BioSystems 235 (2024) 105107

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Short communication

From reinforcement learning to agency: Frameworks for understanding basal cognition

Gabriella Seifert [a,c], Ava Sealander [b,c], Sarah Marzen [c,*], Michael Levin [d,e]

[a] Department of Physics, University of Colorado, Boulder, CO 80309, USA
[b] Department of Electrical Engineering, School of Engineering and Applied Sciences, Columbia University, New York, NY 10027, USA
[c] W. M. Keck Science Department, Pitzer, Scripps, and Claremont McKenna College, Claremont, CA 91711, USA
[d] Department of Biology, Tufts University, Medford, MA 02155, USA
[e] Allen Discovery Center at Tufts University, Medford, MA 02155, USA

ARTICLE INFO

Keywords:
AI
Reinforcement learning
Machine learning
Agency
Goal-directedness
Teleonomy

ABSTRACT

Organisms play, explore, and mimic those around them. Is there a purpose to this behavior? Are organisms just behaving, or are they trying to achieve goals? We believe this is a false dichotomy. To that end, to understand organisms, we attempt to unify two approaches for understanding complex agents, whether evolved or engineered. We argue that formalisms describing multiscale competencies and goal-directedness in biology (e.g., TAME), and reinforcement learning (RL), can be combined in a symbiotic framework. While RL has been largely focused on higher-level organisms and robots of high complexity, TAME is naturally capable of describing lower-level organisms and minimal agents as well. We propose several novel questions that come from using RL/TAME to understand biology as well as ones that come from using biology to formulate new theory in AI. We hope that the research programs proposed in this piece shape future efforts to understand biological organisms and also future efforts to build artificial agents.

1. Introduction

Nature offers many remarkable and inspiring examples of complex structure and function. A paradigmatic example is developmental morphogenesis: a single cell (the fertilized egg) reliably gives rise to a body with exquisite multiscale anatomical order, ranging across body-plans that includes trees, snakes, elephants, and so on. It is largely assumed that this can be understood via the concept of emergent complexity: simple rules governing the behavior of molecular pathways and cells, when executed iteratively by large numbers of local agents, result in complex outcomes (Furusawa and Kaneko, 2002; Halley et al., 2012) such as the aforementioned trees, snakes, and elephants. This kind of behavior is readily observed in cellular automata and other workhorse conceptual tools of complexity science. Importantly, however, this approach has largely not been able to fill two key gaps. First, the inverse problem of deriving low-level interventions that implement a desired system-level goal (Groetsch and Groetsch, 1993) strongly limits advances in regenerative medicine and bioengineering — emergent models make it very hard to know what to change at the level of subunits to get desirable outcomes in anatomy and behavior (Lobo et al., 2014). Second, these formalisms do not address aspects of biological regulation that pertain to flexible problem-solving: robustness, context-sensitive plasticity, and top-down controls via anatomical homeostasis that are able to achieve their adaptive objectives despite changes of composition and environment (Pezzulo and Levin, 2016; Birnbaum and Alvarado, 2008; Levin, 2023b).

Many phenomena in biology are not simply feed-forward (open loop) outcomes of emergence, but rather exhibit remarkable capacity to adjust large-scale outcomes to novel circumstances. For example, early mammalian embryos cut in half give rise to normal monozygotic twins, as each half rebuilds its missing components. This is a special case of a more general phenomenon of regeneration, in which some species' bodies are able to recognize missing structures and activate rapid cell proliferation and remodeling until the correct structure is complete. Salamanders can regenerate their eyes, jaws, limbs, tails (including spinal cord) (McCusker and Gardiner, 2011), while planarian flatworms regenerate any part of their body even from small fragments (Saló et al., 2009). Crucially, the ability to recognize what is missing, construct exactly the structures that are needed, and then stop when the correct target morphology is complete, is a kind of anatomical homeostasis. This efficient error minimization loop is still not well-understood with respect to how the system measures complex state and solves the means-ends problem to reduce distance from the correct target morphology, although progress has been made with

* Corresponding author.
E-mail address: smarzen@cmc.edu (S. Marzen).

https://doi.org/10.1016/j.biosystems.2023.105107
Received 12 September 2023; Received in revised form 17 December 2023; Accepted 17 December 2023
Available online 19 December 2023

respect to the mechanisms that store the anatomical setpoint (Levin, 2023a). Another example is clearly shown by tadpoles, which have to significantly rearrange their faces to become frogs. It was found that if frog embryos are made with faces in a scrambled configuration (eyes, jaws, mouth, etc., in incorrect starting positions), the resulting frogs can be largely normal (Vandenberg et al., 2012) because the organs move in novel, unnatural paths until they get to the correct places and then they stop (Pinet and McLaughlin, 2019). Thus, the genetics does not specify hardwired movements that turn standard tadpoles into standard frogs — instead, it specifies cellular hardware that is able to execute a flexible corrective scheme that implements a kind of means-ends process relative to an anatomical setpoint (Harris, 2018; Levin et al., 2019).

The ability of biological systems to respond to novel conditions goes even deeper than subtractive injury or abnormal starting states. When cells of the newt are artificially increased in size, the resulting animals are normal, showing adjustment and rescaling of organs to a smaller number of cells per structure. The most amazing aspect is the kidney tubule, which in cross-section normally consists of 8 cells working together. When the cells are made bigger in experiments, fewer and fewer cells cooperate to make the same diameter tubules, until the cells are made extremely huge, at which point just one cell wraps around itself to make a lumen of the correct size (Fankhauser, 1945). This example shows that diverse molecular mechanisms (cell–cell communication vs. cytoskeletal bending) can be called up in the service of a largescale anatomical goal. But even the large-scale goals of living forms can be altered on-the-fly, and it does not require changes of the genome. Planarian flatworms can be turned into animals that always produce two heads upon damage (Durant et al., 2017; Oviedo et al., 2010), or indeed produce heads belonging to other species of worms (Emmons-Bell et al., 2015), by a transient modification to the bioelectric memory pattern that encodes their target morphology (Durant et al., 2016; Levin et al., 2019), without transgenes or mutation. Similarly, wild-type skin cells liberated from the instructive influence of their neighbors reboot their multicellularity toward a new motile form: Xenobots (Kriegman et al., 2020): proto-organisms which exhibit novel behaviors (including kinematic self-replication (Kriegman et al., 2021)) and healing after damage to their new Xenobot form.

All of these capacities have been suggested to be the results of the collective intelligence of cellular swarms solving problems in morphospace (Fields and Levin, 2022; Levin, 2022b). Indeed, the definition of intelligence by William James – "same ends through different means"– wisely focuses attention not on anatomical markers (such as brains) or specific material implementations but on the functional invariant of all intelligent agency: goal-directed activity with some degree of competency at handling novelty. And it is not only about anatomical morphospace. Life solves problems in numerous spaces–metabolic, transcriptional, physiological, etc. in addition to the familiar behavioral (3D) space in which conventional intelligences are readily recognized. For example, planaria exposed to barium experience degeneration of their heads, as the barium ion blocks potassium channels and poisons cells. However, within a week or two of living in barium, planarian tails regenerate new heads that are completely adapted to barium: transcriptomic analysis shows that out of their large genomes, the cells have identified just a handful of genes to up- and down-regulate to solve their problem (Emmons-Bell et al., 2019). Planaria do not experience barium in the wild (and thus do not have a specifically selected-for response pathway), nor do their cells turn over fast enough to enable a hill-climbing search by selection through immense numbers of cells that try different solutions (the astronomical set of all possible gene expression responses) and repopulate the head through differential survival. This phenomenon illustrates the still poorly-understood ability of cells to efficiently navigate the very high-dimensional transcriptional space to solve novel physiological problems.

The examples discussed above reveal that evolution does not just result in machines hardwired to address specific problems (solutions fit to specific environments). Long periods of adaptation are not needed to make normal frogs out of divergent tadpole facial configurations, make kidney tubules with much larger cells, or create functional Xenobots. Instead, what evolution makes is hardware that is able to deploy degrees of intelligent problem-solving in a wide range of spaces and novel contexts. This is a critical aspect of living material that, if harnessed, would catapult biomedicine beyond the current limitations of genomics, molecular biology, and stem cell approaches, which are limited by their exclusive focus on the hardware of life. As is well-known in computer science, programming at the level of rewiring the hardware is just the beginning of what is possible.

Any of the examples described above would be considered transformative advances in intelligence if implemented as swarm robotics or AI with such capacities. We propose that a focus on the intelligence of biological systems, generalizing far beyond the familiar neural substrates and behavior, can advance an exciting emerging field at the intersection of the biological and information sciences (Lyon, 2015, 2006). Machine learning, particularly reinforcement learning, will be enriched by new models of generalization and robust plasticity that are not limited by neuromorphic ideas: learning from evolutionarily-ancient problem-solving strategies that are more relevant to generalized intelligence. Conversely, biology and regenerative medicine will be advanced by importing quantitative ideas from the field of AI, to enhance the understanding of how evolution enables multiscale competency and the search for efficient interventions that exert control over large-scale growth and form in biomedical settings (Lagasse and Levin, 2023).

