial neural networks are often used to represent (and learn) a mapping between inputs and outputs (e.g. for classification or regression tasks). One of the simplest 'feed-forward' networks is the single-layer Perceptron where an output node fires if the sum of its weighted inputs exceeds a threshold (more generally the output is some non-linear monotonic function, e.g. a sigmoidal function, of the weighted sum of inputs). This is capable of representing simple input–output relationships and learning to classify inputs according to such relationships. In other cases, connections can be recurrent, that is, connections can form loops and thus states can be influenced by inputs from previous time steps and the system can continue to hold internal state after the input is removed. They can also, thereby, exhibit temporally extended dynamical behaviours. Accordingly, in recurrent networks we are often interested in the dynamical attractors of the system (which are a function of the system's own internal history not just current inputs) rather than instantaneous values of designated outputs or the input–output relationship. The Hopfield network is a simple example (Appendix Box 1). Because connections are symmetric (with no self-connections) in the Hopfield network, its dynamics have only fixed point attractors ('memories'), but more general recurrent architectures may have periodic or chaotic dynamical behaviours.

2. *Deep representations and non-linearly separable functions:* The single-layer Perceptron has important limitations. Specifically, although it can represent 'linearly separable functions' where the response to a change in one input changes *magnitude* depending on the value of another input (i.e. the responses are not independent), it cannot represent *non-linearly separable functions* where the response to a change in one input changes *direction* depending on the value of another input (Watson et al., 2022). This type of interdependence is important because in the linearly separable case, if an input contributes positively to an output in one context, it never contributes negatively in another. This means the single-layer Perceptron can represent cases where 'working together' changes the benefit an individual input can receive (from doing what they were doing anyway), but it cannot represent cases in which working together requires an individual to do the opposite behaviour, move in the other direction or do something opposed to what they were doing when they worked alone or in some other context. Representing non-linearly separable functions requires a network with multiple layers – a multi-layer Perceptron (MLP). In principle, an MLP can represent any function of the inputs given sufficient 'hidden' variables (units that are neither inputs nor outputs but constitute an intermediate layer of representation). In practice, it is frequently useful to employ more layers (with fewer nodes each) because this affords a different inductive bias and generalisation. These are known as *deep networks* (LeCun et al., 2015).

3. *Deep and recurrent networks*: Whilst there are many other architectures used in artificial neural networks, two others are worth mentioning. A deep auto-encoder is a network that compresses a high-dimensional input space into a low-dimensional representation. A decoder decompresses the low-dimensional representation back into the original high-dimensional space. The compressed encoding can be interpreted as a low-dimensional model of the samples observed on the input space. Changes to the variables of the compressed

representation produce large, coordinated changes to the variables in the input space. Lastly, the deep belief network (DBN) (Hinton et al., 2006) is quite a special type of network, and its architecture has particularly relevant properties. The DBN has a layered architecture that can be used to learn compressed representations like the auto-encoder, and within each layer the nodes have recurrent connections. This gives the DBN both the potential to represent low-dimensional recodings of the original input space and to have dynamical attractors that stably retain their state at that higher level of representation.

Implications for evolutionary intelligence and basal cognition

Naturally, for a collection of individuals to exhibit any kind of collective intelligence, it is, at the very least, necessary that the behaviour of one individual has some sensitivity to the behaviour of another. Such interactions can coordinate behaviours to take advantage of scenarios where the benefit/reward or fitness that one individual receives is sensitive to the behaviour of another. However, if this credit-assignment interaction (or fitness epistasis) constitutes a linearly separable function this is not really a difficult problem; although the benefit they receive will vary in different contexts, the behaviour that maximises their benefit is always the same. In contrast, when the credit that one individual receives has an interaction with the credit that another individual receives which constitutes a non-linearly separable function (Watson and Thies, 2019) (or reciprocal sign epistasis, (Weinreich et al., 2005)), this requires that one individual can change its behaviour (or 'do the opposite') depending on the context of what other individuals are doing. For a collective to coordinate behaviours to take advantage of such interactions, it must be able to represent non-linearly separable functions, which requires the interaction structure between individuals to have some depth (Watson et al., 2022).

These are just the kind of relationships that make the credit assignment or fitness of the whole not only different from the sum of the rewards/fitnesses of the parts but also a non-decomposable function. Intuitively, this changes our relationship from 'how good this is for me depends on what you are doing' to 'what is best for me to do depends on what you are doing'. This is important because, when it is reciprocal, the fitness-affecting characteristics of one component only have meaning in the context of the other. In other words, it creates a 'we'; *what we are doing*, for example, whether our behaviours are coordinated or not, becomes a relevant variable (Watson et al., 2022; Watson and Thies, 2019).

Deep representations also have a special significance in recurrent networks. In non-hierarchical networks, the many connections between components can cause the system to hold state over time (i.e. internal states can be maintained as dynamical attractors even when the inputs to the network are removed or have changed). This enables the network to exhibit temporally extended behaviours, but it also has the effect that it becomes difficult to change the system state and, therefore, to be sensitive to system inputs. Getting out of one dynamical basin of attraction and into another can require large and/or specific state perturbations. The system acts as a whole but cannot 'change its mind' easily (Hills et al., 2015; Nash et al., in prep; Watson et al., in review). This is problematic for organismic adaptability and evolutionary variability. In contrast, a hierarchical representation can cause coordinated behaviour in many downstream parts but retain the capacity for small changes to variables in the higher level representation to move all the downstream variables to a new state (Nash et al., in prep). A recent alternative model is provided by a network of neurons that have a 'decision cycle' that repeatedly re-decides which states to adopt with a timing based on learned connections (Watson et al., in review). By learning to synchronise the decision cycle of particular groups of components, this kind of network exhibits multi-scale problem-solving capabilities without having an explicit or pre-defined multi-layer structure.

Cascading control architectures – where a small number of variables cause large coordinated changes in the state of many downstream variables – are common in organisms through many scales from molecular to morphological. This takes explanatory focus away from the collective and onto the units at deeper levels of the causal chain, for example, a gene cues the coordination of other biomolecules within the cell, and the germ line cues coordination of other cells within the organism. However, natural organisms are neither single-layer recurrent networks (with every component connected equally to every other like the Hopfield network) nor strictly feed-forward multi-level hierarchies (with components in one layer only connected to components in the layer below like the MLP). They are not quite like deep-belief networks either, of course, but they do contain elements of both cascading control and recurrent control architectures. This means that different levels of organisation can both be influenced by higher level control variables and be collectives that co-define and sustain their own (non-decomposable) meaning. These considerations suggest that this kind of deep and also partially recurrent architecture is relevant to the multi-scale autonomy observed in complex organisms.

