What Predictive Brains Inherit: An Evolutionary Account of Cognition

Leonel Gómez-Sena

Sección Neurociencias, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

Correspondence: leonel.gomez@gmail.com

Preprint. This manuscript has not been peer reviewed. A version of this work has been submitted for publication.

Declaration of competing interests. The author declares no competing financial or non-financial interests.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data and code availability. The two-timescale model described in the Supplementary Material was implemented in Python. The complete implementation is archived at https://doi.org/10.5281/zenodo.22018805 and developed at https://github.com/leonelfgs/predictive-brains-inherit. It is a single file with no dependencies beyond NumPy and Matplotlib, is fully deterministic, and reproduces every quantitative result reported here exactly.

Abstract

Predictive and Bayesian frameworks explain how organisms weigh expectation against evidence, but not where the ingredients come from — what counts as a signal, what counts as a good outcome, which hypotheses and actions are available at all. We argue these are historical products, not formal assumptions, and hypothesise how they were built.

Cognition is chained together by *conditional necessity*: outcomes are contingent, but each expansion of an organism's possibility space generates demands any viable continuation must meet by some equivalent architecture. Replication requires access to matter and energy, secured by exploiting environmental regularities — what embodied knowledge amounts to. Operational closure turns that knowledge into an Umwelt and norms. Multicellularity demands coordination; mobility demands anticipation; a widening behavioural repertoire demands selection, discrimination and affect; social life turns other agents into recursive variables; language and culture accumulate models outside the individual.

Two mechanisms hold the chain together. Through *encapsulation of competence*, each layer operates on what the layer below has solved without rebuilding its details. A shared directional tendency shifts testing from slower, external, irreversible processes toward faster, internal,

partly reversible ones. Throughout, one computational function persists: constraining possibility so viable action stays probable under limits of time, energy and information.

The framework yields testable expectations: functionally equivalent, structurally distinct solutions in independently evolved lineages; characteristic failures in artificial systems given externally supplied value functions; affective complexity tracking the dimensionality of an Umwelt rather than absolute brain size.

Keywords: evolution of cognition; basal cognition; operational closure; predictive processing; biological computation; niche construction; evolutionary epistemology

Introduction

Predictive and Bayesian accounts of the brain have been remarkably successful (Rao and Ballard, 1999; Clark, 2013; Clark, 2016). They explain how an organism weighs what it expects against what it senses, how it allocates confidence, how it corrects itself when the world surprises it. The historical construction of those terms is usually not their central explanandum: consider what such a model must be handed before it can run — a space of hypotheses, a set of priors over them, a repertoire of available actions, and a function that says which outcomes are better than others. Each is supplied from outside the formalism. Ask a predictive model why *this* variable rather than another counts as evidence, why *this* outcome rather than another counts as bad, and the model has no answer — not because it is wrong, but because the question falls outside what it was built to address.

We take that question to be the biological one. Priors, values and hypothesis spaces are not axioms. They are sediment. Something built them, over a very long time, and the process that built them left its shape in what they are. Our aim is to reconstruct that process.

Many influential accounts begin with nervous systems, representations or learning architectures already in place, treating what precedes them as a nameless antechamber, "pre-cognitive," an undifferentiated before. We invert this, with a growing literature that locates cognition across the tree of life (Lyon et al., 2021; Baluška and Levin, 2016; van Duijn, 2017). Cognition is the management of a problem that is as old as life: an organisation that must rebuild itself cannot rely on unconstrained chance for the matter and energy it needs, and so must exploit regularities in its surroundings. What appears late is not cognition. It is one implementation of it. This shifts the explanandum. The question is no longer *when does mind begin* — a question that invites definitional argument and settles nothing — but *what new demand was each architecture meeting?* Asked that way, capacities that otherwise arrive as mysteries — anticipation, affect, symbols, culture — acquire traceable origins.

Our thesis is that the chain from replication to culture is held together by *conditional necessity*: outcomes are contingent, but each expansion of an organism's possibility space generates functional demands that any viable continuation must meet by some equivalent architecture. Reproduction does not require mobility, but it does require sufficiently regular access to matter and energy. Multicellularity does not require neurons, but it does require coordination among units. Complex mobility does not require any particular neural architecture, but it does require integration and the compensation of delays. Sociality does not require human language, but it does require means of regulating interaction with agents who are themselves anticipating you. Where the chain goes is open. That it must go somewhere is not.

Two mechanisms hold it together, and both answer an obvious objection in advance — that a demand which is merely whatever the survivors turned out to satisfy would make conditional necessity a redescription of what happened, not an explanation of it. First, through the *encapsulation of competence*, each layer operates on what the layer below has already solved without rebuilding its details: this is why cognition carries the historical depth it does, and why priors exist at all — a prior is an encapsulated competence, inherited from a slower clock. Second, testing itself shifts — not as a single currency, but as a shared direction, from the slower and external toward the faster and internal — and this demand can be named in advance of the evidence, without inspecting what evolved. A demand that forbids something, and a demand nameable before the fact, is not a redescription.

Figure 1 sets out the three claims together, since their relation is easier to see than to state: the chain runs left to right, encapsulated competence accumulates beneath each link, and the price of a test falls across the whole span while one computational function runs underneath all of it.

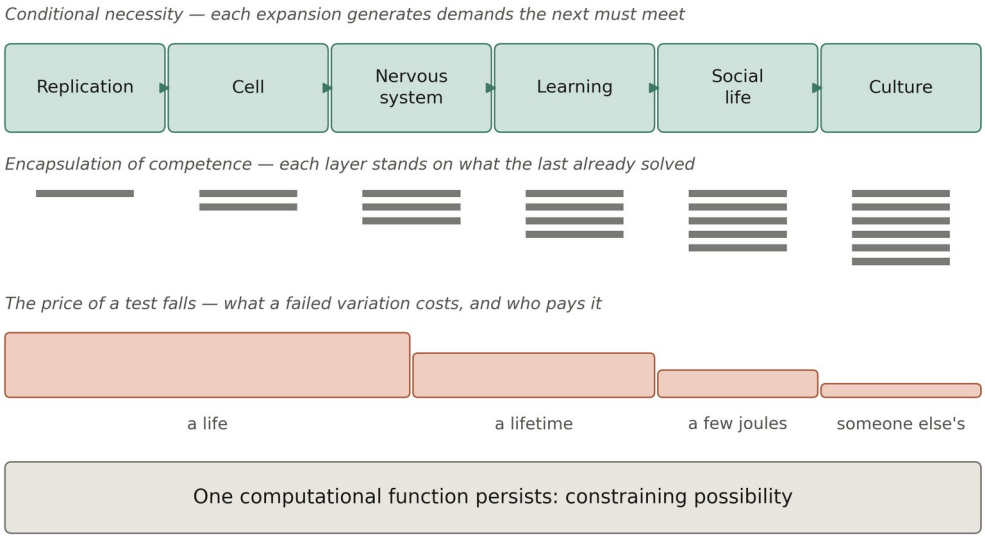

Figure 1. The framework in one view. *Top:* the chain of layers, each expansion generating demands that any viable continuation must meet — conditional necessity. *Second register:* encapsulated competence accumulating beneath each layer, which is why cognition carries the historical depth it does. *Third register:* the falling price of a test. The four rungs deliberately do not correspond one-to-one with the six links, because the claim is a shared direction rather than a common unit: what is risked shifts from a life, to a lifetime, to a few joules of internal rehearsal, to a test someone else already paid for. *Bottom:* the computational function that persists throughout. Multicellularity is folded into 'Cell' for legibility; the chain as developed in the text has more steps than the figure shows.