What is needed now are substrate-independent theoretical tools for understanding multi-scale learning, with enough specificity to drive constructive models (executable code). Here, we first provide a brief introduction to a conceptual framework for thinking about multiscale competencies in biology (TAME) and to reinforcement learning (RL). We then argue that RL is an ideal toolkit for quantifying the key aspects of the TAME framework, and is ready to be applied to problems in multiscale biology. We intend this cross-fertilization of fields to occur in a deep way — RL provides not only a tool for prediction and data analysis, but a formalism for generating insights and actionable algorithms that enable control of complex biological systems. While RL is a powerful and vibrant field, we envision even more vistas of application to, and inspiration from, biology (Neftci and Averbeck, 2019). At stake are better bio-inspired machine learning algorithms, more effective biomedicine (by understanding how tissues can be controlled by stimuli that alter learned responses, not micromanaged), evolutionary methods for design of individual and swarm robotics, and synthetic bioengineering. RL methods can help understand how biology at all scales (pathways, cells, tissues, etc.) learns about itself and its environment, greatly expanding our view of how extant forms arose, and what else is possible via engineering and chimeric technologies in the vast space of life-as-it-can-be (Clawson and Levin, 2022; Langton, 2019).

2. Background

2.1. Introduction to TAME

TAME (Technological Approach to Mind Everywhere) is a framework (Levin, 2022b) designed to facilitate experimental approaches to detecting, understanding, and functionally interacting with diverse intelligences (natural and artificial). It is most well-developed currently around the example of the collective intelligence of biological cells navigating morphospace to solve anatomical tasks during embryogenesis, regeneration, and cancer suppression (Levin, 2019). It focuses on the scaling up of simple homeostatic functions of subcellular components and cells into tissues, organs, and whole organisms via dynamics that allow swarms of agents with small cognitive horizons to assemble into larger-scale agents with larger goals in new problem spaces (Levin, 2019). Its fundamental features include:

- Goal-directed behavior as the primary invariant of all intelligent systems — a functionalist, cybernetic perspective (Rosenblueth et al., 1943) that avoids distinctions based on origin story (evolved vs. engineered) or material composition (protoplasm, silicon, etc.). TAME views such distinctions (e.g., using typical brains as a marker) as the relics of past limitations of technology and highly contingent frozen accidents of the evolutionary trajectory on Earth; the utility of sharp distinctions will not survive the coming decades as chimeric, synthetic, and biohybrid engineering technologies erase remaining distinctions between life and machines.
- A focus on empirical, observer-dependent estimates of agency of any given system based on the optimal efficiency of models needed to predict and control that system by some observer, which also includes the system itself (Bongard and Levin, 2023). This view emphasizes experiments to determine where a system is best placed on a continuum of persuadability (Fig. 1), not philosophical preconceptions of how much cognition should be attributed to a system based on its provenance or structure, or privileged scales of observation (spatial or temporal) which obscure agency in unfamiliar guises and problem spaces (Fields and Levin, 2022).
- A commitment to a continuum of diverse cognitive sophistication; because of gradual evolution and embryonic development, cognitive beings with minds arise from a single cell. Thus, the journey from physics to mind is continuous. Biology offers no support to any bright line separating "true cognitive beings" from "machines that are faking sentience". The ability to make chimeras integrating living tissue, engineered electronics, and software in any configuration means that the space of possible bodies and minds is astronomically vast (Clawson and Levin, 2022), requiring us to develop deep concepts that do not rely on artificial binary categories and quantify the degree and kind of intelligence in any given system. TAME is a framework that is ideally suited to dealing with agents that change over time as it recognizes that cognitive media (whether brains in metamorphosing insects, or refactored digital hardware) can be altered on-the-fly, implying that we need to understand not only the perspective of constant, well-defined agents but also of ones that are splitting, merging, and changing during the agent's lifetime (Blackiston et al., 2015).
- A recognition that biological systems consist of nested modules that are themselves goal-seeking agents. Bodies are made of organs which are made of tissues which are made of cells which are made of molecular networks; more important than the structural modularity is the fact that each of these levels consists of active agents that themselves have goals and are solving problems in various spaces (physiological, metabolic, transcriptional, anatomical, and behavioral). This multiscale competency architecture is proposed to be the source of biology's incredible robustness and problem-solving ability.

In the TAME framework, an agent's Self is determined by the system-level goals it can pursue (Fig. 1). A Self is first and foremost the subject of preferences (which may be as simple as the setpoint of a homeostatic loop) and has the ability to take action in some problem space (and expend energy) to reduce the delta between a current state and a target goal state as is typical of a control theory (Kirk, 2004) setup, though other capabilities can be included as well. An agent's degree of sophistication is the size, in space and time, of the biggest goal states it can possibly represent and work toward; more generally, an agent's degree of sophistication is the modeling and behavioral apparatus associated with the actions that it takes. This demarcates that agent's cognitive horizon; agents' boundaries can shift (grow or shrink) as the result of events that modify the agent's goal space, and agents can exist as nested (multi-scale) systems with cooperation and competition within and across scales.

TAME has been applied to a wide variety of biological phenomena and has made numerous predictions that are successfully guiding novel work at the interaction of regenerative biology and cognitive science. For example, it facilitates the use of techniques from behavioral neuroscience to understand and control morphogenesis as the behavior of the collective intelligence of cellular swarms in anatomical morphospace, as well as the processes during which normal morphogenesis or morphostasis (integration of cells into a network that pursues organ-level goals) falls apart into cancer (when individual cells disconnect from the information network and allow their goals to recede back to the evolutionarily ancient unicellular goals of metabolism and proliferation) (Levin, 2023a; Pezzulo and Levin, 2015). However, TAME v1.0 is largely qualitative. What is needed is rigorous development of theory quantifying key dimensions of TAME, including the space of goals, preferences, and algorithms by which the homeostatic loops of individual agents are coupled to result in collectives capable of pursuing much larger-scale goals.

Specifically, one open area is to understand how components work together to allow the emergent higher-order agent to acquire memories and skills, and work in a problem space, that belong to it, and not to any of its parts. For example, once "a rat" has learned to press a lever to get a reward, the associative memory belongs not to the skin cells of the paw that interacted with the lever, nor to the gut cells which received the nutritious reward, but to the collective since no single cell ever had the experience of both events. Because learning is observed across the web of life — from chemical networks (Biswas et al., 2021; Watson et al., 2010) and microbes (Baluška and Levin, 2016; Boussard et al., 2019; Wolf et al., 2008; Yang et al., 2020) to entire ecosystems (Power et al., 2015), this kind of credit assignment among the parts is an essential aspect of understanding the origin and function of composite, emergent agents (Watson et al., 2022). Reinforcement learning (RL) (Sutton and Barto, 2018) is one powerful tool to advance this set of questions.

Fundamental open questions include the algorithms guiding the computations within the composite agent, and credit assignment among diverse parts, all of which enables learning by collective **intelligences** (Couzin, 2007; Solé et al., 2016). Here, we suggest that the field of reinforcement learning provides computational tools to extend TAME in the necessary ways, as well as itself benefiting from aspects of this biological framework – such as the response to novel perturbations of the environment – that have not yet received sufficient attention in engineering.

2.2. Introduction to reinforcement learning: an ideal complement to TAME

Briefly, reinforcement learning (RL), a subfield of machine learning (ML), is a mathematical formalization that can greatly assist the TAME framework, and more broadly, efforts to recognize, manipulate, and build agents in novel embodiments. Specifically, it provides a simple mathematical formalization of the notion of a goal-directed agent.

The most common mathematical formulation is that of a Markov Decision Process (MDP), in which the environment and agent and their interaction are defined by a few quantities: the reward function $r(s, a)$ that specifies how much reward the agent gets when the environment is in state s and the agent takes action a; the transition probabilities $p(s'|s, a)$ that specifies how likely it is for the environment to transition to state s' when it is in state s previously and when the agent takes an action a; a discount factor γ that specifies how much less valued rewards are when they are not given immediately; and the agent's action policy, $\pi(a|s)$, that specifies how likely it is for the agent to take action a when the world is in state s. Other related frameworks also allow for time delays and partial observations of the world state (Bertsekas, 2012; Katsikopoulos and Engelbrecht, 2003; Sawaya et al., 2023). A reinforcement learning agent is any agent that attempts to "solve the MDP", or maximize the total sum of discounted rewards over the course of its lifetime, which could be arbitrarily long. Solving the MDP leads