Learning the structure of interactions

The previous section discussed how the types of relationships, and their organisation, might influence the type of information integration and coordinated action that could be

exhibited by a collective. But how do such organisations arise? For this, we turn our attention from connectionist architectures to models of connectionist *learning*. A number of issues and observations are relevant to collective intelligence:

Gradient methods versus stochastic local search, supervised learning versus reinforcement learning. For many learning tasks, it is useful to express the error in the output (with respect to an input and a target) as a function of the connection strengths in the network. If this function is differentiable, then this can be used (in artificial machine learning methods) to define a gradient method which computes a change in the weights of the network that will systematically reduce the error. In biological evolution or emergent collective intelligence, there is no explicit target or desired output predetermined by an external agent or teacher. There is therefore no 'error' function, as such. The more relevant type of learning is *reinforcement* learning, where different outputs confer different rewards but the 'correct' output, or the pattern that maximises reward for a given input, is not used explicitly in training (and may be unknown). Natural selection can be used to increase the fit of an organism to its environment or improve rewards by adjusting weights in the same way. These basic observations are the basis of the formal equivalence between learning and evolution by natural selection (Campbell, 1956; Harper, 2009; Shalizi, 2009; Skinner, 1981; Watson and Szathmáry, 2016).

What makes learning systems smart, however, is not merely the ability to increase the fit of model parameters to data; what makes such systems interesting is that the parameters they adjust and the data to which they fit are *not* in the same space (Buckley et al., in prep; Watson and Szathmáry, 2016). For example, the quality of the network output is, in a direct sense, a function of the network outputs and how well these fit the environmental needs. But the parameters that are adjusted during learning are not these output variables per se. Rather they are the parameters of a model that produces these outputs – namely, the network of interactions connecting one node to another. This separation between 'model parameters' and 'solution space' is crucial because without it there is no possibility of using past experience to respond appropriately in *novel* situations, that is, *generalisation* (Watson and Szathmáry, 2016).

Generalisation is fundamental to learning and intelligence. Without it, a system can only respond to current inputs in a manner consistent with past rewards. At one extreme, if the future is going to be exactly like the past, this is fine. At the other extreme, if the future has nothing at all in common with the past, then there is not much that can be done about that. But, in other cases, the future is not the same as the past, but it shares some kind of underlying regularity in common with it. These are the cases where intelligence has some meaning. Specifically, a system that can generalise can act in a manner that is consistent with long-term rewards, even when this appears to oppose immediate or short-term interests. For individuals that interact with others in a collective, the ability to act in a manner that is consistent with *long-term individual* interest is frequently aligned with the ability to act in a manner that is consistent with *collective* interest (though it may be opposed by individual short-term interest). Although this ability might seem quite sophisticated and mysterious, connectionist models of cognition demonstrate that this does not require the parts to become more intelligent; only that the relationships between them are adjusted appropriately, which can be implemented by simple incremental gradient following (Appendix Box 1).

Unsupervised learning. It might seem curious that any kind of learning can occur without supervision or system-level reward feedback of some kind. How can a learning system know what to learn if nothing tells it what it is supposed to learn? *Unsupervised learning* builds a low-dimensional model of the input data. The changes to connections are not motivated by error minimisation or reward maximisation but purely by the fit of the model to the data. Hebbian learning ('neurons that fire together wire together' (Buckley et al., in prep; Watson and Szathmáry, 2016; Watson et al., 2014) (Appendix Box 1) reduces the effective degrees of freedom in the network dynamics in a manner that 'mirrors' the degrees of freedom induced by past experience – without being rewarded for that purpose or using an error function that targets it.

The level of credit assignment in reinforcement learning and collectives. Consideration of unsupervised learning has direct significance for the evolution and reward of collective intelligence. This is because reinforcement learning acting on the individual characteristics affecting their connections to others can result in dynamics that are equivalent to unsupervised learning at the system scale (Davies et al., 2011; Power et al., 2015; Watson et al., 2011a). Intuitively, if B is rewarded for being activated, then one of the ways it can increase its reward is to increase the strength of its connection from A (e.g. when A and the connection are positive). This increases the individual reward B receives right now, but it also makes the future activation of B correlated with the activation of A (the principle of Hebbian learning in another guise). The same considerations apply to A and its connection from B. Note that neither component is making the connection with the other because it is interested in the collective reward that A and B receive together, nor because it makes the future dynamics of the AB pair more consistent with their past correlation. Nonetheless, it does make the future dynamics of the AB pair more consistent with their past correlation (Watson et al., 2014).

This observation creates a fundamental linkage between the principles of individual learning or individual utility-maximisation and the principles of system-level or collective intelligence (Watson et al., 2011a). Note that the mechanism of Hebbian learning was identified by Donald Hebb to explain neural learning because it is the right way to modify synaptic connections if you want the network to model observed correlations. This equivalent mechanism, in contrast, is motivated bottom-up – it is a consequence of components that are incentivised only by *short-term self-interest* (given that they have connections with others that they can modify). In the same way that this distributed learning does not require system-level reinforcement, it also occurs in evolutionary systems without system-level selection (Power et al., 2015). Individual selection acting on members of an ecological community produces the same structural changes to connections (inter-specific interactions) simply because each is incentivised by selection to maximise its individual growth rate. This has the same consequences for the ecological assembly rules and succession dynamics as it does for the dynamical attractors of neural activations in the Hopfield network (Power et al., 2015).

How does distributed learning effect system-level rewards and credit assignment?. This distributed learning is not *motivated* by system-level rewards (total utility), nor does it involve system-level selection, but it has a systematic relationship to system-level rewards and fitness nonetheless. In multi-agent systems, the original dynamics, given a system of constrained interactions, are much like a ball rolling down hill – each individual decides how to act to maximise individual reward as determined by the constraints with others. This finds a local optimum in total utility, but only a *local* optimum. As the individuals modify their connections from others, the dynamics of the systems are channelled into trajectories that mirror the structure of past experience. If the system is subject to repeated shocks or perturbation, or experiences an episodic stress, causing it to visit a distribution of attractors over time, then what it 'learns' is a generalised model of the constraints it has experienced. Because this model is based on interactions between components and not on independent parameters, it is a correlation model that has the potential to generalise – responding in a way that resolves constraints between individuals better than any previous attractor visited (Buckley et al., in prep; Watson et al., 2011a). In this way, short-sighted self-interested agents form relationships with one another that sometimes cause them to make different decisions (given the new weightings of the options created by the new relationships). Also, these new choices better optimise the long-term collective interests of the system as a whole (Buckley et al., in prep; Watson, accepted; Watson et al., in review; Watson et al., 2011a).