Three traditions have carried pieces of this argument without being made to speak to one another. Dennett (1995, 2017) established that evolution accumulates enormous competence without comprehension — continuity seen from outside. Maturana and Varela (1980) established the complementary half: an organisation that must continuously produce itself generates norms of its own, not norms an observer finds it useful to ascribe to it — continuity

seen from inside. Piaget (1971), in *Biology and Knowledge*, anticipated the encapsulation of competence developed here, though the precise operation-on-operation formulation is worked out more fully in his later writings; his epistemological programme did not survive the collapse of his stage theory, and went down with it undeservedly. Our account requires all three, and supplies what none alone does.

The paper proceeds as follows. We first trace the chain itself, from the origin of implicit knowledge in replication through operational closure, the nervous system as a change of clock, and the layered demands running from multicellularity to language and culture. We then name the currency that orders that chain — the falling price of a test — and defend the sense of computation the argument requires, one broad enough to include the bacterium and narrow enough to exclude the stone rolling downhill. We close with the predictions the account makes and the observations that would count against it.

Replication and the origin of implicit knowledge

The minimal starting point is a structure that produces others sufficiently like itself. Reproduction requires stability — a recognisable organisation must survive across copies — and it requires variation, since without heritable difference there is no space of evolutionary change. Variants differ in their capacity to persist and to reproduce under particular conditions, and so the composition of populations shifts across generations. It is useful to think of this as a search, provided the metaphor is not overread: there is no prior representation of the problem, no global objective, no goal toward which the process is directed. There is a distributed exploration of configurations, produced by variation and differential reproduction.

Replication by itself does not yet carry this argument. A bare replicator — a molecule copied by machinery it did not build, a virus executed by a host — can persist without exploiting anything on its own account; the relevant regularities are exploited by whatever copies it, not by the replicator. What follows applies to a narrower case: an organisation that must maintain, through its own activity, the material and energetic conditions of its own reconstruction. It is this autonomous organisation, not replication in isolation, that cannot rely on unrestricted chance.

But even the most elementary such organisation faces an immediate problem. To produce another structure like itself it needs matter and energy, and it must persist long enough to complete the reconstruction. Viability depends on regular exchanges with conditions the system does not fully control. It is exceedingly improbable that a complex biological organisation could maintain itself if the acquisition of each component depended on unrestricted chance. Activity may contain stochastic elements, but they must be biased, channelled or corrected by some relation to the state of the environment.

A bacterium may explore space through random components in its locomotion. But when the duration of its runs is modulated by the temporal evolution of a chemical concentration, the search ceases to be unrestricted (Colin et al., 2021). The system exploits a regularity: if conditions improve while a direction is held, it pays to prolong the run; if they worsen, it pays to change it. The organisation need not represent a food source or compute a map. It must incorporate a relation between environmental change, internal state and activity. In this elementary sense there is already knowledge — not conscious, not propositional, not symbolic,

but a material incorporation of the differences and covariations relevant to the system's continuation (Lyon et al., 2021). It is distributed across membrane permeability, receptor specificity, metabolic and regulatory networks, morphology and the possibilities of action (Bray, 2009).

Life thus involves, from the outset, two inseparable dimensions. One is reconstructive: how to produce the organisation again. The other is operative-ecological: how to obtain the conditions that make reconstruction possible. The genetic code alone does not suffice; it operates within a cellular machinery that already knows, in a functional sense, how to recognise materials, transform energy, repair components and respond to perturbations. Reconstructive information and operative knowledge presuppose one another.

Evolution as accumulated embodied knowledge

Differential selection has a cumulative consequence. Organisations that persist leave behind not only sequences but configurations that work under certain conditions. Evolution conserves competence. Following Dennett (2017), competence precedes comprehension: an organisation can solve a problem without identifying it as such, and exploit a regularity without representing it. The absence of reflective understanding does not remove the knowledge embodied in the solution.

Persistent biological structures contain a compressed history of relations with the environment. A receptor expresses the recurrence of certain molecules. A metabolic pathway incorporates exploitable chemical relations. A sensory structure expresses physical properties of the medium. A body form incorporates relations among material, gravity, friction, fluid and movement. No isolated trait is a proposition about the world; the knowledge lies in the organised whole and in the relations that make a competence possible. A receptor without transduction, effectors or metabolism resolves nothing on its own. We can therefore state what "embodied knowledge" designates: the set of recurrent environmental relations incorporated into an organisation capable of using them to maintain, regulate and reproduce its own dynamics.

This expansion is at once material and cognitive. A new capacity to inhabit water, land or air is not merely a change of location. It constructs a different relation to gravity, pressure, temperature, movement and other organisms. Each form of life constitutes a new set of relevant differences and possibilities of action — a point we develop as the Umwelt in 'Operational closure, normativity, and the Umwelt'.

The genetic code. The clearest instance of embodied historical knowledge — and the one on which much of the subsequent chain depends — is the genetic code itself. The sequence that is copied does not resemble the structures and activities it eventually helps to produce. A codon bears no pictorial resemblance to the amino acid it specifies, just as a word need not resemble its referent. The correspondence is chemically implemented and historically stabilised, but it is not dictated by resemblance alone; its meaning depends on a translation system that relates one class of molecular differences to another (Barbieri, 2015). It is, in this sense, one of the foundational transitions Maynard Smith and Szathmáry (1995) traced through the history of life — the one every later transition in this chain presupposes.

This relative independence between sequence and consequence is not a limitation of the code. It is one of its principal evolutionary powers. A sequence can be copied, recombined and varied without first constructing and evaluating the complete phenotype to which it may contribute. The production of variants is thereby partially separated from the test of what those variants do. Their functional consequences are realised later, through expression, development, organism–environment interaction and differential reproduction.

The separation is not absolute. Replication, mutation and expression all have material costs, and many sequence changes are constrained by the molecular machinery that generates them. Nevertheless, the code makes a crucial redistribution possible: the proposal of a variation can be comparatively cheap, while the more expensive evaluation of its consequences occurs on another organisational level and often on another timescale.