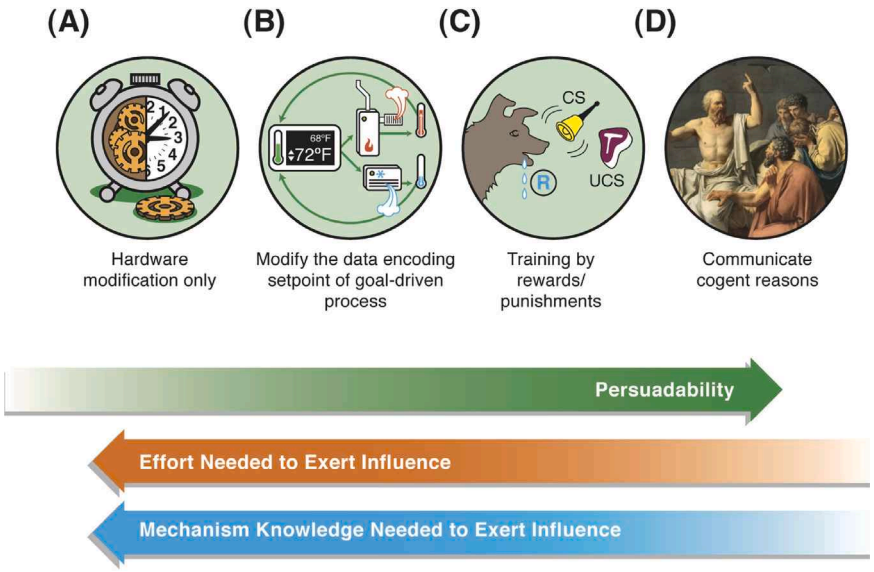

Fig. 1. The axis of persuadability. This is a visualization of a continuum of agency, framed from the perspective of an engineer (or a biological agent) seeking to control some system. What kind of techniques provide optimal control for that system? Here are shown only a few representative waypoints. On the far left are the simplest physical systems, e.g., mechanical clocks (A). These cannot be persuaded, argued with, or even rewarded/punished—only physical hardware-level "rewiring" is possible if one wants to change their behavior. On the far right (D) are human beings (and perhaps others to be discovered) whose behavior can be radically changed by a communication that encodes a rational argument that changes the motivation, planning, values, and commitment of the agent receiving this — it relies heavily on the high cognitive competency of the system. Between these extremes lies a rich and diverse set of intermediate agents, such as simple homeostatic circuits (B) which have setpoints encoding goal states, and more complex systems such as animals which can be controlled by signals, stimuli, training, etc., (C). They can have some degree of plasticity, memory (change of future behavior caused by past events), various types of simple or complex learning, anticipation/prediction, etc.

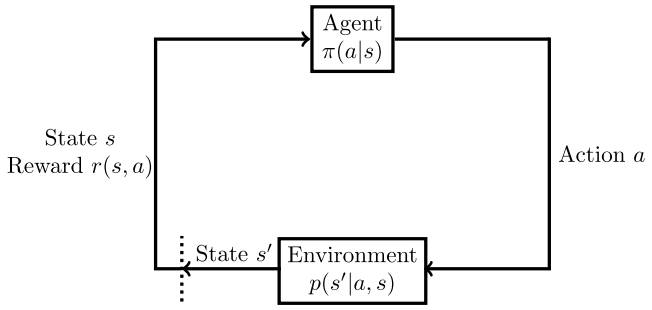

Fig. 2. The reinforcement learning agent (organism) takes actions a based on an action policy $\pi(a|s)$ and moves to new states s of the environment, collecting rewards as it moves. The environment changes states according to a transition probability $p(s'|a,s)$, which can be influenced by the actions of the agent. Its goal is to maximize a discounted sum total of rewards collected over its lifetime.

to an action policy that maximizes the sum of discounted rewards. In TAME, this could for example then lead to $x(t)$ matching the target $\hat{x}(t)$ as well as possible, and this policy is called the "optimal action policy" (see Fig. 2).

The actions that the agent takes now affect the state that it is in many timesteps later, and as a result, the agent must effectively plan over long time horizons. Explicitly doing this results in a curse of dimensionality: the agent must keep track of trajectories, the number of which increases exponentially with the number of timesteps. Furthermore, rewards might be sparse, so the agent only receives a reward at the end of its lifetime or at a few points during its lifetime, e.g. number of viable children. This makes learning difficult over the lifetime of the organism, but not over the span of many generations. There are, however, canonical methods for solving the MDP that might inform the TAME framework and, in a wider scope, provide inspiration for ways to think about how biological agents are able to achieve specific goals despite significant perturbations (robust problem-solving in diverse spaces (Fields and Levin, 2022)).

RL researchers have identified two different approaches to solving the MDP: model-free and model-based (Sutton and Barto, 2018). In model-free approaches, there is no explicit model in the agent of the environment, i.e. no explicit recognition of what the transition probability $p(s'|s,a)$ might be. Instead, the agent implicitly has a model of the environment that it uses via dynamic programming to choose the optimal action policy. Dynamic programming is a clever way of dealing with a curse of dimensionality associated with the exponential increase in the number of possible trajectories as one's lifetime increases, recursively relating the value of the current state to the value of the next states. The details of this are a Bellman equation that allows one to find a state–action value function that gives the "value" of taking an action a when the world is in state s. There are various ways of finding this state–action value function, all with benefits and drawbacks, such as Monte Carlo approaches, temporal difference learning, or the TD-λ approaches that interpolate between the two. Model-based approaches to finding the optimal action policy have an explicit model of the environment that they use to simulate long trajectories of states and actions, which they then use to choose the optimal action policy. The newer successor representation (Gershman, 2018) sits between these two extremes, and it is based on understanding the transitions from one state to another state in the environment and using that to calculate the value of a state.

One familiar touchstone for such approaches is the brain — an uncontroversial example of a collection of cells that solves the credit assignment problem for an emergent cognitive agent. There is evidence for brains using model-free, model-based, and successor representation approaches. Famously, there is evidence that dopaminergic signals are related to temporal difference learning (Schultz et al., 1997), though see Ref. Jeong et al. (2022). Serotonin is thought to be associated with RL as well by modulating the learning rate (Iigaya et al., 2018). The hippocampus has place cells that are thought to encode the successor representation (Stachenfeld et al., 2017). Historically, research in RL has seen benefits in drawing insights from cognitive biology. The field has roots in early psychological experiments with reinforcements (Sutton and Barto, 1981), and RL researchers often find ways in nature to improve algorithms. For example, episodic memory allowed RL

researchers to develop artificial agents with super-human performance in complex board games, most impressively (Blundell et al., 2016).

Both the TAME framework and RL specify that the agent is reward-seeking. For TAME, the reward is achieving a goal state, and thus there is a direct map from Attribute 1 of the TAME framework to the fundamental assumption of reinforcement learning. RL offers TAME a mathematical formalization through MDPs and related generalizations: the reward function can be specified to be 0 until the goal state is reached, at which point you get a reward of 1. These sparse rewards are typically hard to deal with, but there are well-known ways of approaching this problem from reinforcement learning, with likely more techniques to discover as TAME is better understood. The reinforcement learning agent might have intrinsic curiosity that drives it to new parts of the state space that allow it to find that sparse reward that comes from reaching the goal state, e.g. as in Ref. Pathak et al. (2017); the reinforcement learning agent might be asked to solve incremental tasks that lead it to solving the full task (Florensa et al., 2018), essentially allowing for reaching of intermediate goal states first; or the reinforcement learning agent might be asked to solve a different task whose solution correlates with that of the original task (Riedmiller et al., 2018). Conversely, Attributes 2–4 of TAME can constitute new directions of research in reinforcement learning, as elaborated upon subsequently in this manuscript.

3. Novel questions: from biology to RL and back

RL has its roots in psychology — in particular, in understanding how reward and punishment could shape the weights of a network that decided behavior in real biological experiments as part of the Rescorla-Wagner model (Rescorla, 1988)– so perhaps it is no surprise that thinking about biology gives us even more new directions for research in RL. In this manuscript, we advocate for thinking about reinforcement learning in lower-level organisms, too– not just the higher-level organisms that have preferentially populated the reinforcement learning literature. This leads to frameworks that are not MDPs or even their more complex variants.

Typically, when reinforcement learning researchers think about biology, they think about higher-level organisms: mice, monkeys, humans. In these brains, they find model-free and model-based reinforcement learning systems (Daw et al., 2011a). These brains have so many neurons that they can essentially implement any calculation, in the same way that neural networks with a large enough number of neurons (and enough depth) can approximate any function (Hornik et al., 1989). Any of the reinforcement learning algorithms that have been developed can be implemented, although exactly how is a question for future research.

However, lower-level organisms (or indeed, parts of organisms, such as cells, tissues, and organs (Levin, 2019)) might also implement reinforcement learning algorithms. In Ref. Celani and Vergassola (2010), it appears that bacteria are solving a slightly different reinforcement learning-like objective function. In a space with many bacteria, any chemoattractant gradient that is created quickly vanishes, as the bacteria swim up the gradient and consume the chemoattractant. This amounts to each bacterium being in a worst-case scenario, in which potential rewards quickly vanish. This is a slightly different objective function than what is usually considered in reinforcement learning, but it still falls under the heading of reinforcement learning— essentially, the bacterium deals with a minimax objective, where the goal is to maximize reward in the worst-case environment. In Ref. Celani and Vergassola (2010), they show that the bacteria's response to chemoattractant pasts is designed to optimize this minimax objective. Additional learning mechanisms (mediated by bioelectricity and other mechanisms) have been reported in bacterial colonies (Ben-Jacob, 2009; Lee et al., 2018; Yang et al., 2020). The variety of learning mechanisms does not preclude RL from describing these colonies, as RL is more a statement of the goal rather than how the goal is achieved.

Given that, might there be other lower-level organisms or other biological subsystems that are reinforcement learners (Dexter et al., 2019; Gershman et al., 2021)? Might asking how they learn lead to new algorithms and novel questions?

For example, might bacteria actually also use reinforcement learning to solve the minimax objective? In humans, it is known that both model-free and model-based reinforcement learning systems are used, and the results combined. In bacteria, a model-based reinforcement learning system would require prediction. They could do this using a reservoir, from reservoir computing (Lukoševičius and Jaeger, 2009; Schrauwen et al., 2007)– using a simple genetic regulatory circuit (Katz and Springer, 2016; Katz et al., 2018). A new, and simple, computation shows that the reservoir of mRNAs and proteins corresponding to a simple genetic regulatory circuit can be used for prediction. The goal of what follows is to suggest that complex computations (reservoir computing) could exist in a simple unicellular organism, so that it is not too far-fetched to think that lower-level organisms are solving reinforcement learning problems.