This bottom-up incremental adjustment of relationships can thus increase system-level welfare. It does so in a manner that is functionally equivalent to distributed learning mechanisms familiar in artificial neural networks, without presupposing system-level rewards or credit assignment. Moreover, in so doing, it creates a non-decomposable whole (attractors that are non-linearly separable functions of the inputs and depend on the system's own internal history), which means that credit assignment or reward at the level of individual parts and their individual behaviours becomes ineffective. Instead, credit assignment (if it applies at all) and any possibility of effecting modified behaviours through reward become meaningful only at the higher level of organisation.

Modelling collective intelligence and basal cognition: Evolutionary individuality, organismic individuality and cognition are coextensive

As discussed above, the basic computational elements of such distributed learning are substrate-agnostic and common to a wide range of biological networks (Cervera et al., 2018; Pietak and Levin, 2016, 2017). However, the conditions for distributed learning are non-trivial; not all of these networks may meet them. The important thing to note is that there is no requirement for an incentive to model long-term or collective consequences of individual actions, or for a system-level incentive to model the structure or pattern of observations. We do not yet know which of these biological systems might meet these conditions and the extent to which this influences their collective intelligence. But it is known that organismic individuality evolved through a bottom-up process of collective intelligence, resulting in information integration and coordinated action so well-organised that we observe a new level of organism and evolutionary individuality. The principles of connectionist cognition and learning described above provide a roadmap of gaps and opportunities that future research might explore to better understand how such emergent individuality occurs. In particular, the architecture of the interactions – whether they are feed-forward or recurrent, capable of representing non-linearly separable functions or not, shallow or deep or some mixture of these characteristics – has important consequences for the type of cognitive model they can represent.

The ecological models developed thus far demonstrate that connectionist learning principles are relevant to collective intelligence in systems that are not (yet) evolutionary units. They fall short, however, of demonstrating the spontaneous evolution of a new level of individuality. In algorithmic terms, such models cannot do the 'chunking' of the search space or rescaling of the search process that is facilitated by the induction of deep models (Caldwell et al., 2018; Mills, 2010; Mills et al., 2014; Watson et al., 2011b; Watson et al., 2016; Watson et al., 2009). We hypothesise that this is because they are single-level networks of

symmetric interactions; our roadmap supports the idea that the evolutionary transitions in individuality correspond to deep interaction structures (Czegel et al., 2019; Watson et al., 2022) or perhaps other mechanisms of multi-scale dynamics (Watson, accepted; Watson et al., in review).

We propose that some of the gaps in this picture might be addressed by exploring the hypothesis that evolutionary individuality, organismic individuality and cognition are coextensive (Watson et al., 2022). The idea is that acting in a manner consistent with long-term collective interests, in particular when this conflicts with short-term self-interest, is not just a hallmark of collective intelligence but is in fact what constitutes cognition and individuality at the collective level. This can perhaps be formalised through the consideration of non-linearly separable functions. Specifically, if a system of functional interactions among the parts represents a non-linearly separable function, then the incentive of the whole *is* related to the incentives of the parts *only* in a non-decomposable way (Watson et al., 2022).

Conclusions

Commonalities between cognitive and evolutionary processes and those that shape growth and form have been hinted at in the past (Grossberg, 1978; Pezzulo and Levin, 2015; Spemann, 1967). We argue that conceptual advances in the links between machine learning and evolution now provide quantitative formalisms with which to begin to develop testable models of collective intelligence across scales. From subcellular processes, to cellular swarms during morphogenesis, to ecological dynamics on evolutionary timescales – all of these processes are driven by the scaling of reward dynamics that bind subunits into collectives that better navigate novel problem spaces.

In addition to shedding light on biological evolution, a better understanding of the origin and operation of collective intelligences would have a number of practical applications. Molecular medicine today is focused almost entirely on the micro-hardware of life – modifying DNA and rewiring molecular pathways – with limited success due to difficult inverse problems (Lobo et al., 2014). The capacity to manipulate the collective intelligence of cell groups might offer powerful ways to guide native and synthetic morphogenesis top-down (Pezzulo and Levin, 2016). Insights gleaned from biological systems could also significantly enhance the engineering of intelligent robots whose behaviour results from cooperation, competition and merging of subunits across multiple levels of organization.

Harnessing the native capability of collective intelligence in the service of biomedicine or bioengineering will require a much better understanding of how to identify, characterise and motivate emergent agents in anatomical, physiological and transcriptional spaces (Levin, 2022a; Pezzulo and Levin, 2015). As a starting point, we need to develop appropriate formalisms for top-down control of multi-scale intelligent agents of diverse composition. We argue that the tools and concepts of machine learning, behavioural neuroscience and evolutionary biology apply to problems of collective intelligence at multiple scales and offer a promising way forward.

There is a deep, fundamental symmetry between the origin of new evolutionary individuals from competent subunits and the assembly of an integrated cognitive agent as a collective intelligence composed of sub-agents. Future experimental and in silico work will quantitatively identify the necessary and sufficient relationships that effect such transitions. Such work has the potential to drive a flourishing sub-field of collective intelligence with implications ranging from basic evolutionary biology to regenerative medicine and artificial intelligence.

ORCID iD

Michael Levin https://orcid.org/0000-0001-7292-8084

References

Abramson CI (1994) *A Primer of Invertebrate Learning : The Behavioral Perspective*. 1st edition. Washington D.C.: American Psychological Association.

Baluška F and Levin M (2016) On having no head: cognition throughout biological systems. *Frontiers in Psychology* 7: 902. DOI: 10.3389/fpsyg.2016.00902.

Beane WS, Morokuma J, Lemire JM, et al. (2013). Bioelectric signaling regulates head and organ size during planarian regeneration [Research Support, N.I.H., Extramural Research Support, U.S Gov't, Non-P.H.S]. *Development*, 140(2), 313–322. DOI: 10.1242/dev.086900.

Bechtel W and Bich L (2021) Grounding cognition: heterarchical control mechanisms in biology. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376(1820): 20190751. DOI: 10.1098/rstb.2019.0751.

Best JB (1965) Behaviour of planaria in instrumental learning paradigms. *Animal Behaviour Supplement* 13(Suppl. 1): 69–75.