In this sense, the genetic system provides an early and particularly literal instance of competence encapsulation. The copying machinery need not predict what a protein will do in order to preserve and reproduce the sequence that can generate it. Nor must each new organism reconstruct the historical process through which the mapping between codons and amino acids became established. That correspondence is inherited as an operative competence of the system.

The genetic code also anticipates the pattern later described as the falling cost of a test. A modification becomes cheaper to propose to the extent that the vehicle in which it is proposed can be manipulated independently of the full consequences it will eventually produce. Evolutionary testing remains expensive — variants must ultimately confront development, physiology and the environment — but the generation and transmission of candidate variations no longer require the prior construction of every possible outcome.

What this separation opened is hard to calibrate, because nothing before it in the history of life gives a scale to measure it against. Earlier forms of chemical memory were more directly continuous with the processes they retained: a concentration, a molecular conformation or a self-maintaining reaction network stored history in the same physical dynamics through which that history remained effective. Template-based sequence inheritance introduced a more discrete and recombinable register. With a small alphabet and combinatorial syntax, biological systems could generate an open-ended space of sequences vastly larger than the set any lineage could ever explore.

The mature genetic code therefore marks a decisive transition toward digital biological memory. Four nucleotide bases, organised into triplets and interpreted through a translation apparatus, support a combinatorial space capable of specifying an enormous diversity of proteins. The importance of this transition is not simply quantitative. It changes the relation between memory, variation and consequence: hereditary descriptions can be recombined according to rules that are partially independent of the phenotypic operations they later help to construct.

Sexual reproduction adds a further operation on that recombinable memory. Clonal lineages explore mainly by modifying an inherited genome as a whole; recombination instead assembles components that arose and were tested separately, in different lineages — changing not just the rate of variation but its generative structure, since hypotheses can now be composed from parts selection has already touched. The advantage is not automatic: recombination breaks

apart favourable combinations as readily as it creates them (Otto and Lenormand, 2002). Where it works, it is evolvability as encapsulation of competence — what one lineage already tested becomes material for another's proposal.

The cell, and the encapsulation of competence

The cell is one of the great solutions in the history of life. It brings together delimitation, selective exchange, energy transformation, synthesis, repair, regulation and reproduction within an organisation that actively produces and conserves its own components. The membrane separates without isolating; it permits some flows and impedes others, sustains gradients, and helps define an operative individuality. The cell already contains an extraordinary quantity of embodied knowledge: it distinguishes molecules, conserves electrochemical differences, responds to damage, coordinates phases of its cycle and modifies its activity according to available resources. Every multicellular organism will be built from units that already possess these capacities.

New layers do not annul the previous ones; they use them. Multicellularity does not replace the cell; it reorganises it within a wider architecture. Cells go on metabolising, regulating gene expression, maintaining gradients and repairing components — but they must now also coordinate, differentiate, limit their proliferation and contribute to an organisation that exceeds them.

We call this the *encapsulation of competence*, and it does more work in this paper than any other single idea. A new layer can operate on competences that are already resolved without controlling their details. The multicellular organism does not reinvent protein synthesis. The nervous system does not compute the metabolism of its neurons. The planning of an action does not administer each muscular contraction or ionic exchange. The term does not imply that lower layers are sealed off — influences run upward and downward, and metabolic state can profoundly modify neural activity. What is encapsulated is the *operative competence*: a new layer can use what the previous one already knows how to do without reconstructing its mechanisms.

The cell is therefore not a mere support for thought. Its possibilities and constraints run through every later form of organisation. Neural excitability derives from properties of membranes and channels; plasticity uses ancient molecular mechanisms; cognitive activity depends on energetic budgets, ionic regulation, repair and development. Understanding cognition requires acknowledging that historical depth.

Operational closure, normativity, and the Umwelt

Operational closure (Maturana and Varela, 1980; Varela et al., 1991; Thompson, 2007) lets us see why biological knowledge cannot be described as the simple reception of information from the world. The environment does not deposit objects, nutrients, threats or opportunities into the organism already constituted as such. It produces physical perturbations — chemical, mechanical, thermal, electromagnetic, social — whose effects depend on the organisation that receives them. Glucose is not food in itself for any organisation whatever; it is food for a system able to detect, incorporate and use it. A sound frequency is not a social signal by physics alone;

it becomes one within an organism and a history able to produce and recognise that relation (Barbieri, 2015).

Closure does not mean the organism is independent of the world or that it constructs reality arbitrarily. It is radically open to matter and energy and depends utterly on external conditions. But those conditions do not specify its internal states directly. The perturbation triggers transformations within a network organised by the system itself — a more formal restatement of the same point distinguishes constraints, which conserve their form across the relevant timescale, from the processes they act on, and treats closure as the mutual dependence of a set of such constraints (Montévil and Mossio, 2015). And this organisation introduces normativity (Barandiaran et al., 2009). The possible states are not equivalent: some conserve the dynamics, others disorganise it. The difference between favourable and unfavourable arises from the precariousness of an organisation that must be maintained continuously. Organisms need not represent an abstract goal of survival; they operate through proximal goals — conserve a concentration, repair a lesion, approach a source, complete a developmental sequence, protect an offspring, preserve a relation. Preservation and reproduction are not explicit orders issued to the organism. They are the historical conditions under which architectures of regulation and action persisted.

That the difference between favourable and unfavourable is the organism's own difference — not a distinction an observer finds useful — is what the design stance leaves open and what the rest of the argument requires. Value has an organisational origin: a formalism whose value function is supplied from outside cannot recover it, and 'The status of computation' uses this to separate natural computation from its artificial limiting case.

The concept of Umwelt names the domain of differences, actions and relations that exists for a given form of life (von Uexküll, 2010). It is not a selection carved out of a complete, ideal perceptual reality that was initially available; that totality never exists for the organism. A particular organisation constitutes, through its sensitivity, body, activity and needs, its own domain of relations. The Umwelt of a bacterium includes gradients, metabolisable substances and conditions that modify its motility. That of a mobile animal incorporates trajectories, obstacles, surfaces, refuges and other organisms. That of a social animal includes identities, hierarchies, bonds and expectations. The human Umwelt adds institutions, obligations, narratives, absent objects, imagined futures, explanations and norms. These are not arbitrary or incommunicable worlds; all are realised within one materiality. But each organisation makes certain differences — and not others — functionally exist.

The Umwelt comprises possible actions as well. To perceive a surface as walkable, an object as manipulable or an individual as an ally depends on the body's capacities and the agent's history. And so evolution does not merely conquer pre-given niches. Variants that exploit a new resource, produce a new form of locomotion or establish a new relation to other organisms modify the very space of problems and opportunities (Odling-Smee et al., 2003). They open new niches and new Umwelten. Each solution creates difficulties that did not exist before. Mobility increases access to resources but demands orientation and bodily control. Predation opens an energy source but generates new pressures of detection, pursuit, concealment and defence. Cooperation offers protection and learning but introduces deception, conflict, reputation and dependence. Through the activity of organisms, evolution modifies the space of possibilities it must go on exploring.