As an example, E. coli is a model unicellular organism. E. coli's preferred food source is glucose, but if glucose is absent, E. coli is able to consume other forms of sugar instead. In order to do this, it needs to produce the protein lactase by transcribing its lac operon gene. Ideally, the lac operon gene is only transcribed when glucose is absent so that extra energy is not wasted on transcription. To optimize the production of lactase, E. coli should predict its environment rather than just respond to it.

Following a standard model of gene regulatory networks (Thattai and Van Oudenaarden, 2001), the state of the genetic system can be described at time t by the total number of mRNA molecules (j) that have been transcribed from the lac operon gene. The mRNA transcribe at a rate of α per mRNA and degrade at a rate of β per mRNA. The change in probability that there are j mRNA at time t (here, dp/dt) can be described with four components:

- The probability that one timestep dt ago, there were $j-1$ mRNA and one was transcribed, which is αdt;
- The probability that dt ago, there were $j+1$ mRNA and one degraded, which is $\beta(j+1)dt$
- The probability that in dt, one more mRNA will be transcribed, which is αdt;
- The probability that in dt, one mRNA will degrade, which is $\beta j dt$.

We set α to be the intensity of the middle pixel of a naturalistic video collected by a GoPro camera. As a result, the number of mRNAs can contain information about the future intensities of the naturalistic video's middle pixel. This rate of mRNA transcription is described by the van Oudenaarden equation:

$$\frac{d}{dt}p(j,t) = \alpha \cdot p(j-1,t) + \beta \cdot (j+1) \cdot p(j+1,t) - \alpha \cdot p(j,t) - \beta \cdot j \cdot p(j,t) \quad (1)$$

This model of mRNA transcription in E. coli can be used to predict complex stimuli, such as video input. Fig. 3 graphs the squared correlation coefficient of the predicted video frame and the video frame itself with a time lag given by that of the x-axis.

If a simple E. coli organism equipped with a simple genetic regulatory circuit can predict complex input like that of a GoPro video, then even the simplest of organisms might be a reinforcement learner. Indeed, it was recently shown that genetic regulatory networks exhibit learning and memory (Biswas et al., 2021; Watson et al., 2010), implying that any organism can be a reinforcement learner.

3.1. RL learners composed of many RL learners

Biological organisms are composed of parts, many of which used to be independent organisms and have retained many computational features needed for survival. Engineered agents are usually not multiscale in that sense (at least, not yet — swarm robotics is beginning

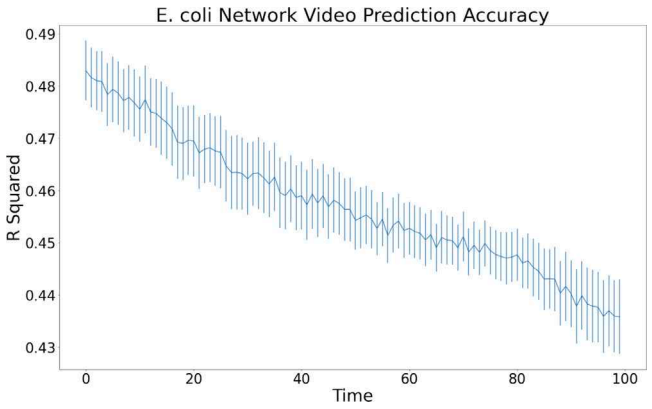

Fig. 3. The squared correlation coefficient between the number of mRNAs and the future intensity of the middle pixel of a naturalistic video captured by a GoPro camera, separated by a time lag given by the x-axis, when these intensities are used as the transcription rate of the gene. Standard errors are found by bootstrapping.

to exploit this multiscale competency architecture (Brambilla et al., 2013)). Biological organisms currently have some advantages over engineered agents as a result.

For example, individual bacteria are reinforcement learners, trying to maximize reward in the worst-case environment (Celani and Vergassola, 2010). When they coalesce to form a biofilm, they begin to act as one goal-directed agent. The unification of this goal is evidenced by the survival of the interior of the biofilm. If every bacterium acted selfishly for its own survival, the bacteria on the ends would have the best access to food while the bacteria in the center would suffer. But potassium waves lead to electrical communication that synchronizes eating patterns so that the bacteria in the center get enough food (Martinez-Corral et al., 2019; Prindle et al., 2015). Hence, the biofilm acts as one single organism, to the benefit of all bacteria.

Tumors are another example of a biological structure composed of individual organisms. Tumors can eventually develop a resistance to chemotherapy. It is thought that the tumor initially contains a few cancer cells that have an innate resistance to chemotherapy (i.e., a selectionist explanation of group resistance by differential survival and repopulation (Gatenby and Brown, 2018)); but it is possible that actually some cancer cells learn to become resistant to chemotherapy. The ability of the individual cancer cells to learn to survive then could enable the tumor to learn to survive. The "boundary of the self" model (Levin, 2019) proposes that cancer arises when cells disconnect from the electrical network that binds cells toward larger-scale morphological goals and revert to their unicellular behavior: on this view, they are not more selfish than normal cells, but instead their "selves" are smaller (the boundaries between self and external environment have shrunk back to the level of a single cell (Levin, 2021b)). In this context, the question of whether individual cells learn and/or the community learns focuses attention on the scale of the Agent and the credit assignment that needs to occur at different levels for RL to occur (Watson et al., 2022).

The portion of RL literature devoted to understanding multi-agent reinforcement learning is growing but relatively small (Hernandez-Leal et al., 2019). Similarly, swarm robotics (in which there is a swarm of robots that are often hand-designed for a task) is growing rapidly, but still, there are open questions (Brambilla et al., 2013). In multi-agent RL, there are four focuses of research (Hernandez-Leal et al., 2019): analysis of emergent behaviors, meaning that we unravel what actually happens when these agents are placed in the same environment; learning communication, meaning that we understand what happens when these learners can share information to cooperate; learning cooperation, meaning that we understand how to get these learners to cooperate without communication; and agents modeling agents, meaning that we understand how these learners end up modeling the other agents that are essentially part of their environment. These concepts play into an understanding of how a composition of organisms can be a more effective emergent organism, but further insight is likely to be gained from biology.

For example, it is an open question how to design the agents in the multi-agent RL scenario or the robots in the swarm so that they interact to achieve the desired goal, especially given that the environment from the individual agent's perspective is nonstationary (Brambilla et al., 2013). If one instead decides to set the agents to have some learning strategy, some behavior emerges which may be difficult to model. Biology (as you might have gathered from the examples above) has found a way to fine-tune the "robots" in its swarm or the RL agents in its multi-agent RL scenario so that the whole functions and survives. When we understand how, we are likely to see major advances in swarm robotics.

Also, one of the key questions that biophysicists like to ask in such studies is how efficacy scales with various resources– here, the number of cells in the multicellular organism and the rate of communication. So we can ask, how does the efficacy of the multicellular organism scale with the number of cells? In other words, are most robot swarms *too* redundant and wasteful of their resources in an attempt to achieve robustness? Without access to tight lower bounds on performance with number of components, it is hard to understand how well one's system is scaling. With respect to mechanisms, like bioelectric signaling (Levin, 2021a), that bind individual cells into networks with larger-scale goals in different problem spaces (thus enlarging their cognitive light cone), it is still poorly understood how their computational capacities scale with cell number and topological relationships.

Answering these questions would lead to advances in robotics to engineer agents that compete and cooperate to more robustly achieve some group objective, as well as better communication strategies with cellular composite agents that could improve approaches in regenerative medicine.

3.2. Environments with agency

Environments that contain multiple agents are complicated. The environment might be: beneficial, in which part of the environment (a mentor) trains the agent in question to succeed; benign, in which the environment does not train the agent at all and instead lets the agent learn how to succeed in the environment on their own; or adversarial, in which the environment actively tries to make it hard to succeed and survive (Celani and Vergassola, 2010). Environments with agency (Fields and Levin, 2023) include the first and last examples, whereas the second example shows little agency. The need to ascertain degrees of agency "laterally" (within one's level of organization) is compounded by the possibility of agency above (is the agent a cog in a much larger being with its own emergent agency?) and below (are the agent's own parts also agents, the behavior of which could be manipulated via behavior-shaping signals). In all of these cases, it is imperative for an agent to estimate how many agents exist in any learning interaction: just one (the agent is driving "learning" from a mechanical, uncaring world) or more (the agent is being "trained", by other agents with agendas).

Biological organisms exist in all three types of environments, whereas the reinforcement learners typically seen in the literature exist in the second type of environment. For example, E. coli when in natural ecological settings appear to exist in adversarial environments (Celani and Vergassola, 2010). The objective function that they appear to optimize is a minimax– a drive for maximum performance in the worst-case environment– and comes from the existence of many other E. coli in the same environment that chew away at chemoattractant gradients the moment they appear. Meanwhile, monkeys in a cued task followed by a reward (Schultz et al., 1997) show, in their dopaminergic activity, an ability to code a typical model-free reinforcement learning algorithm

called temporal difference learning. This would make sense for a benign environment, in which the expected sum of discounted rewards is to be maximized.