Birnbaum KD and Sánchez Alvarado A (2008) Slicing across kingdoms: regeneration in plants and animals. *Cell* 132(4): 697–710. DOI: 10.1016/j.cell.2008.01.040.

Biswas S, Manicka S, Hoel E, et al. (2021) Gene Regulatory Networks Exhibit Several Kinds of Memory: Quantification of Memory in Biological and Random Transcriptional Networks. *iScience* 24(3): 102131. DOI: 10.1016/j.isci.2021.102131.

Blackiston D, Lederer E, Kriegman S, et al. (2021) A cellular platform for the development of synthetic living machines. *Science Robotics* 6(52): eabf1571. DOI: 10.1126/scirobotics.abf1571.

Blackiston DJ and Levin M (2013) Ectopic eyes outside the head in Xenopus tadpoles provide sensory data for light-mediated learning. *The Journal of Experimental Biology* 216(Pt 6): 1031–1040. DOI: 10.1242/jeb.074963.

Brook Chernet ML and Levin M (2014) Endogenous voltage potentials and the microenvironment: bioelectric signals that reveal, induce and normalize cancer. *Journal of Clinical & Experimental Oncology* s1: S1. DOI: 10.4172/2324-9110.S1-002.

Brun-Usan M, Rago A, Thies C, et al. (2020a) Developmental models reveal the role of phenotypic plasticity in explaining genetic evolvability. *bioRxiv*. DOI: 10.1101/2020.06.29.179226.

Brun-Usan M, Thies C and Watson RA (2020b). How to fit in: The learning principles of cell differentiation. *PLOS Computational Biology*, 16(4), e1006811. DOI: 10.1371/journal.pcbi.1006811.

Bryant DM, Sousounis K, Farkas JE, et al. (2017) Repeated removal of developing limb buds permanently reduces appendage size in the highly-regenerative axolotl. *Developmental Biology* 424(1): 1–9. DOI: 10.1016/j.ydbio.2017.02.013.

Buckley CL, Lewens T, Levin M, et al. (in prep). Natural induction: spontaneous adaptive organisation in physical networks.

Caldwell JR, Watson RA, Thies C, et al. (2018) Deep optimisation: solving combinatorial optimisation problems using deep neural networks, p. 1811. arXiv.00784.

Campbell DT (2007) Adaptive-Behavior from Random Response. *Behavioral Science* 1(2): 105–110. DOI: 10.1002/bs.3830010204.

Campbell JO (2016) Universal Darwinism as a Process of Bayesian Inference. *Frontiers in Systems Neuroscience* 10: 49. DOI: 10.3389/fnsys.2016.00049.

Cervera J, Pietak A, Levin M, et al. (2018, Apr 21). Bioelectrical coupling in multicellular domains regulated by gap junctions: a conceptual approach. *Bioelectrochemistry*, 123, 45–61. DOI: 10.1016/j.bioelechem.2018.04.013.

Chastain E, Livnat A, Papadimitriou C, et al. (2014) Algorithms, games, and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 111(29): 10620–10623. DOI: 10.1073/pnas.1406556111.

Clarke E (2016) A levels-of-selection approach to evolutionary individuality. *Biology & Philosophy* 31(6): 893–911. DOI: 10.1007/s10539-016-9540-4.

Cooke J (1979) Cell number in relation to primary pattern formation in the embryo of *Xenopus laevis*. I: The cell cycle during new pattern formation in response to implanted organisers. *Journal of Embryology and Experimental Morphology* 51: 165–182. https://dev.biologists.org/content/51/1/165.full.pdf

Cooke J (1981). Scale of body pattern adjusts to available cell number in amphibian embryos. *Nature*, 290(5809), 775–778. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7219562 http://www.nature.com.ezp-prod1.hul.harvard.edu/nature/journal/v290/n5809/pdf/290775a0.pdf

Czegel D, Zachar I and Szathmáry E (2019) Multilevel selection as Bayesian inference, major transitions in individuality as structure learning. *Royal Society Open Science* 6(8): 190202. DOI: 10.1098/rsos.190202.

Davies A. P., Watson R. A., Mills R., et al. (2011) "If you can't be with the one you love, love the one you're with": how individual habituation of agent interactions improves global utility. *Artificial Life* 17(3): 167–181. DOI: 10.1162/artl_a_00030.

Deisboeck TS and Couzin ID (2009). Collective behavior in cancer cell populations. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 31(2), 190–197. DOI: 10.1002/bies.200800084.

Dinet C, Michelot A, Herrou J, et al. (2021) Linking single-cell decisions to collective behaviours in social bacteria. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376(1820): 20190755. DOI: 10.1098/rstb.2019.0755.

Doursat R, Sayama H and Michel O (2013). A review of morphogenetic engineering. *Natural Computing*, 12(4), 517–535. DOI: 10.1007/S11047-013-9398-1.

Draghi J and Wagner GP (2008). Evolution of evolvability in a developmental model. *Evolution; International Journal of Organic Evolution*, 62(2), 301–315. DOI: 10.1111/j.1558-5646.2007.00303.x.

Draghi J and Wagner GP (2009). The evolutionary dynamics of evolvability in a gene network model. *Journal of Evolutionary Biology*, 22(3), 599–611. DOI: 10.1111/j.1420-9101.2008.01663.x.

Durant F, Lobo D, Hammelman J, et al. (2016). Physiological controls of large-scale patterning in planarian regeneration: a molecular and computational perspective on growth and form. *Regeneration (Oxf)*, 3(2), 78–102. DOI: 10.1002/reg2.54.

Durant F, Morokuma J, Fields C, et al. (2017) Long-term, stochastic editing of regenerative anatomy via targeting endogenous bioelectric gradients. *Biophysical Journal* 112(10): 2231–2243. DOI: 10.1016/j.bpj.2017.04.011.

Ebrahimkhani MR and Levin M (2021). Synthetic living machines: A new window on life. *iScience*, 24(5), 102505. DOI: 10.1016/j.isci.2021.102505.

Egeblad M, Nakasone ES and Werb Z (2010). Tumors as organs: complex tissues that interface with the entire organism [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review]. *Developmental Cell*, 18(6), 884–901. DOI: 10.1016/j.devcel.2010.05.012.

Emmons-Bell M, Durant F, Hammelman J, et al. (2015) Gap junctional blockade stochastically induces different species-specific head anatomies in genetically wild-type girardia dorotocephala flatworms. *International Journal of Molecular Sciences* 16(11): 27865–27896. DOI: 10.3390/ijms161126065.

Evans CG, O'Brien J, Winfree E, et al. (2022) Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly. arXiv DOI: 10.48550/arXiv.2207.06399.