The organism as a dynamic model

If an organisation incorporates regularities of the environment, it constitutes an implicit model of its world — but "model" must be made precise. It is not necessarily a separate internal representation, an image, or a static correspondence between a structure and an external property. The biological model is fundamentally dynamic. The organism must capture how the relevant variables vary and covary: which changes tend to precede others, which spatial configurations permit which actions, what trajectory an object may follow, what sensory consequences its own movements produce. Knowing a gradient need not mean representing a map; it may consist in relating temporal changes of concentration to changes of movement. Knowing an object may consist in anticipating how signals will transform as the organism moves around it: an electric fish reading its self-generated field identifies an object less from any single electric image than from how that image flows as it swims past (Hofmann et al., 2013). The correspondence is not between static things but between dynamics.

This connects with the dynamical-systems tradition in cognition, in which cognitive states are trajectories, attractors and transitions within coupled systems rather than stored units processed in sequence (Thelen and Smith, 1994; van Gelder, 1998). Organism and environment form a system of continuous interactions across multiple timescales, with delays and feedback loops. The existence of delays makes anticipation necessary. Transduction, propagation, processing and motor execution all take time; a system that reacted only to an already-registered state would always act on a world partly past. Coordination requires producing present states that anticipate, within some margin of error, future conditions. Catching a moving object requires estimating a trajectory. Perceiving during movement requires anticipating the sensory changes generated by one's own action. Forward models, efference copies and corollary discharges are neural solutions to this problem: the system combines its current state with the intended action to predict the sensory consequences, and the discrepancy between predicted and actual lets it distinguish external change from self-produced change, correct movements and update the model (Wolpert et al., 1995). Weakly electric fish make the mechanism explicit. Each electric organ discharge is triggered by a command whose pathway also issues corollary discharges (Grant et al., 1999); in cerebellum-like structures of the electrosensory lobe these build a negative image of the expected self-generated input and subtract it, leaving the sensory surface free to register the world rather than the fish's own signal (Bell et al., 1997). The general logic, however, is older than nervous systems. Cellular regulatory networks already relate recent history, internal state and external conditions. Nervous systems expand that capacity within the individual life.

The organism-as-model is not a metaphor we are free to take or leave. It is a theorem of regulation: a regulator that holds a variable within bounds against a range of disturbances must embody, in its own organisation, a model of the regularities that produce those disturbances. Anything that regulates well is, to that extent, a model of what it regulates — a result we return to because it grounds the claim that knowledge is coextensive with viable organisation rather than with representation. (Note 2.)

Bayesian and predictive formulations offer one way to make this dynamic conception precise (Rao and Ballard, 1999; Clark, 2013). Perception does not reconstruct a stimulus from the bottom up; it combines prior regularities with changing evidence to estimate the most plausible causes of the signals. This is compatible with operational closure: the data do not specify their

own cause or meaning; they are interpreted from a prior organisation (Allen and Friston, 2018). But — and this is the pivot of the paper — the biological priors do not begin as explicit neural beliefs. Evolution has already delimited what signals can be detected, what associations can be learned, what actions are available, what states must be regulated, what transformations are expectable. Morphology, sensory systems and motor repertoires function as prior constraints on the space of hypotheses and actions. Development calibrates those dispositions; ontogenetic learning introduces faster adjustments; perception and action produce real-time updates. The predictive machinery describes how expectations and evidence combine. It is the evolutionary and developmental history that explains where the relevant variables, the priors, the value functions and the action space came from.

That developmental calibration is not a black box; Piaget gave it a mechanism, and one that anticipates the shape of the account offered here. A schema, for Piaget, is not a stored copy of the world but an organisation of action — a repeatable way of assimilating some class of encounters to what the organism can already do (Piaget, 1971). Development proceeds by two coupled processes: assimilation, which folds a new encounter into an existing schema, and accommodation, which adjusts the schema when the encounter resists it. Neither operates alone; a schema that only assimilated would never change, and one that only accommodated would never generalise. What makes the developmental sequence cumulative rather than merely successive is a further move, reflective abstraction: a later stage does not discard the coordinations achieved by an earlier one but takes them as its object, operating on operations rather than rebuilding them (Piaget, 2001). This is competence encapsulation on a faster clock: the same logic installing itself within a lifetime rather than across generations.

A remark on the most ambitious version of this picture is in order. The free-energy principle (Friston, 2010) casts an organism as a system that minimises the discrepancy between the states it expects to occupy and those it finds itself in, and in doing so recovers much of what we have said: that to persist is to embody a model of one's world, that perception is inference, that action serves prediction. Our account is not a rival to it but a question put to its premise. The free-energy formulation describes the dynamics of a system already equipped with a generative model — with priors, with a distribution over the states it treats as characteristic of itself. It takes that model as given and derives the behaviour of a system that has one. We ask where the model came from. The principle does gesture at an answer — evolution and development are said to supply the generative model, and an organism's very form is, in a sense, a model of its niche — but the gesture is left largely undeveloped, a boundary condition we take as our subject rather than an object of study. What installs the priors, what fixes the states an organism expects to occupy, what makes some outcomes count as surprising in the first place: these are not features of the formalism but sediment of a history, and the mechanisms that laid down that sediment — the encapsulation of competence across timescales, the falling cost of a test — are what the formalism receives ready-made. 'The status of computation' makes the same point about Turing machines themselves, whose alphabet and halting criterion are just as thoroughly supplied from outside.

The layered chain: from multicellularity to culture

The history of cognition can be reconstructed as a sequence of layers, each adding possibilities and, with them, new necessities. The realisations are contingent; the demands generated by the

preceding conditions are more general. We state the chain compactly, since the point is the pattern rather than any single link.

From cell to multicellularity. Associating cells allows greater size, specialised function and distributed labour, but it generates problems of coordination, communication, development, repair and control of proliferation. A population of cells is not yet an integrated organism. Different multicellular lineages evolved different mechanisms; but once a stable multicellular organisation exists, some form of coordination and differentiation becomes necessary.

From multicellularity to nervous systems. In small, slow organisms, local chemical or electrical signals can suffice. Increased size, mobility and the need to coordinate distant effectors create new temporal demands: signalling must be fast, selective and able to integrate perturbations from different parts of the body. Nervous systems are a particularly effective realisation of these demands — not the only conceivable one, and possibly not of single origin (Moroz and Kohn, 2016). The nervous system did not invent biological electricity; it inherited it. Voltage-gated channels and action potentials are molecular tools that long predate neurons — the depolarisation–contraction–secretion coupling at the core of neuromuscular circuits is reconstructed as already present in unicellular stem eukaryotes (Brunet and Arendt, 2016) — and large unicellular organisms already used them to coordinate their own movement. What changed with nervous systems was not electricity but its organisation.