It is unclear how to choose the agent's objective function to interpolate between these extremes in objective function, but in multi-agent settings, this must be done in order to understand the individual agents and how a full organism can fruitfully emerge from a combination of individual organisms.

Note that environments are even more complicated than this, since what we define as the reinforcement learning agent may be unintuitive and may change over time. For instance, when talking about Planarian regeneration as a reinforcement learning agent, one can fruitfully say that the error correction module is part of the reinforcement learning environment, while the part of the Planaria that acts to change electric fields is the reinforcement learning agent.

The idea that an environment has agency (Fields and Levin, 2023) brings up the important notion, central to our main point, that we can ascribe agency to something that may not have what we would guess to be its goal as its actual goal. And yet, it is a useful philosophical construct to describe the environment as having agency with its imputed goal in mind, if we can describe its behavior accurately. This idea applies to the organisms in question as well.

We believe that there is no need to claim that any particular goal is the objectively correct goal for an agent (e.g., polycomputing (Bongard and Levin, 2023; Levin, 2022b)). Goals are interpretation frames assigned by observers, who make hypotheses about the behavior of agents and use goal-directed frameworks to achieve prediction and control. For complex agents, even the goals set by an engineer (or evolution) may or may not be what they actually end up pursuing. Importantly, complex agents have the ability to build internal models of other agents' goals, which they can apply to their own actions. In other words (paralleling the work on confabulation in cognitive neuroscience, and theories such as those in Ref. Chater (2018)), the agent itself is also an observer with its own self-model of its goals.

3.3. How to perform well in fluctuating environments quickly

Organisms typically do not live in stationary environments. Stationary environments are those that do not materially change with time. Indeed, cells are surrounded by other cells that naturally make the environment of each cell nonstationary. (The environment of a cell includes the environment of other learning cells.) More glaring examples of nonstationarity include transitions of the agent from caterpillar to butterfly. As such, it is imperative to ask how agents perform in fluctuating environments, not just in static ones.

As noted in Ref. Neftci and Averbeck (2019), this implies that the heuristics used in biology might be useful while high-powered reinforcement learning algorithms fail. Indeed, biological organisms can still outperform engineered agents in adapting quickly to a new environment. While AlphaGo was able to achieve super-human performance, it was not data-efficient, requiring thousands of hours of game-play to learn strategies that took adept humans a few games to learn (Tsividis et al., 2017). Rats quickly learn to do a credit assignment in experiments in which they are rewarded for maintaining a temperature difference between their ears (how do the rats know they were not rewarded for digesting well, or for taking only so many sips of water? etc.) The same remarkable example of credit assignment is also seen in human experiments with biofeedback (Vital et al., 2021; Zimet, 1979) - as long as consistent reward is provided, the system is able to bring even autonomic functions under control (using effectors of which the system was previously unaware and crossing levels to control physiological processes normally not accessible to behavioral control). Engineered agents, if not told that temperature differentials were an important variable to consider, might take much longer than the rats.

If one deals with a fluctuating environment, one must quickly learn a new action policy. As we change jobs and move cities, we must find a new favorite coffee shop, a new route to work, new friends at work, and so on. How? There are many routes to doing this, but some of them seem to fail. For example, in Ref. Marzen (2019), the policy gradient method often fails to adapt to new environments. Trying to find methods that quickly adapt to fluctuating environments while retaining lessons of previous environments (metalearning Wang et al., 2018) may lead to advances in reinforcement learning, and if we study organisms that do manage to adapt quickly, this may give us inspiration for artificial agents– episodic memory (Botvinick et al., 2019) being only one of many possible methods.

3.4. Robustness: how to survive if someone kills half your intelligence

There are a number of beautiful examples of organisms losing major parts of their brain and/or body and still surviving, regrowing, and regaining their target morphologies. For example, planaria regenerate if you cut them into pieces. Human embryos generate fully-functional twins, not half-bodies, if the early embryo is split in half. During the caterpillar to butterfly transition, the brain of the caterpillar gets taken apart and put back together, but the butterfly remembers whatever the caterpillar was trained to learn. Salamanders regenerate their eyes, jaws, spinal cords, ovaries, and limbs. Ground squirrels need a lot of memory for understanding a complex social structure, but when it gets cold in the winter, the brain loses a third to a half of its mass because the squirrels cannot get food; yet, brains inflate again in the spring with no loss to memory of social information (see Blackiston et al. (2015) and Levin (2022a) for reviews of these kinds of robustness examples in morphological and behavioral spaces).

We are not always robust to every aspect of the environment that might change. For instance, it is possible to achieve an above-average IQ with much less brain volume than normal (Lorber, 1981), but humans have not evolved to do so despite the obvious survival advantages of a smaller head size for human fetuses. Bioelectric signals can exert significant control over form and function (reviewed in Ref. Levin (2021a)), but organisms are robust to a remarkably wide variety of conditions, such as different levels of ions in their aqueous medium.

The equivalent of this in reinforcement learning would be some form of Dropout: at the time of testing, connections are dropped randomly, nodes are killed randomly, and yet the neural network performs as you have trained it to. The key here is that Dropout is a regularization technique employed during training, while we ask: what would happen if a stringent version of Dropout was to be used during testing, and maybe even testing alone? Could we drop perhaps half of the nodes in an artificial neural network that estimates the value function and get similar performance? Could we develop a way of training neural networks that achieves this level of robustness? That is what certain biological organisms can do.

4. Novel research questions and programs in biology from RL/TAME

As RL was birthed from psychology, RL has already been used famously to interpret biological data, e.g. Ref. Schultz et al. (1997). There may even be a way to interpret what biological organisms are doing in terms of the most recent RL algorithms (Botvinick et al., 2019). But it pays to look toward research questions and programs in biology that are made possible by theoretical advances in TAME/RL.

The first and most fundamental question we have to answer before we begin is: are biological organisms goal-directed agents? Do the frameworks of TAME and RL even apply? It is always possible to infer from behavior a reward function (Zhifei and Joo, 2012)– but how do we know that it is right to infer a reward function in the first place? The way to identify learning agents is empirical testing: can training strategies (Abramson and Levin, 2021), applied to unconventional and diverse embodied agents, afford efficient prediction and control? This can uncover basal agency in surprising guises, such as in systems as

simple as gene-regulatory networks (Biswas et al., 2021; Watson et al., 2010) and in whole populations/ecosystems (Power et al., 2015).

To show that the organism is a reinforcement learner, one must show that the organism's action policy changes in a way that benefits them as it moves to a new environment. Once this has been established, the theoretical frameworks of TAME and RL can be brought to bear on the behavior and neuronal or pre-neuronal correlates in biological organisms.

4.1. Estimating cognitive capacity based on the action policy of the organism

Cognitive capacity is not a binary thing– does an organism have it or not– but instead should be viewed as a continuum of degree and kind (Levin, 2022b). Bacteria are likely to have little cognitive capacity, while humans have a lot, but bacteria still have a little (Celani and Vergassola, 2010; Ben-Jacob, 2009; Jacob et al., 2006) and can be given more, including RL capability, by synthetic biology efforts (Racovita et al., 2022).

Can we ascertain what this cognitive capacity is based on the action policy of the organism?

The action policy of the organism, under the right conditions, implies certain cognitive traits, such as look-ahead and planning. To see these traits and to differentiate one learning strategy for an action policy from another, one likely has to place the organisms in a new environment and watch the organism learn.

There is currently no well-defined scale for cognitive capacity, but looking at the changing action policies of organisms in fluctuating environments may allow us to create such a scale. The question we have to ask ourselves is: how quickly do organisms learn, and what cognitive capacity is implied by such learning? This is likely to be correlated with whether or not the organism in question is composed of organisms itself.

It should be said that it may be dangerous to have such a scale, as certain ethical restrictions are partly based on an intuitive understanding of such a scale. For instance, IRB regulations and how much harm you can do to organisms of a particular type in service of scientific progress depends partly on how much pain we think they can feel (National Research Council et al., 2010). If we decide that an animal has such a low cognitive capacity that we should not even worry about its pain threshold in service of science, we could inadvertently violate ethical principles.

4.2. Finding neuronal and pre-neuronal correlates of RL/TAME signals

Quantitative biology has ushered in a new era of measurement, in which deep learning combined with microscopy can track in great detail the exact pose of a fly over time, e.g. as in Ref. Günel et al. (2019), and image aspects of the brain with varying degrees of spatiotemporal resolution. Imagine that we can take detailed measurements of the behavior of an organism, whether it be a Planaria or tadpole or salamander or mouse, in novel environments. We might be able to better determine what kind of reinforcement learner we are examining– a policy gradient learner, or a Q-learner, or a SARSA learner, and so on Daw et al. (2011b), or with modifications as in Ref. Ashwood et al. (2020).