Fankhauser G (1945a) The effects of changes in chromosome number on amphibian development. *Quarterly Review of Biology* 20(1): 20–78. DOI: 10.1086/394703

Fankhauser G (1945b) Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. *The Journal of Experimental Zoology* 100(3): 445–455. DOI: 10.1002/jez.1401000310.

Fields C, Bischof J and Levin M (2020) Morphological coordination: a common ancestral function unifying neural and non-neural signaling. *Physiology (Bethesda)* 35(1): 16–30. DOI: 10.1152/physiol.00027.2019.

Fields C and Levin M (2018a). Are planaria individuals? what regenerative biology is telling us about the nature of multicellularity. *Evolutionary Biology*, 45(3), 237–247. DOI: 10.1007/s11692-018-9448-9.

Fields C and Levin M (2018b), Multiscale memory and bioelectric error correction in the cytoplasm-cytoskeleton-membrane system. *WIREs Systems Biology and Medicine*, 10(2), e1410-n/a, Article e1410. DOI: 10.1002/wsbm.1410.

Fields C and Levin M (2020). Scale-free biology: integrating evolutionary and developmental thinking. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 42(8), e1900228. DOI: 10.1002/bies.201900228.

Fields C and Levin M (2022) Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments. *Entropy (Basel)* 24(6): 819. DOI: 10.3390/e24060819.

Frank SA (2009) Natural selection maximizes fisher information. *Journal of Evolutionary Bi*ology 22(2): 231–244. DOI: 10.1111/j.1420-9101.2008.01647.x.

Friston K, Levin M, Sengupta B, et al. (2015) Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface* 12(105): 20141383. DOI: 10.1098/rsif.2014.1383.

Gawne R, McKenna KZ and Levin M (2020) Competitive and coordinative interactions between body parts produce adaptive developmental outcomes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 42(8): e1900245. DOI: 10.1002/bies.201900245.

Godfrey-Smith P (2009) *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.

Grossberg S (1978) Communication, memory, and development. In: Rosen R. and Snell F. (eds), *Progress in Theoretical Biology*, 5. https://www.semanticscholar.org/paper/Communication%2C-Memory%2C-and-Development-Grossberg/53641a9fd078df1fd314f441064a6eb3218b2339

Hansen TF (2006) The evolution of genetic architecture. *Annual Review of Ecology, Evolution, and Systematics* 37: 123–157. DOI: 10.1146/annurev.ecolsys.37.091305.110224.

Hansen TF, Álvarez-Castro JM, Carter AJR, et al. (2006, Aug). Evolution of genetic architecture under directional selection. *Evolution; International Journal of Organic Evolution*, 60(8), 1523–1536. DOI: 10.1111/j.0014-3820.2006.tb00498.x.

Harper M (2009). The replicator equation as an inference dynamic. arXiv, 0911.1763, Article arXiv:0911.1763.

Harris AK (2018). The need for a concept of shape homeostasis. *Bio Systems*, 173, 65–72. DOI: 10.1016/j.biosystems.2018.09.012.

Herrera-Delgado E, Perez-Carrasco R, Briscoe J, et al. (2018). Memory functions reveal structural properties of gene regulatory networks. *PLOS Computational Biology*, 14(2), e1006003. DOI: 10.1371/journal.pcbi.1006003.

Herrera-Rincon C, Guay JA and Levin M (2018) Bioelectrical coordination of cell activity toward anatomical target states: an engineering perspective on regeneration. In: Gardiner DM (ed), *Developmental Biology: Principles and Applications*. Boca Raton: CRC Press, pp. 55–112.

Hills TT, Todd PM, Lazer D, Cognitive Search Research Group, et al. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, 19(1), 46–54. DOI: 10.1016/j.tics.2014.10.004.

Hinton GE, Osindero S and Teh YW (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554. DOI: 10.1162/neco.2006.18.7.1527.

Hopfield JJ (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8), 2554–2558. DOI: 10.1073/pnas.79.8.2554.

Hopfield JJ and Tank DW (1986) Computing with neural circuits: a model. *Science* 233(4764): 625–633. DOI: 10.1126/science.3755256.

James W (1890) *The Principles of Psychology*. New York: H. Holt and company. http://catalog.hathitrust.org/api/volumes/oclc/1862859.htmlHathiTrust

Jones AG, Arnold SJ and Bürger R (2007). The mutation matrix and the evolution of evolvability. *Evolution; International Journal of Organic Evolution*, 61(4), 727–745. DOI: 10.1111/j.1558-5646.2007.00071.x.

Jones AG, Bürger R and Arnold SJ (2014) Epistasis and natural selection shape the mutational architecture of complex traits. *Nature Communications* 5(1): 3709. DOI: 10.1038/ncomms4709.

Kashtan N., Mayo A. E., Kalisky T., et al. (2009). An analytically solvable model for rapid evolution of modular structure. *PLOS Computational Biology*, 5(4), e1000355. DOI: 10.1371/journal.pcbi.1000355.

Keijzer F, van Duijn M and Lyon P (2013). What nervous systems do: early evolution, input-output, and the skin brain thesis. *Adaptive Behavior*, 21(2), 67–85. DOI: 10.1177/1059712312465330.

Kounios L, Clune J, Kouvaris K, et al. (2016) Resolving the paradox of evolvability with learning theory: How evolution learns to improve evolvability on rugged fitness landscapes. DOI: 10.48550/arXiv.1612.05955

Kouvaris K, Clune J, Kounios L, et al. (2017) How evolution learns to generalise: Using the principles of learning theory to understand the evolution of developmental organisation. *PLOS Computational Biology* 13(4): e1005358. DOI: 10.1371/journal.pcbi.1005358.

Kriegman S, Blackiston D, Levin M, et al. (2020) A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences of the United States of America* 117(4): 1853–1859. DOI: 10.1073/pnas.1910837117.

LeCun Y, Bengio Y and Hinton G (2015) Deep learning. *Nature* 521(7553): 436–444. DOI: 10.1038/nature14539.

Levin M (2014) Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. *The Journal of Physiology* 592(11): 2295–2305. DOI: 10.1113/jphysiol.2014.271940.

Levin M (2019). The computational boundary of a "self": developmental bioelectricity drives multicellularity and scale-free cognition [hypothesis and theory]. *Frontiers in Psychology*, 10, 2688. DOI: 10.3389/fpsyg.2019.02688.

Levin M (2021a). Life, death, and self: fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. *Biochemical and Biophysical Research Communications*, 564, 114–133. DOI: 10.1016/j.bbrc.2020.10.077.