From inherited control to individual learning. Inherited mechanisms resolve recurrent conditions, but a mobile organism meets local variations that cannot be specified in advance. Learning modifies behaviour according to individual history. Forms of plasticity and habituation exist in systems without neurons (Lyon et al., 2021; van Duijn, 2017); nervous systems amplify and organise this capacity on a new scale. Associative learning detects contingencies among signals, events, actions and consequences: an initially neutral signal can acquire predictive value, and the organism can act on what a signal announces before the relevant event arrives. Complex integrative centres evolved convergently in insects, cephalopods, birds and mammals, using distinct architectures to solve problems of memory, learning and multimodal integration (Roth, 2015). The convergence suggests that behavioural expansion generates recurrent functional necessities whose anatomical realisations are not unique. This architectural diversity itself follows a recognisable sequence of transitions — networked, centralised, recurrent, laminated and reflective — each marking a change in computational architecture rather than in the underlying physics, and each building on the encapsulated competence of the one before (Barron et al., 2023).

Learning what to learn. A system cannot process everything with equal depth. Learning requires data-acquisition mechanisms that determine what to observe, for how long, and under what conditions (Lotem and Halpern, 2012). The rules of learning and the mechanisms of acquisition must coevolve: the statistical structure of the environment cannot be exploited unless the organism directs its activity toward the appropriate variables (Godfrey-Smith, 1996). Attention, curiosity, exploration and perceptual predispositions are part of the cognitive problem. The organism does not simply receive a dataset; it acts to produce the data from which it will learn.

From physical niche to social world. Other organisms are part of the environment from very early on — as resources, competitors, predators, symbionts or mates. But stable social life

introduces a particular complexity: the relevant variables are other agents, themselves able to act, learn, anticipate and modify their behaviour. The social environment is not passive; each participant changes as a function of the others, and the regularities are recursive — what one agent does depends on what it expects another to do, which may depend in turn on what that other attributes to the first. This generates new cognitive necessities: recognising individuals, anticipating intentions, evaluating alliances and conflicts, detecting deception, managing reputation. These capacities do not replace the earlier ones; they reorganise perception, memory, bodily regulation, learning and affect within a more complex relational space. Biological viability comes to be mediated by social organisation.

Language, culture and extra-corporeal accumulation. Language transforms the cognitive space because it lets differences be stabilised, combined and transmitted without depending on the immediate presence of their referents. Humans can coordinate around absent objects, past events, possible futures, norms and abstract entities. Language does not arise outside bodily and social capacities — it reuses auditory and visual perception, motor control, memory, categorisation and joint attention — but, once constituted, it opens a new space of possibilities: to condense experience, share models, formulate alternatives, preserve instructions, represent counterfactual relations, and transmit knowledge across generations. Culture extends accumulation beyond the individual organisation: knowledge can be stored in tools, practices, narratives, texts and institutions (Clark, 1997), so that no individual need reconstruct from scratch the solutions used by a community. A new encapsulation appears: the agent uses cultural competences without understanding all the historical and technical processes that produced them. And culture creates problems that did not exist before: humans must manage identity, obligation and legitimacy, and can suffer over symbolic losses and institutional exclusions. We are organisms and, simultaneously, social and cultural agents. A threat to reputation can recruit ancient bodily mechanisms; a norm can redistribute biological resources; a belief can reorganise reproductive or alimentary behaviour.

Affect, across the chain. One demand recurs at every level above the simplest and deserves to be named on its own, because it is easy to treat as a late colouring of an otherwise neutral process, and it is not. When multiple signals, actions and outcomes are possible at once, the organism must determine what matters *now*. Life already introduces biological value: some states favour the continuation of the organisation and others compromise it. With nervous systems able to integrate multiple needs, temporal horizons and courses of action, that valuation becomes affect. Affect does not add an emotional tone to a prior, neutral cognition; it participates in defining what will be perceived, learned, remembered and selected. It raises the priority of some signals, modifies the precision assigned to some predictions, changes the rate of learning, regulates effort, favours exploration or avoidance, and sustains goals across time.

Selection has an input side as well, and it makes a demand of its own. A widening repertoire of possible actions is worthless unless the situations that call for different actions can be told apart. Selection therefore makes two demands at once, not one: affect, to weigh what matters among the options on the output side; and a correspondingly rich and discriminating sensory report on the input side, fine enough to distinguish the situations the options answer to. The two must grow together — requisite variety read in both directions (Ashby, 1956) — since discrimination without a repertoire to exploit it is a cost without a return, and a repertoire without discrimination to guide it is a capacity that cannot be aimed. This is why the elaboration of the senses and the elaboration of behaviour track one another across the tree of

life: an animal does not evolve a finer retina or a richer chemosensory array for its own sake, but because a body able to do more with the world needs, in equal measure, to tell more of the world apart.

Available actions are specified and compete in parallel, modulated by value, cost and probability of success (Cisek, 2007); dopaminergic prediction-error signals update values according to the discrepancy between expected and obtained outcomes (Schultz, 2016); affective, attentional and executive networks are integrated rather than separate (Pessoa, 2017). Affect, on this account, is the dynamic organisation of value and relevance within the space of states, actions and possible futures of an agent — and the higher the dimensionality of the Umwelt, the larger the problem it solves. Language and culture will expand that space extraordinarily: a person can be affected by absent events, hypothetical futures, norms, symbolic losses and conflicts of identity. Human affect keeps its ancient regulatory roots while operating within social and symbolic worlds.

The price of a test

We have described the chain and named the demands that link its steps. We have not yet named the currency in which those demands are paid. There is one: selection retains whichever architecture pays it more cheaply, and this is why the sequence has the order it has.

Call it a currency, but not a common unit. What evolution risks in a failed test is a life; what a nervous system risks in a rehearsed action is a few joules and a moment's delay; what a culture risks in transmitting a practice is the possibility of the practice being wrong. These are not exchangeable amounts of a single stuff, and nothing below trades on the pretence that they are. What they share is a direction: from the external, the slow and the irreversible toward the internal, the fast and the partly reversible. That direction is the pattern we track.

No system can know in advance which variation will work — not, at least, when the problem is of the kind biology faces. This has to be stated carefully, because it is not true of *every* system that changes. Where the problem is transparent — linear, stationary, low-dimensional, its structure known in advance — a viable modification can be computed rather than tried: the answer is available in closed form, and no test need be paid for. Engineering does this routinely. The cost of a test is a function of the *opacity* of the problem, not of change as such. What makes biological problems the expensive kind is precisely that they are not transparent: they are high-dimensional, non-stationary, and — because the organism's own activity remakes its niche — they do not sit still long enough to be solved once. It is for problems of that class, and only for them, that a viable variation must be proposed and tested against the world rather than derived beforehand; and cognition lives entirely inside that class.