Perhaps more importantly, we might then be able to use our new quantitative measurements of brain and other activity to ascertain which neuromodulators and hormones and other molecules are carrying information about the corresponding reinforcement learning signals. Although brains are required to carry such signals, if lower-level organisms are indeed reinforcement learners as well, we should expect to see neuronal and pre-neuronal architectures carrying these signals as well. This could revolutionize understanding of the behavior and workings of various lower-level organisms. The ability of planaria for example, converted to a 2-headed form by transient physiological experience, to continue to form 2-headed animals on future rounds of regeneration, may be modeled by RL algorithms that use the bioelectric circuit. Similarly, salamander limbs eventually give up regeneration following continued amputation (Bryant et al., 2017), one reason for which could be learning on the part of the cellular collective of the body. Could interventions be developed to retrain planarian cells toward a different target morphology — to exploit the built-in RL capacities of their hardware but provide a different target? It is already known that planarian cells can build heads appropriate to other species (Emmons-Bell et al., 2015), but we are only beginning to develop approaches that manipulate morphogenesis top-down (using the tools of behavioral sciences) to complement mainstream bottom-up (molecular medicine) approaches.

4.3. Regenerative medicine

Birth defects, traumatic injury, cancer, aging, degenerative disease — all of these pressing biomedical needs would be resolved if we perfected ways to control the anatomical structures toward which cellular collectives build and repair. The biomedicine of the future will look less like pathway rewiring and more like communication, biofeedback, and training — a kind of somatic psychiatry (Pio-Lopez et al., 2022) that parallels the move in computer science from programming via rewiring the hardware to taking advantage of software learning capacities and high-level control languages. Bottom-up approaches such as CRISPR and genome editing will hit inevitable limitations (beyond low-hanging fruit of single-gene diseases) because it is in general impossible to compute what genes must be altered to achieve a desired complex system-level effect (Lobo et al., 2014). Instead, it is imperative to learn to take advantage of the causal structure of multiscale learning agents to train them toward desired form and function (Mathews and Levin, 2018).

4.4. Synthetic biosciences

Moving forward from synthetic biology (reprogramming cells via novel molecular circuits), we must enter the domain of synthetic morphology: reprogramming large-scale form and function (Davies and Levin, 2023; Glykofrydis et al., 2021). Since bioengineers, like evolution, work with an agential material (cells Davies and Levin, 2023), not a passive one, the roadmap for this field is to begin to understand the kinds of learning that cellular collectives are capable of, and develop strategies to re-specify their goals in anatomical and behavioral spaces. Along with the next-level task of actually increasing the RL capacity of living tissues, these strategies will greatly potentiate the construction of arbitrary desired synthetic living machines: going beyond restoring standard morphologies (regenerative medicine) to complete control over growth and form to create whatever novel structures are needed for engineering uses.

5. Conclusion

TAME is a framework for understanding, and learning to manipulate, the robust functional capabilities of multiscale agents (whether evolved, designed, or hybrid). RL promises to be its mathematical instantiation. However, RL as it stands today, misses a few of TAME's attributes– in particular, its emphasis on the multi-agent setups that are common in biology. As such, these attributes promise to deliver new questions in reinforcement learning.

RL may not seem to be appropriate for addressing organism behavior because goal-directedness appears to violate causality, but this has been refuted, for example in Ref. Heylighen (2023). Likewise, determinism in physical rules does not imply a lack of agency because the agent can still be the source of its actions (see Ref. Babcock and McShea (2023)). Moreover, RL may appear to be a less useful instantiation of TAME for studying organisms because so many learning modalities are involved (Corning et al., 2023) and so few actual

rewards and punishments are meted out in real organisms. However, RL is more about the goals and less about how they are achieved, and so the plethora of learning modalities that exists in nature is completely aligned with the possibility that RL algorithms could be used to mathematically describe organism behavior. And, even if we only receive rewards every so often, such as when we have children or when we eat, "sparse rewards" are allowable within the RL framework. One thing that we have not commented upon and that may be an interesting future direction is to imagine that there are multiple reward functions, e.g. one for sustenance, one for shelter, one for mating, and so on, although perhaps organism behavior ultimately boils down to maximization of the number of viable children, however one defines "viable".

RL also took much inspiration from biology and psychology, and so many new research directions and advances in RL come from incorporating biological attributes. An example of this is Ref. Wang et al. (2018), in which meta-learning is thought to be performed by a recurrent neural network in the prefrontal cortex. This paper is an example of exactly the kind of interdisciplinary research we hope to foster with this piece. Indeed, one could view evolution as a meta-learning mechanism for the lower level organisms that we have been concerned with.

We suggest that several questions driven by the TAME framework could advance the field of RL. For instance, many questions arose from considering artificial agents that were composed of reinforcement learners, and while some work has been done in this direction (on multi-agent reinforcement learning), we would argue– not enough. Also, little to no work has been done on quantifying the cognitive capabilities of an RL agent with a scale, which would help quantify the ways in which biological organisms exhibit degrees of intelligence at multiple scales and in multiple problem spaces (Lyon, 2006). On the other hand, the RL and TAME frameworks provide new insights into biological organisms and on how to modify function at many levels in regenerative medicine contexts by resetting the homeodynamic goals toward which cells, tissues, and organs regulate (Lagasse and Levin, 2023). New quantitative measurements and interventional experiments enable one to test these insights.

Thus, we believe that the interplay between RL algorithm development in engineered systems and the study of biological agents via the TAME framework will lead to advances in both biology and engineering. Moreover, this is likely just one example of concepts and tools that can be ported across fields in the study of natural and artificial intelligence. RL is not the only mathematical instantiation of TAME principles, and coexists naturally with other possible formalisms such as the prediction-centric expected free energy principle of Friston and colleagues (Smith et al., 2022). We believe it provides interesting and practical insights into one of the most fascinating problems facing science and philosophy — the deep principles of diverse intelligence.

CRediT authorship contribution statement

Gabriella Seifert: Formal analysis, Software, Writing – review & editing, Writing – original draft. **Ava Sealander:** Data curation, Software, Writing – review & editing. **Sarah Marzen:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Michael Levin:** Conceptualization, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgments

S.M. was supported by the Air Force Office for Scientific Research, Award FA9550-19-1-0411. M.L. gratefully acknowledges the support of Grant 62212 from the John Templeton Foundation and support by the Air Force Office of Scientific Research (AFOSR) under award number FA9550-22-1-0465. The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

References

Abramson, Charles I., Levin, Michael, 2021. Behaviorist approaches to investigating memory and learning: A primer for synthetic biology and bioengineering. Commun. Integr. Biol. 14 (1), 230–247.

Ashwood, Zoe, Roy, Nicholas A, Bak, Ji Hyun, Pillow, Jonathan W, 2020. Inferring learning rules from animal decision-making. Adv. Neural Inf. Process. Syst. 33, 3442–3453.

Babcock, Gunnar, McShea, Daniel W., 2023. Resolving teleology's false dilemma. Biol. J. Linnean Soc. 139 (4), 415–432.

Baluška, František, Levin, Michael, 2016. On having no head: cognition throughout biological systems. Front. Psychol. 7, 902.

Ben-Jacob, Eshel, 2009. Learning from bacteria about natural information processing. Ann. New York Acad. Sci. 1178 (1), 78–90.

Bertsekas, Dimitri, 2012. Dynamic Programming and Optimal Control: Volume I. vol. 4, Athena scientific.

Birnbaum, Kenneth D., Alvarado, Alejandro Sánchez, 2008. Slicing across kingdoms: regeneration in plants and animals. Cell 132 (4), 697–710.

Biswas, Surama, Manicka, Santosh, Hoel, Erik, Levin, Michael, 2021. Gene regulatory networks exhibit several kinds of memory: Quantification of memory in biological and random transcriptional networks. Iscience 24 (3), 102131.

Blackiston, Douglas J., Shomrat, Tal, Levin, Michael, 2015. The stability of memories during brain remodeling: a perspective. Commun. Integr. Biol. 8 (5), e1073424.

Blundell, Charles, Uria, Benigno, Pritzel, Alexander, Li, Yazhe, Ruderman, Avraham, Leibo, Joel Z, Rae, Jack, Wierstra, Daan, Hassabis, Demis, 2016. Model-free episodic control. arXiv preprint arXiv:1606.04460.

Bongard, Joshua, Levin, Michael, 2023. There's plenty of room right here: Biological systems as evolved, overloaded, multi-scale machines. Biomimetics 8 (1), 110.

Botvinick, Matthew, Ritter, Sam, Wang, Jane X, Kurth-Nelson, Zeb, Blundell, Charles, Hassabis, Demis, 2019. Reinforcement learning, fast and slow. Trends Cogn. Sci. 23 (5), 408–422.

Boussard, Aurèle, Delescluse, Julie, Pérez-Escudero, Alfonso, Dussutour, Audrey, 2019. Memory inception and preservation in slime moulds: the quest for a common mechanism. Philos. Trans. R. Soc. B 374 (1774), 20180368.

Brambilla, Manuele, Ferrante, Eliseo, Birattari, Mauro, Dorigo, Marco, 2013. Swarm robotics: a review from the swarm engineering perspective. Swarm Intell. 7, 1–41.

Bryant, Donald M, Sousounis, Konstantinos, Farkas, Johanna E, Bryant, Sevara, Thao, Neng, Guzikowski, Anna R, Monaghan, James R, Levin, Michael, Whited, Jessica L, 2017. Repeated removal of developing limb buds permanently reduces appendage size in the highly-regenerative axolotl. Dev. Biol. 424 (1), 1–9.

Celani, Antonio, Vergassola, Massimo, 2010. Bacterial strategies for chemotaxis response. Proc. Natl. Acad. Sci. 107 (4), 1391–1396.

Chater, Nick, 2018. Mind Is Flat: The Remarkable Shallowness of the Improvising Brain. Yale University Press.