Levin M (2021b) Bioelectric signaling: reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell* 184(8): 1971–1989. DOI: 10.1016/j.cell.2021.02.034.

Levin M (2021c) Bioelectrical approaches to cancer as a problem of the scaling of the cellular self. *Progress in Biophysics and Molecular Biology* 165: 102–113. DOI: 10.1016/j.pbiomolbio.2021.04.007.

Levin M (2022) Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. *Frontiers in Systems Neuroscience* 16: 768201. DOI: 10.3389/fnsys.2022.768201.

Levin M (2023) Collective Intelligence of Morphogenesis as a Teleonomic Process. In: Corning PA (ed), *Teleology*. Cambridge: MIT Press. https://psyarxiv.com/hqc9b/

Levin M, Keijzer F, Lyon P, et al. (2021) Uncovering cognitive similarities and differences, conservation and innovation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376(1821): 20200458. DOI: 10.1098/rstb.2020.0458.

Levin M, Pietak AM and Bischof J (2019). Planarian regeneration as a model of anatomical homeostasis: recent progress in biophysical and computational approaches. *Seminars in Cell and Developmental Biology*, 87, 125–144. DOI: 10.1016/j.semcdb.2018.04.003.

Lobo D, Solano M, Bubenik GA, et al. (2014) A linear-encoding model explains the variability of the target morphology in regeneration. *Journal of the Royal Society Interface* 11(92): 20130918. DOI: 10.1098/rsif.2013.0918.

Lyon P (2015) The cognitive cell: bacterial behavior reconsidered. *Frontiers in Microbiology* 6: 264. DOI: 10.3389/fmicb.2015.00264.

Lyon P (2020) Of what is "minimal cognition" the half-baked version? *Adaptive Behavior* 28(6): 407–424. DOI: 10.1177/1059712319871360.

Lyon P, Keijzer F, Arendt D, et al. (2021) Reframing cognition: getting down to biological basics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376(1820): 20190750. DOI: 10.1098/rstb.2019.0750.

Manicka S and Levin M (2019a) The cognitive lens: a primer on conceptual tools for analysing information processing in developmental and regenerative morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 374(1774): 20180369. DOI: 10.1098/rstb.2018.0369.

Manicka S and Levin M (2019b) Modeling somatic computation with non-neural bioelectric networks. *Scientific Reports* 9(1): 18612. DOI: 10.1038/s41598-019-54859-8.

Manicka S and Levin M (2022) Minimal developmental computation: a causal network approach to understand morphogenetic pattern formation. *Entropy (Basel)* 24(1): 107. DOI: 10.3390/e24010107.

Mathews J and Levin M (2017) Gap junctional signaling in pattern regulation: Physiological network connectivity instructs growth and form. *Developmental Neurobiology* 77(5): 643–673. DOI: 10.1002/dneu.22405.

Maynard Smith J and Szathmáry E (1997) *The Major Transitions in Evolution*. Oxford: Oxford University Press.

McCabe C, Watson R, Prichard J, et al. (2011). The web as an adaptive network: coevolution of web behavior and web structure. In 3rd International Web Science Conference, Koblenz, Germany, June 2011.

Michod RE (2000) *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton University Press.

Mills R (2010) *How Micro-Evolution Can Guide Macro-Evolution: Multi-Scale Search via Evolved Modular Variation*. Southampton: University of Southampton. https://eprints.soton.ac.uk/156549/

Mills R, Jansen T and Watson RA (2014). Transforming evolutionary search into higher-level evolutionary search by capturing problem structure. *Ieee Transactions on Evolutionary Computation*, 18(5), 628–642. DOI: 10.1109/Tevc.2014.2347702.

Moczek AP (2019) The shape of things to come: evo devo perspectives on causes and consequences in evolution. In: Uller T and Laland KN (eds), *Evolutionary Causation: Biological and Philosophical Reflections*. MIT Press. DOI: 10.7551/mitpress/11693.003.0005.

Nanos V and Levin M (2022) Multi-scale Chimerism: An experimental window on the algorithms of anatomical control. *Cells & Development* 169: 203764. DOI: 10.1016/j.cdev.2021.203764.

Nash F., Kounios L., Thies C., et al. (in prep) Scaling-up evolutionary variability: the causes and consequences of developmental hierarchy.

Newman SA and Bhat R (2008) Dynamical patterning modules: physico-genetic determinants of morphological development and evolution. *Physical Biology* 5(1): 015008. DOI: 10.1088/1478-3975/5/1/015008.

Okasha S (2006) *Evolution and the Levels of Selection*. Oxford: Oxford University Press. http://www.loc.gov/catdir/toc/ecip075/2006039679.html

Pai VP, Cervera J, Mafe S, et al. (2020) HCN2 channel-induced rescue of brain teratogenesis *via* local and long-range bioelectric repair. *Frontiers in Cellular Neuroscience* 14(136): 136. DOI: 10.3389/fncel.2020.00136.

Pai VP and Levin M (2022). HCN2 channel-induced rescue of brain, eye, heart and gut teratogenesis caused by nicotine, ethanol and aberrant notch signalling. *Wound Repair and Regeneration*. DOI: 10.1111/wrr.13032.

Pai VP, Pietak A, Willocq V, et al. (2018) HCN2 rescues brain defects by enforcing endogenous voltage pre-patterns. *Nature Communications* 9(1): 998. DOI: 10.1038/s41467-018-03334-5.

Palacios-Prado N and Bukauskas FF (2009) Heterotypic gap junction channels as voltage-sensitive valves for intercellular signaling. *Proceedings of the National Academy of Sciences of the United States of America* 106(35): 14855–14860. DOI: 10.1073/pnas.0901923106.

Parter M, Kashtan N and Alon U (2008) Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLOS Computational Biology* 4(11): e1000206. DOI: 10.1371/journal.pcbi.1000206.

Pavličev M and Cheverud JM (2015) Constraints evolve: context dependency of gene effects allows evolution of pleiotropy. *Annual Review of Ecology, Evolution, and Systematics* 46: 413–434. DOI: 10.1146/annurev-ecolsys-120213-091721.

Pezzulo G, LaPalme J, Durant F, et al. (2021) Bistability of somatic pattern memories: stochastic outcomes in bioelectric circuits underlying regeneration. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376(1821): 20190765. DOI: 10.1098/rstb.2019.0765.

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. *Integrative Biology: Quantitative Biosciences from Nano to Macro*, 7(12), 1487–1517. DOI: 10.1039/c5ib00221d.