For such problems the constraint is informational and unavoidable. To know in advance that a modification will succeed, a system would need a model of its world complete and accurate enough to evaluate the modification without trying it — and a system already in possession of such a model would have no need to vary. The formal statement is familiar: absent prior assumptions about the structure of the problem, no search strategy outperforms any other (Wolpert and Macready, 1997). Selection is post hoc because, for problems of this class, it cannot be otherwise.

The point is old. Campbell (1960) made it the core of evolutionary epistemology — blind variation and selective retention as the engine of every genuine increase in fitted knowledge. Popper (1979) made it epistemology proper. Dennett (1995, 2017) built the staircase: Darwinian, Skinnerian, Popperian and Gregorian creatures, each with a better way of generating and testing. What the staircase has lacked is an organising principle. Dennett catalogues the floors; he does not say what quantity varies systematically across them. We propose that one does — the cost of a test — and that its fall is what the ordering of the floors tracks.

Every layer we describe tests more cheaply than the last: an architecture that does outcompetes one that does not, whenever both are available to selection.

Evolution tests through differential reproduction. A variant that fails outright is removed by death; far more often it simply reproduces less, and the test is paid in descendants rather than in the organism itself. Either way the unit of testing is a life, spent or discounted, and the clock runs in generations. Individual learning tests by trying: a failed trial can still be costly, sometimes lethal, but typically it costs an injury, a missed meal, a wasted afternoon — cheaper than a lineage, because it need not end one — and the clock falls from generations to a lifetime. Nothing about knowing has changed — the price has. A nervous system with forward models tests inside: it runs a candidate action against its own model and discards it before a muscle moves, so what is risked falls again, to a few joules and a moment's delay, and the clock to milliseconds. This is precisely the Popperian move: letting our theories die in our stead (Popper, 1979). Language and culture test with someone else's experience: the cost falls sharply for whoever inherits the result, though never to zero — transmission, uptake and correction all cost something — because most of what was originally risked was paid by another, sometimes by someone dead a thousand years. Cumulative culture is a mechanism for inheriting tests one did not run at their original price.

A minimal two-timescale demonstration. To test the internal coherence of the proposed mechanism, we implemented a minimal two-timescale model, specified in full in the Supplementary Material (Sections S1–S2). The model is not offered as a novel learning architecture, nor as evidence that biological evolution followed this particular implementation; its structure is familiar from the Baldwin-effect and meta-learning traditions (Baldwin, 1896; Hinton and Nowlan, 1987). Its purpose is narrower: to determine whether competence acquired on a slower clock can be encapsulated as a prior for a faster one, and whether that transfer reduces the number of population-level selective evaluations required to attain a common level of competence. In brief, the set-up is a world with two parts: structure that holds across lives, and detail that changes from one life to the next. Two clocks run on it. The slow one can only ever discover what is stable, because that is all it sees repeated; the fast one must handle whatever the slow one could not have known. The question the model puts is whether the slow clock's product becomes the fast clock's starting point — whether a prior is, as we have claimed, an encapsulated competence inherited from a slower process — and whether that inheritance lowers the number of selective evaluations the slow clock must pay.

Results: encapsulation of environmental structure. The inherited initialisation progressively approached the stable mean of the task family. The Euclidean distance ||m − mu|| *fell from 1.676 to 0.720, and the final cosine similarity between m and mu* was 0.964. This does not mean that the inherited prior solved each task: individual tasks still varied around mu* and required within-life adjustment. It means that the slow clock captured statistical structure

shared across lives. When the evolved prior and a zero initialisation were evaluated on exactly the same task realisations, each with its independently selected learning rate, the evolved prior reduced the time to the within-life criterion by 3.10 steps on average (paired bootstrap 95% CI 2.15–4.12). The result operationalises encapsulation: information acquired by the slow selective process is made available as the starting condition of the faster learning process (Supplementary Figure S1A).

Results: prior and plasticity as a functional bundle. A 2x2 ablation crossed inherited prior present/absent with lifetime plasticity present/absent. The observed mean fitness of the combined prior-plus-plasticity condition was the only point estimate to exceed the externally fixed criterion (−1.244; 95% bootstrap CI −1.281 to −1.207). Plasticity without the evolved prior improved performance but remained below criterion (−1.976; 95% CI −2.032 to −1.920). The evolved prior without plasticity could capture the stable family mean but not lifetime-specific departures from it (−4.518; 95% CI −4.664 to −4.373). Removing both components produced the poorest performance (−7.434; 95% CI −7.660 to −7.213). The combined condition's confidence interval narrowly overlaps the criterion, so the robust claim is complementarity rather than an all-or-none threshold effect: the prior captures what is stable across lives, while plasticity captures what varies within them (Supplementary Figure S1B).

Results: lifetime learning lowers slow-loop selective cost. The selective-cost experiment compared otherwise matched evolutionary runs with and without within-life learning. All six runs with lifetime learning crossed the common competence criterion, at generations 13, 6, 5, 10, 11, and 12. The median crossing was generation 10.5, corresponding to approximately 345 population-level selective evaluations when the initial generation is included. None of the six runs without lifetime learning reached the same criterion within the 50-generation budget. The result supports a precise form of the falling-price claim. It does not show that learning is costless, nor that an ontogenetic trial is equivalent to a biological death. It shows that adding a fast adaptive loop reduces the number of slow-loop selective evaluations required to reach an identical, independently specified level of competence. Part of the adjustment that would otherwise have to be purchased through changes in population composition is transferred to a faster and more reversible process within individual lives (Supplementary Figure S2C).

Results: a common demand admits multiple implementations. To distinguish functional necessity from commitment to a single mechanism, the lifetime update was implemented using two structurally distinct rules: ordinary stochastic gradient descent (SGD) and normalized least mean squares (normalized LMS). Across four independent runs per rule, both implementations exceeded the same competence criterion. Mean validation fitness was −1.096 (SD 0.025) for SGD and −0.923 (SD 0.021) for normalized LMS. The rules did not produce identical performance and are not claimed to exhaust the possible architectures. Their relevance is narrower: a common functional demand can be met by more than one update mechanism (Supplementary Figure S2D).

Interpretation and scope. The model is a demonstration of internal consistency, not a historical reconstruction and not empirical evidence that natural evolution uses these particular algorithms. Its components are deliberately familiar. The contribution lies in what the model is used to test: whether the same minimal system can exhibit the four relations proposed by the framework. It does. The slow clock encodes stable task-family structure in a prior; that prior accelerates adjustment on the fast clock; prior and plasticity play

complementary roles; the fast loop reduces the slow-loop selective budget required to attain a common competence; and distinct update rules can satisfy the same functional demand. The model therefore gives computational form to the paper's central interpretation of priors: the inherited vector m is the compressed record of regularities that selection has already paid to discover. Lifetime learning begins from that record rather than from an unstructured space, and the number of population-level tests needed to attain competence falls accordingly.