Clawson, Wesley P., Levin, Michael, 2022. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. Biol. J. Linnean Soc. 139 (4), 457–486.

Corning, Peter A, Kauffman, Stuart A, Noble, Denis, Shapiro, James A, Vane-Wright, Richard I, 2023. Evolution" on Purpose": Teleonomy in Living Systems. MIT Press.

Couzin, Iain, 2007. Collective minds. Nature 445 (7129), 715.

Davies, Jamie, Levin, Michael, 2023. Synthetic morphology with agential materials. Nat. Rev. Bioeng. 1 (1), 46–59.

Daw, Nathaniel D, Gershman, Samuel J, Seymour, Ben, Dayan, Peter, Dolan, Raymond J, 2011a. Model-based influences on humans' choices and striatal prediction errors. Neuron 69 (6), 1204–1215.

Daw, Nathaniel D., et al., 2011b. Trial-by-trial data analysis using computational models. Decis. Mak. Affect Learn.: Atten. Perform. XXIII 23 (1).

Dexter, Joseph P., Prabakaran, Sudhakaran, Gunawardena, Jeremy, 2019. A complex hierarchy of avoidance behaviors in a single-cell eukaryote. Curr. Biol. 29 (24), 4323–4329.

Durant, Fallon, Lobo, Daniel, Hammelman, Jennifer, Levin, Michael, 2016. Physiological controls of large-scale patterning in planarian regeneration: a molecular and computational perspective on growth and form. Regeneration 3 (2), 78–102.

Durant, Fallon, Morokuma, Junji, Fields, Christopher, Williams, Katherine, Adams, Dany Spencer, Levin, Michael, 2017. Long-term, stochastic editing of regenerative anatomy via targeting endogenous bioelectric gradients. Biophys. J. 112 (10), 2231–2243.

Emmons-Bell, Maya, Durant, Fallon, Hammelman, Jennifer, Bessonov, Nicholas, Volpert, Vitaly, Morokuma, Junji, Pinet, Kaylinnette, Adams, Dany S, Pietak, Alexis, Lobo, Daniel, et al., 2015. Gap junctional blockade stochastically induces different species-specific head anatomies in genetically wild-type girardia dorotocephala flatworms. Int. J. Mol. Sci. 16 (11), 27865–27896.

Emmons-Bell, Maya, Durant, Fallon, Tung, Angela, Pietak, Alexis, Miller, Kelsie, Kane, Anna, Martyniuk, Christopher J, Davidian, Devon, Morokuma, Junji, Levin, Michael, 2019. Regenerative adaptation to electrochemical perturbation in planaria: A molecular analysis of physiological plasticity. Iscience 22, 147–165.

Fankhauser, Gerhard, 1945. Maintenance of normal structure in heteroploid salamander larvae, through compensation of changes in cell size by adjustment of cell number and cell shape. J. Exp. Zool. 100 (3), 445–455.

Fields, Chris, Levin, Michael, 2022. Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments. Entropy 24 (6), 819.

Fields, Chris, Levin, Michael, 2023. Regulative development as a model for origin of life and artificial life studies. Biosystems 229, 104927.

Florensa, Carlos, Held, David, Geng, Xinyang, Abbeel, Pieter, 2018. Automatic goal generation for reinforcement learning agents. In: International Conference on Machine Learning. PMLR, pp. 1515–1528.

Furusawa, Chikara, Kaneko, Kunihiko, 2002. Origin of multicellular organisms as an inevitable consequence of dynamical systems. Anat. Record: Off. Publ. Am. Assoc. Anat. 268 (3), 327–342.

Gatenby, Robert, Brown, Joel, 2018. The evolution and ecology of resistance in cancer therapy. Cold Spring Harb. Perspect. Med. 8 (3), a033415.

Gershman, Samuel J., 2018. The successor representation: its computational logic and neural substrates. J. Neurosci. 38 (33), 7193–7200.

Gershman, Samuel J, Balbi, Petra EM, Gallistel, C Randy, Gunawardena, Jeremy, 2021. Reconsidering the evidence for learning in single cells. Elife 10, e61907.

Glykofrydis, Fokion, Cachat, Elise, Berzanskyte, Ieva, Dzierzak, Elaine, Davies, Jamie A, 2021. Bioengineering self-organizing signaling centers to control embryoid body pattern elaboration. ACS Synth. Biol. 10 (6), 1465–1480.

Groetsch, Charles W., Groetsch, C.W., 1993. Inverse Problems in the Mathematical Sciences. vol. 52, Springer.

Günel, Semih, Rhodin, Helge, Morales, Daniel, Campagnolo, João, Ramdya, Pavan, Fua, Pascal, 2019. DeepFly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult drosophila. Elife 8, e48571.

Halley, Julianne D, Smith-Miles, Kate, Winkler, Dave A, Kalkan, Tuzer, Huang, Sui, Smith, Austin, 2012. Self-organizing circuitry and emergent computation in mouse embryonic stem cells. Stem Cell Res. 8 (2), 324–333.

Harris, Albert K., 2018. The need for a concept of shape homeostasis. Biosystems 173, 65–72.

Hernandez-Leal, Pablo, Kartal, Bilal, Taylor, Matthew E., 2019. A survey and critique of multiagent deep reinforcement learning. Auton. Agents Multi-Agent Syst. 33 (6), 750–797.

Heylighen, Francis, 2023. The meaning and origin of goal-directedness: a dynamical systems perspective. Biol. J. Linnean Soc. 139 (4), 370–387.

Hornik, Kurt, Stinchcombe, Maxwell, White, Halbert, 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2 (5), 359–366.

Iigaya, Kiyohito, Fonseca, Madalena S, Murakami, Masayoshi, Mainen, Zachary F, Dayan, Peter, 2018. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. Nat. Commun. 9 (1), 2477.

Jacob, Eshel Ben, Shapira, Yoash, Tauber, Alfred I., 2006. Seeking the foundations of cognition in bacteria: From Schrödinger's negative entropy to latent information. Physica A 359, 495–524.

Jeong, Huijeong, Taylor, Annie, Floeder, Joseph R, Lohmann, Martin, Mihalas, Stefan, Wu, Brenda, Zhou, Mingkang, Burke, Dennis A, Namboodiri, Vijay Mohan K, 2022. Mesolimbic dopamine release conveys causal associations. Science 378 (6626), eabq6740.

Katsikopoulos, Konstantinos V., Engelbrecht, Sascha E., 2003. Markov decision processes with delays and asynchronous cost collection. IEEE Trans. Autom. Control 48 (4), 568–574.

Katz, Yarden, Springer, Michael, 2016. Probabilistic adaptation in changing microbial environments. PeerJ 4, e2716.

Katz, Yarden, Springer, Michael, Fontana, Walter, 2018. Embodying probabilistic inference in biochemical circuits. arXiv preprint arXiv:1806.10161.

Kirk, Donald E., 2004. Optimal Control Theory: An Introduction. Courier Corporation.

Kriegman, Sam, Blackiston, Douglas, Levin, Michael, Bongard, Josh, 2020. A scalable pipeline for designing reconfigurable organisms. Proc. Natl. Acad. Sci. 117 (4), 1853–1859.

Kriegman, Sam, Blackiston, Douglas, Levin, Michael, Bongard, Josh, 2021. Kinematic self-replication in reconfigurable organisms. Proc. Natl. Acad. Sci. 118 (49), e2112672118.

Lagasse, Eric, Levin, Michael, 2023. Future medicine: from molecular pathways to the collective intelligence of the body. Trends Mol. Med. 29 (9), 687–710.

Langton, Christopher G., 2019. Artificial life. In: Artificial Life. Routledge, pp. 1–47.

Lee, Calvin K, De Anda, Jaime, Baker, Amy E, Bennett, Rachel R, Luo, Yun, Lee, Ernest Y, Keefe, Joshua A, Helali, Joshua S, Ma, Jie, Zhao, Kun, et al., 2018. Multigenerational memory and adaptive adhesion in early bacterial biofilm communities. Proc. Natl. Acad. Sci. 115 (17), 4471–4476.

Levin, Michael, 2019. The computational boundary of a "self": developmental bioelectricity drives multicellularity and scale-free cognition. Front. Psychol. 10, 2688.

Levin, Michael, 2021a. Bioelectric signaling: Reprogrammable circuits underlying embryogenesis, regeneration, and cancer. Cell 184 (8), 1971–1989.

Levin, Michael, 2021b. Bioelectrical approaches to cancer as a problem of the scaling of the cellular self. Progress Biophys. Mol. Biol. 165, 102–113.

Levin, Michael, 2022a. Collective intelligence of morphogenesis as a teleonomic process.

Levin, Michael, 2022b. Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. Front. Syst. Neurosci. 16, 768201.

Levin, Michael, 2023a. Bioelectric networks: the cognitive glue enabling evolutionary scaling from physiology to mind. Animal Cogn. 1–27.

Levin, Michael, 2023b. Darwin's agential materials: evolutionary implications of multiscale competency in developmental biology. Cell. Mol. Life Sci. 80 (6), 142.

Levin, Michael, Pietak, Alexis M., Bischof, Johanna, 2019. Planarian regeneration as a model of anatomical homeostasis: recent progress in biophysical and computational approaches. In: Seminars in Cell & Developmental Biology. vol. 87, Elsevier, pp. 125–144.