Pezzulo G and Levin M (2016). Top-down models in biology: explanation and control of complex living systems above the molecular level. *Journal of the Royal Society Interface*, 13 (124), 20160555. DOI: 10.1098/rsif.2016.0555.

Pietak A and Levin M (2016) Exploring instructive physiological signaling with the bioelectric tissue simulation engine. *Frontiers in Bioengineering and Biotechnology* 4: 55. DOI: 10.3389/fbioe.2016.00055.

Pietak A and Levin M (2017). Bioelectric gene and reaction networks: computational modelling of genetic, biochemical and bioelectrical dynamics in pattern regulation. *Journal of the Royal Society Interface*, 14, 20170425(134). DOI: 10.1098/rsif.2017.0425.

Power DA (2019) *Distributed Associative Learning in Ecological Community Networks*. Southampton: University of Southampton.

Power DA, Watson RA, Szathmáry E, et al. (2015). What can ecosystems learn? Expanding evolutionary ecology with learning theory. *Biology Direct*, 10(1), 69. DOI: 10.1186/s13062-015-0094-1.

Radisky D, Hagios C and Bissell MJ (2001) Tumors are unique organs defined by abnormal signaling and context. *Seminars in Cancer Biology* 11(2): 87–95. DOI: 10.1006/scbi.2000.0360.

Rago A, Kouvaris K, Uller T, et al. (2019) How adaptive plasticity evolves when selected against. *PLOS Computational Biology* 15(3): e1006260. DOI: 10.1371/journal.pcbi.1006260.

Rescorla RA and Solomon RL (1967) Two-process learning theory: relationships between pavlovian conditioning and instrumental learning. *Psychological Review* 74(3): 151–182. DOI: 10.1037/h0024475.

Rubenstein M, Cornejo A and Nagpal R (2014) Robotics. Programmable self-assembly in a thousand-robot swarm. *Science* 345(6198): 795–799. DOI: 10.1126/science.1254295.

Shalizi CR (2009) Dynamics of Bayesian updating with dependent data and misspecified models. *Electronic Journal of Statistics* 3: 1039–1074. DOI: 10.1214/09-Ejs485.

Skinner BF (1981) Selection by consequences. *Science* 213(4507): 501–504. DOI: 10.1126/science.7244649.

Slavkov I, Carrillo-Zapata D, Carranza N, et al. (2018) Morphogenesis in robot swarms. *Science Robotics* 3(25): eaau9178. DOI: 10.1126/scirobotics.aau9178.

Sonnenschein C and Soto AM (1999) *The Society of Cells : Cancer Control of Cell Proliferation*. Berlin: Springer.

Soto AM, Sonnenschein C and Miquel PA (2008) On physicalism and downward causation in developmental and cancer biology. *Acta Biotheoretica* 56(4): 257–274. DOI: 10.1007/s10441-008-9052-y.

Spemann H (1967) *Embryonic Development and Induction*. London: Yale University Press; H. Milford Oxford University Press.

Stern M and Murugan A (2022) *Learning without Neurons in Physical Systems*. arXiv DOI: 10.48550/arXiv.2206.05831.

Stone JR (1997). The spirit of D'arcy Thompson dwells in empirical morphospace [Research Support, Non-U.S. Gov't Review]. *Mathematical Biosciences*, 142(1), 13–30. http://www.ncbi.nlm.nih.gov/pubmed/9125858 http://ac.els-cdn.com/S0025556496001861/1-s2.0-S0025556496001861-main.pdf?_tid=f4529142-2748-11e2-9812-00000aacb362&acdnat=1352120686_7803a1a866f8fd8d892db6cfa8bf6fcd

Sullivan KG and Levin M (2016) Neurotransmitter signaling pathways required for normal development in Xenopus laevis embryos: a pharmacological survey screen. *Journal of Anatomy* 229(4): 483–502. DOI: 10.1111/joa.12467.

Szabó Á, Vattay G and Kondor D (2012). A cell signaling model as a trainable neural nanonetwork. *Nano Communication Networks*, 3(1), 57–64. http://www.sciencedirect.com/science/article/pii/S1878778912000038

Tareen A and Kinney JB (2020) *Biophysical Models of Cis-Regulation as Interpretable Neural Networks*. DOI: 10.1101/835942.

Valiant L (2013) *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. Basic Books.

Vanchurin V, Wolf YI, Katsnelson MI, et al. (2021) *Towards a Theory of Evolution as Multilevel Learning*. ArXiv. arXiv:2110.14602.

Vandenberg LN, Adams DS and Levin M (2012). Normalized shape and location of perturbed craniofacial structures in the Xenopus tadpole reveal an innate ability to achieve correct morphology. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 241(5), 863–878. DOI: 10.1002/dvdy.23770.

Watson JB (1967) *Behavior; An Introduction to Comparative Psychology*. Holt.

Watson RA (accepted). Agency, goal-directed behaviour and part-whole relationships in biological systems. *Biological Theory*.

Watson RA, Buckley CL and Mills R (2011a). Optimization in "self-modeling" complex adaptive systems. *Complexity*, 16(5), 17–26. DOI: 10.1002/cplx.20346.

Watson RA, Levin M and Buckley CL (2022) Design for an individual: connectionist approaches to the evolutionary transitions in individuality [hypothesis and theory]. *Frontiers in Ecology and Evolution* 10. DOI: 10.3389/fevo.2022.823588.

Watson RA, Levin ML, Buckley CL, et al. (in review). An ability to respond begins with inner alignment: How phase synchronisation effects transitions to higher levels of agency.

Watson RA, Mills R and Buckley CL (2011b) Global adaptation in networks of selfish components: emergent associative memory at the system scale. *Artificial Life* 17(3): 147–166. DOI: 10.1162/artl_a_00029.

Watson RA, Mills R and Buckley CL (2011c) Transformations in the scale of behavior and the global optimization of constraints in adaptive networks. *Adaptive Behavior* 19(4): 227–249. DOI: 10.1177/1059712311412797.

Watson RA, Mills R, Buckley CL, et al. (2016) Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evolutionary Biology* 43(4): 553–581. DOI: 10.1007/s11692-015-9358-z.

Watson RA, Palmius N, Mills R, et al. (2009) Can selfish symbioses effect higher-level selection? In: Kampis G., Karsai I. and Szathmáry E. (eds), *Advances in Artificial Life. Darwin Meets von Neumann European Conference on Artificial Life 2009*. Budapest, Hungary.