Read this way, the hierarchy of timescales introduced in 'The organism as a dynamic model' as a hierarchy of clocks turns out to be a consequence rather than a premise. The ladder of clocks is a ladder of prices. Speed is not the goal; speed is what one gets when a test stops costing a life.

This also supplies what a chain of conditional necessities would otherwise lack: a unifying principle. The demand to make testing cheaper can be stated in advance, without inspecting what evolved. Any lineage whose behavioural repertoire expands faces a search space growing faster than the supply of lives available to search it, and must therefore either shrink the space or cheapen the test. That is a demand, and it is prior to every solution to it — the promised second answer to the charge of circularity from 'Two mechanisms': a demand nameable before the fact does not redescribe it.

Vision need not rediscover that light falls from above — the discovery was purchased, in lives, long ago.

One caution, which turns into an argument. We do not claim that variation is blind. The claim is contested and we do not need it: facilitated variation, mutational bias and developmental constraint all indicate that variation is channelled toward regions where viable phenotypes are likelier to be found (Kirschner and Gerhart, 2005). The weaker claim suffices — variation is not *guaranteed*. Nothing establishes in advance that a variant will work, and post hoc testing is therefore ineliminable. But consider what channelled variation implies. If the space an organism explores is already biased toward the viable, then there are priors on variation itself. Evolvability is the encapsulation of competence applied to the generation of novelty: the record of what worked constrains what is tried. And this gives the same mechanism at three levels, which we did not set out to find — priors over hypotheses, which is perception; priors over actions, which is behaviour; priors over variations, which is evolvability. These are not three analogies. They are one mechanism, running on three clocks.

The status of computation

We claim that one computational function runs from bacterium to human. The claim is worthless unless the word is given a definition with teeth, and the available definition is contested.

On the classical account, computation is the rule-governed manipulation of discrete symbols over a well-defined domain. On that criterion a bacterium does not compute, and our chain is severed at its first link. The opposite temptation is to use the word loosely, as a synonym for "behaves in an orderly fashion," at which point it does no work at all and we have earned the charge routinely levelled at basal cognition: a metaphor stretched until it means nothing (Lyon et al., 2021).

We take a third route. Computation, in the sense that matters here, is the transformation of states and differences so as to constrain a space of possibilities. A chemotactic network transforms a temporal gradient into a bias on tumbling frequency, and thereby cuts an unbounded search down to a tractable one. A metabolic circuit transforms substrate availability into enzyme expression, and thereby forecloses reactions that would waste what is scarce (Bray, 2009). A cortical population transforms a pattern of afferent activity into a distribution over causes (Rao and Ballard, 1999; Friston, 2010), discarding the indefinitely many world-states that would have produced the same input — the same reduction of uncertainty, in purely informational terms, that Shannon formalised for communication (Shannon, 1948). The operation is the same in each case. What changes is speed, scope, and the timescale on which the constraint was installed.

This is broader than Turing, and breadth invites the obvious objection: a stone rolling downhill also has its trajectory constrained, and we do not wish to say that it computes. Two conditions do the excluding. The constraint must be *normatively oriented* — bound to the viability of the system that implements it, such that some outcomes are better *for that system* than others. And it must be *historically acquired* — sedimented by selection, development or learning, rather than merely instantiated by physical law. The stone's trajectory answers to neither condition. The bacterium's answers to both. Computation, on this account, is constraint plus norm plus history, and it excludes things: the stone; the thermostat, whose norms belong to whoever installed it; and, as we shall see, rather more than one might expect.

Consider what a Turing machine actually is. Its alphabet, its state space, its halting criterion, and the well-defined domain over which it operates are all handed to it from outside, by a designer who has already carved the problem. In this comparison, a Turing machine functions as the limiting case of externally supplied priors. It is the condition our account says organisms do not satisfy, and the condition our second prediction attributes to artificial systems whose value functions are set from without. To insist that Turing computation is the only legitimate sense of the term therefore inverts the order of explanation: it takes the most artificial case — the one in which the framing of the problem was solved at the factory — as the paradigm for the natural ones, which had to do their own framing and had nothing to do it with but their own history. The question of what computation is and the question in our title turn out to be one question.

This also disposes of the frame problem, and it is worth being exact about how. On our account the frame problem is not solved. It does not arise. The problem arises only for a system that must compute over a pre-given universe of possibly relevant facts and determine, from within that universe, which of them matter. No organism, and nothing in biology, has ever confronted an unstructured space. Each layer inherits a space already carved by the layers beneath it — by evolution, by development, by morphology, by the body's own repertoire of possible actions — and affect weights relevance among whatever survived the carving. This is what the encapsulation of competence buys: the organism arrives late to a world already cut down to size. The frame problem is an artefact of the classical formulation, not a fact about cognition — a diagnosis anticipated by Dreyfus (1992) and developed by Wheeler (2005), and one that follows directly from the account given here.

One apparent exception proves instructive. A system with language can, it seems, treat anything as relevant to anything, and there the space is not pre-carved — the frame problem

appears to return. But the human case is not an exception to the account; it is its most demanding instance. Relevance is carved here too, only now the carving is done by culture: by the practices, institutions and inherited categories that make some considerations salient and leave the rest unthought. And when that cultural carving comes loose from the bodily and material substrate that anchored it — when the symbolic layer runs free — the system does not gain omniscience. It fails, in the specific way that unanchored symbolic systems fail. The frame problem does not return for the cultural animal; the *vulnerability* of a carving that has forgotten what it rests on does.

Our position is close to mechanistic accounts of physical computation, which legitimise computation outside the Turing paradigm by identifying it with mechanisms that manipulate vehicles according to rules sensitive only to vehicle properties (Piccinini, 2015). We depart on one point, and the departure is not incidental. Such accounts typically require *medium-independence*: the same computation could in principle be implemented in any substrate. Biological computation as we describe it is deliberately *medium-dependent*. The membrane does not implement the computation; the membrane *is* the computation. Gradients, channel kinetics, diffusion times and metabolic cost are not implementation details lying beneath an algorithm — they are constitutive of what is computed, because they are what make some transformations cheap, others slow, and others impossible. A computation whose constraints were installed by the history of a body cannot be lifted free of that body without ceasing to be the computation it is.

Predictions and how they could fail

A synthesis earns the name of hypothesis only if something would count against it. Three expectations follow from the account, and they are not of equal strength; we say so, because a framework that presents all its consequences as equally secure invites disbelief in all of them.