Lobo, Daniel, Solano, Mauricio, Bubenik, George A, Levin, Michael, 2014. A linear-encoding model explains the variability of the target morphology in regeneration. J. R. Soc. Interface 11 (92), 20130918.

Lorber, John, 1981. Is your brain really necessary? Nurs. Mirror 152 (18), 29–30.

Lukoševičius, Mantas, Jaeger, Herbert, 2009. Reservoir computing approaches to recurrent neural network training. Comp. Sci. Rev. 3 (3), 127–149.

Lyon, Pamela, 2006. The biogenic approach to cognition. Cogn. Process. 7, 11–29.

Lyon, Pamela, 2015. The cognitive cell: bacterial behavior reconsidered. Front. Microbiol. 6, 264.

Martinez-Corral, Rosa, Liu, Jintao, Prindle, Arthur, Süel, Gürol M, Garcia-Ojalvo, Jordi, 2019. Metabolic basis of brain-like electrical signalling in bacterial communities. Philos. Trans. R. Soc. B 374 (1774), 20180382.

Marzen, Sarah E., 2019. Novelty detection improves performance of reinforcement learners in fluctuating, partially observable environments. J. Theoret. Biol. 477, 44–50.

Mathews, Juanita, Levin, Michael, 2018. The body electric 2.0: recent advances in developmental bioelectricity for regenerative and synthetic bioengineering. Curr. Opin. Biotechnol. 52, 134–144.

McCusker, Catherine, Gardiner, David M., 2011. The axolotl model for regeneration and aging research: a mini-review. Gerontology 57 (6), 565–571.

National Research Council, et al., 2010. Recognition and alleviation of pain in laboratory animals.

Neftci, Emre O., Averbeck, Bruno B., 2019. Reinforcement learning in artificial and biological systems. Nat. Mach. Intell. 1 (3), 133–143.

Oviedo, Néstor J, Morokuma, Junji, Walentek, Peter, Kema, Ido P, Gu, Man Bock, Ahn, Joo-Myung, Hwang, Jung Shan, Gojobori, Takashi, Levin, Michael, 2010. Long-range neural and gap junction protein-mediated cues control polarity during planarian regeneration. Dev. Biol. 339 (1), 188–199.

Pathak, Deepak, Agrawal, Pulkit, Efros, Alexei A, Darrell, Trevor, 2017. Curiosity-driven exploration by self-supervised prediction. In: International Conference on Machine Learning. PMLR, pp. 2778–2787.

Pezzulo, Giovanni, Levin, Michael, 2015. Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. Integr. Biol. 7 (12), 1487–1517.

Pezzulo, Giovanni, Levin, Michael, 2016. Top-down models in biology: explanation and control of complex living systems above the molecular level. J. R. Soc. Interface 13 (124), 20160555.

Pinet, Kaylinnette, McLaughlin, Kelly A., 2019. Mechanisms of physiological tissue remodeling in animals: Manipulating tissue, organ, and organism morphology. Dev. Biol. 451 (2), 134–145.

Pio-Lopez, Léo, Kuchling, Franz, Tung, Angela, Pezzulo, Giovanni, Levin, Michael, 2022. Active inference, morphogenesis, and computational psychiatry. Front. Comput. Neurosci. 16, 988977.

Power, Daniel A, Watson, Richard A, Szathmáry, Eörs, Mills, Rob, Powers, Simon T, Doncaster, C Patrick, Czapp, BłaŻej, 2015. What can ecosystems learn? Expanding evolutionary ecology with learning theory. Biol. Direct 10, 1–24.

Prindle, Arthur, Liu, Jintao, Asally, Munehiro, Ly, San, Garcia-Ojalvo, Jordi, Süel, Gürol M, 2015. Ion channels enable electrical communication in bacterial communities. Nature 527 (7576), 59–63.

Racovita, Adrian, Prakash, Satya, Varela, Clenira, Walsh, Mark, Galizi, Roberto, Isalan, Mark, Jaramillo, Alfonso, 2022. Engineered gene circuits capable of reinforcement learning allow bacteria to master gameplaying. bioRxiv, pp. 4–5.

Rescorla, Robert A., 1988. Behavioral studies of pavlovian conditioning. Annu. Rev. Neurosci. 11 (1), 329–352.

Riedmiller, Martin, Hafner, Roland, Lampe, Thomas, Neunert, Michael, Degrave, Jonas, Wiele, Tom, Mnih, Vlad, Heess, Nicolas, Springenberg, Jost Tobias, 2018. Learning by playing solving sparse reward tasks from scratch. In: International Conference on Machine Learning. PMLR, pp. 4344–4353.

Rosenblueth, Arturo, Wiener, Norbert, Bigelow, Julian, 1943. Behavior, purpose and teleology. Philos. Sci. 10 (1), 18–24.

Saló, Emili, Abril, Josep F., Adell, Teresa, Cebriá, Francesc, Eckelt, Kay, Fernández-Taboada, Enrique, Handberg-Thorsager, Mette, Iglesias, Marta, Molina, M Dolores, Rodríguez-Esteban, Gustavo, 2009. Planarian regeneration: achievements and future directions after 20 years of research. Int. J. Dev. Biol. 53 (8-9-10), 1317–1327.

Sawaya, Yorgo, Issa, George, Marzen, Sarah E., 2023. Framework for solving time-delayed Markov decision processes. Phys. Rev. Res. 5 (3), 033034.

Schrauwen, Benjamin, Verstraeten, David, Van Campenhout, Jan, 2007. An overview of reservoir computing: theory, applications and implementations. In: Proceedings of the 15th European Symposium on Artificial Neural Networks. P. 471-482 2007. pp. 471–482.

Schultz, Wolfram, Dayan, Peter, Montague, P. Read, 1997. A neural substrate of prediction and reward. Science 275 (5306), 1593–1599.

Smith, Ryan, Friston, Karl J., Whyte, Christopher J., 2022. A step-by-step tutorial on active inference and its application to empirical data. J. Math. Psychol. 107, 102632.

Solé, Ricard, Amor, Daniel R., Duran-Nebreda, Salva, Conde-Pueyo, Núria, Carbonell-Ballestero, Max, Montañez, Raul, 2016. Synthetic collective intelligence. Biosystems 148, 47–61.

Stachenfeld, Kimberly L, Botvinick, Matthew M, Gershman, Samuel J, 2017. The hippocampus as a predictive map. Nature Neurosci. 20 (11), 1643–1653.

Sutton, Richard S., Barto, Andrew G., 1981. Toward a modern theory of adaptive networks: expectation and prediction.. Psychol. Rev. 88 (2), 135.

Sutton, Richard S., Barto, Andrew G., 2018. Reinforcement Learning: An Introduction. MIT Press.

Thattai, Mukund, Van Oudenaarden, Alexander, 2001. Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. 98 (15), 8614–8619.

Tsividis, Pedro A, Pouncy, Thomas, Xu, Jaqueline L, Tenenbaum, Joshua B, Gershman, Samuel J, 2017. Human learning in atari. In: 2017 AAAI Spring Symposium Series.

Vandenberg, Laura N., Adams, Dany S., Levin, Michael, 2012. Normalized shape and location of perturbed craniofacial structures in the xenopus tadpole reveal an innate ability to achieve correct morphology. Dev. Dyn. 241 (5), 863–878.

Vital, Jose Edimosio Costa, de Morais Nunes, Adriele, New, Beatriz Souza de Albuquerque Cacique, de Sousa, Barbara Dayane Araujo, Nascimento, Micaele Farias, Formiga, Magno F, Fernandes, Ana Tereza NSF, et al., 2021. Biofeedback therapeutic effects on blood pressure levels in hypertensive individuals: A systematic review and meta-analysis. Complement. Ther. Clin. Pract. 44, 101420.

Wang, Jane X, Kurth-Nelson, Zeb, Kumaran, Dharshan, Tirumala, Dhruva, Soyer, Hubert, Leibo, Joel Z, Hassabis, Demis, Botvinick, Matthew, 2018. Prefrontal cortex as a meta-reinforcement learning system. Nature Neurosci. 21 (6), 860–868.

Watson, Richard, Buckley, Christopher L, Mills, Rob, Davies, Adam, 2010. Associative memory in gene regulation networks.

Watson, Richard A., Levin, Michael, Buckley, Christopher L., 2022. Design for an individual: connectionist approaches to the evolutionary transitions in individuality. Front. Ecol. Evol. 10, 64.

Wolf, Denise M, Fontaine-Bodin, Lisa, Bischofs, Ilka, Price, Gavin, Keasling, Jay, Arkin, Adam P, 2008. Memory in microbes: quantifying history-dependent behavior in a bacterium. PLoS One 3 (2), e1700.

Yang, Chih-Yu, Bialecka-Fornal, Maja, Weatherwax, Colleen, Larkin, Joseph W, Prindle, Arthur, Liu, Jintao, Garcia-Ojalvo, Jordi, Süel, Gürol M, 2020. Encoding membrane-potential-based memory within a microbial community. Cell Syst. 10 (5), 417–423.

Zhifei, Shao, Joo, Er Meng, 2012. A review of inverse reinforcement learning theory and recent advances. In: 2012 IEEE Congress on Evolutionary Computation. IEEE, pp. 1–8.

Zimet, Gregory D., 1979. Locus of control and biofeedback: a review of the literature. Percept. Mot. Skills 49 (3), 871–877.