Watson RA and Szathmáry E (2016). How can evolution learn? *Trends in Ecology & Evolution*, 31(2), 147–157. DOI: 10.1016/j.tree.2015.11.009.

Watson RA and Thies C (2019) Are developmental plasticity, niche construction, and extended inheritance necessary for evolution by natural selection? the role of active phenotypes in the minimal criteria for darwinian individuality. In: Uller T and Laland KN (eds), *Evolutionary Causation: Biological and Philosophical Reflections*. MIT Press.

Watson RA, Wagner GP, Pavličev M, et al. (2014, Apr). The evolution of phenotypic correlations and "developmental memory". *Evolution; International Journal of Organic Evolution*, 68(4), 1124–1138. DOI: 10.1111/evo.12337.

Weinreich DM, Watson RA and Chao L (2005) Perspective:Sign Epistasis and Genetic Constraint on Evolutionary Trajectories. *Evolution; International Journal of Organic Evolution* 59(6): 1165–1174. DOI: 10.1554/04-272.

West SA, Fisher RM, Gardner A, et al. (2015) Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences of the United States of America* 112(33): 10112–10119. DOI: 10.1073/pnas.1421402112.

Appendix

Box 1: The Hopfield network and collective behaviour

The Hopfield network is a neural network model (and a general model of dynamical systems in many domains from ferro-magnets to ecological communities) described by a set of nodes (either binary threshold units or sigmoidal continuous response units) connected to each other with symmetric connections (and no self-connections). Each node 'fires' if the weighted sum of inputs from other nodes is sufficiently strong. Given that some weights can be zero, this is a fairly general concept of a dynamical system described by a set of interactions between variables (Figure A1). The special qualification that the weights are symmetric (and no self-weights) is important, however, because it means that the dynamics can be described by the local minimisation of an energy function and the attractors of the dynamics are fixed points (i.e. state configurations where no units change value).

One interesting behaviour of this kind of network is the ability to store multiple patterns of activation in the connections of the network and generate, recognise or recall stored patterns through associative memory. A pattern (such as an image, or a set of features, describing a food type, habitat or a predator) can be stored by setting the units to match the (signed) pattern values and then applying Hebbian learning to the weights such that a change in the connection between two neurons is proportional to the product of the state values (a.k.a. *neurons that fire together wire together*). This kind of change to a connection makes it easier (lower energy) for the two states it connects to fire together in future. For example, if both states are firing at the same time, the connection strength is increased, meaning that activation in one stimulates activation in the other, making it more likely that they both activate together in future. This has the effect of lowering the energy of this configuration, drawing the network state towards this pattern in future – that is, forming a memory of the pattern. One network can store multiple patterns simultaneously, and

stored patterns can be generated from the network (from the set of patterns that it has stored) by initialising the state values at random and running the network to an attractor. A given pattern can also be recalled or recognised by presenting a partial or noisy input – causing the network to complete or recreate the entire/uncorrupted pattern that was closest to this stimulus (i.e. a content-addressable memory (Hopfield, 1982). Over a set of patterns stored in this way, connections in the network model the correlations (commonly occurring combinations) of state values. This 'associative model' of past state configurations can also generalise – for example, generate a pattern that has the same underlying structural relationships as those observed during learning but is nonetheless novel, that is, different from any specific pattern observed during training.

These networks can also exhibit problem-solving behaviour. If the connections of the network correspond to the constraints of a problem (i.e. the agreement or disagreement of two variables confers a change in solution quality proportional to the magnitude of the weight between them), then the natural dynamics of the state variables is to change in a manner that decreases violated constraints, causing the network to discover locally optimal solutions to the problem (Hopfield and Tank, 1986). Moreover, under certain conditions, the addition of relatively slow Hebbian learning to the weights, applied whilst the state variables visit a distribution of such locally optimal solutions, causes the network to form an associative memory of its own behaviour (a 'self-modelling dynamical system' (Buckley et al., in prep; Watson et al., 2011a), or a memory of the locally optimal solutions it visits. Because this associative memory can generalise, it can change its own dynamics in a manner that improves the ability of the network to resolve problem constraints, and with positive feedback, it can thus learn to discover high-quality solutions more reliably over time (reinforcement) and also find solutions that are better than any solutions found before the application of such learning (i.e. true optimisation).

Note that the memory, recall/recognition and problem-solving behaviour of the network, and the learning
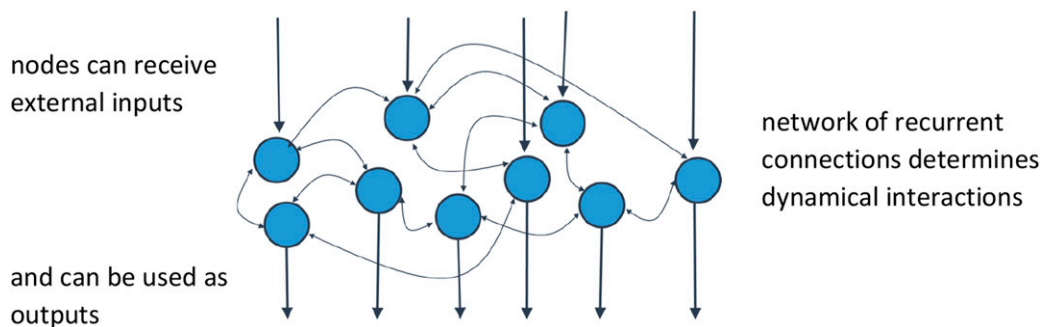

Figure A1. Hopfield network architecture.

mechanisms that organise the connections to achieve this, are fully distributed and decentralised. During recall, each neuron fires if its inputs are strong enough, without centralised control. And during learning, the update to each connection is proportional to the product of activation in the two neurons it connects, without reference to any global feedback, performance measure or testing of consequences from this change. Crucially, these recognition/recall and problem-solving behaviours can be exhibited by the network as a whole but cannot be exhibited by the individual components therein (nor explained by any average or sum of their individual behaviours). Neither do these new system-level behaviours result from changes to the behaviours of individual units but only from a change to the organisation of connections between them. These observations are important for collective intelligence for the following reasons. Where individuals have behaviours that are sensitive to the behaviours of others, adjustments are made to the organization of these relationships, either in terms of their selective strength (rHN-s, Watson et al., 2011a), their generation of variability (rHN-g, Watson et al., 2011c), or their timing (rHN-t, Watson et al., in prep). Such adjustments, which are made using only local information, are sufficient to produce non-trivial collective behaviours (collective memory, recognition, learning, generalization and problem solving) without centralized control or global feedback on performance.