Capacities should come in bundles, not à la carte. The demands we describe do not arrive singly. Moving a large body quickly through a fluid does not merely require fast signalling; it requires, indissociably, coordination across distant effectors, anticipation to cover transduction and conduction delays, a means of telling world-caused from self-caused sensory change, and sensory discrimination fine enough to tell apart the situations the expanded repertoire must respond to differently. If these demands are genuinely entailed by the expansion rather than independently invented, the framework prohibits a specific class of organism: one with sophisticated motor anticipation and no reafference-cancelling mechanism of any kind; one with a broad behavioural repertoire and a sensory surface too coarse to distinguish the conditions that repertoire is for. The bundle is the prediction; any lineage exhibiting one member should exhibit the others, however unrecognisably implemented — and the claim fails if a lineage is found that took one and left the rest. This can be sought where the evidence is not yet in: in understudied lineages, in the order in which these capacities appear during development, and in the sequence in which they appear phylogenetically. Convergent architectures in cephalopods, hymenopterans and vertebrates are consistent with the framework, but they were known in advance (Roth, 2015); the bundle prohibition is not.

Externally supplied value should produce a characteristic failure signature. Systems whose value functions are supplied externally should show a failure class that scale does not repair. Perceptual and pattern failures yield to more data and more parameters; failures of

relevance constitution — determining which variables should count when the objective underdetermines the situation — should not. The prediction is comparative and falsifiable: if the same scaling that fixes perception also dissolves relevance failures, the claim that value has an organisational origin loses its bite. This restates the Turing-machine argument of 'The status of computation' as an empirical wager.

Systems whose priors are assembled through layered encapsulation — each level consolidating what a prior level already solved, rather than re-deriving it from an undifferentiated search — should reach a fixed competence criterion with fewer evaluations and lower computational cost than systems of comparable capacity whose priors were built in one flat, single-stage process. The comparison is not compositional structure as such — a library of primitives specified by a designer's own prior analysis of the domain remains, in our terms, externally supplied, whatever its internal architecture — but structure the system itself earned by solving problems. Program-synthesis systems that grow their own library of reusable components through repeated problem-solving, rather than inheriting one from a designer, already show the pattern: solution speed and success rate improve as the self-discovered library deepens (Ellis et al., 2021).

Affective complexity should track Umwelt dimensionality, not brain size. Affective complexity — operationalised independently, for instance as the diversity of neuromodulatory systems capable of independent regulation — should covary with ecological and social demand heterogeneity, itself measured independently of the nervous system (group size, dietary breadth, foraging horizon, parental investment), after controlling for neuron number. This is the framework's most exposed claim. If absolute neural mass accounts for the variance and demand heterogeneity adds nothing, the account of affect as the solution to selection in high-dimensional possibility spaces is wrong. The operationalisations must be independent of one another, on pain of circularity — affective repertoire measured by neuromodulatory diversity on one side, demand heterogeneity by ecology on the other.

The three are not equally secure. The convergence evidence is partly retrodictive, and only the bundle prohibition genuinely risks something not yet seen. The artificial-systems prediction is short-horizon and hostage to how a fast-moving field develops. The affective prediction is the one most able to be shown false, and we regard it as the framework's principal exposure rather than its safest illustration.

Conclusion

The history of cognition begins when the persistence of an organisation stops depending on wholly random encounters and starts exploiting regularities of the world. A structure that can reproduce needs matter, energy and stability; to obtain them regularly it must discriminate, anticipate and act; and in that organisation there is already implicit knowledge. Evolution accumulates and transforms that knowledge, and new layers use what earlier ones already know how to do without rebuilding it. Operational closure means the world does not enter as a description; the organism constitutes an Umwelt in the coupling between its organisation and the perturbations it can register, and its knowledge is a dynamic model of covariations, sequences, trajectories and the consequences of action. Nervous systems widen the timescale of updating; affect organises value and relevance; sociality turns other agents into dynamic

components of the world; language and culture let models accumulate outside the individual, and generate, in turn, new spaces of uncertainty, conflict and regulation.

Two mechanisms give the chain its shape. Encapsulation lets each layer stand on the solved competence of the last, which is why cognition carries its historical depth and why priors exist at all. And testing itself shifts — not as a single currency, but as a shared direction, from the external and irreversible toward the internal and reversible — which is why the sequence runs from the organism that tests by dying to the culture that inherits tests it never ran. Through all of it, one computational function persists: to constrain and structure the space of possibilities so that viable action becomes probable under limits of time, energy and information — where computation means constraint bound to a system's own norms and installed by its own history, and where the Turing machine, far from being the standard, functions as the artificial limiting case in which that history was supplied from outside.

There is no necessary march toward the human. There are chains of local necessity. Each organisation opens possibilities, and those possibilities generate new problems; the concrete solutions may be many, but any continuation must answer, in some way, the demands its own expansion created. Human thought is not an exception standing outside natural history. It is a recent realisation of a much older process by which life built organisations able to know, dynamically, the transformations of their worlds, and to act within them. Where the priors of the predictive brain come from is not a mystery the formalism forgot to address. It is a fact about a body with a history — the compressed record of every test that history has already paid for.

Notes

Note 1. The severance is visible in the most ambitious recent synthesis of the field. Dennett (2017), *From Bacteria to Bach and Back*, traverses the full arc from prokaryote to Bach without substantively engaging with Piaget, Maturana or Varela — the traditions that place the construction of knowledge in the organism. Von Uexküll is present, credited with prefiguring the idea that each organism has its own ontology (von Uexküll, 2010). The Umwelt was taken; the machinery by which an organism builds it was not.

Note 2. Conant and Ashby (1970). We adopt the theorem's direction — good regulation entails an internal model — while declining its usual reading as an argument for representation: the model in question is the regulator's organisation, not a symbol it manipulates. On requisite variety, see Ashby (1956).

Declaration of Generative AI and AI-assisted technologies in the writing process

The author is not a native English speaker, and prepared this manuscript with substantial assistance from Claude (Anthropic), which drafted and refined the English prose from the author's direction. The author specified each argument and revision, reviewed and edited everything, and approved the final text.

The theoretical framework and its articulation are the author's own. The model also assisted in implementing the computational model reported in the Supplementary Material, in verifying citations against Crossref and publisher records — which corrected several errors — and in

producing the figures. The author takes full scientific responsibility for the content, the argument, and any errors.

References

Allen, M., Friston, K.J., 2018. From cognitivism to autopoiesis: towards a computational framework for the embodied mind. Synthese 195, 2459–2482.

Ashby, W.R., 1956. An Introduction to Cybernetics. Chapman & Hall, London.

Baldwin, J.M., 1896. A new factor in evolution. Am. Nat. 30, 441–451, 536–553.

Baluška, F., Levin, M., 2016. On having no head: cognition throughout biological systems. Front. Psychol. 7, 902. doi:10.3389/fpsyg.2016.00902.

Barandiaran, X.E., Di Paolo, E.A., Rohde, M., 2009. Defining agency: individuality, normativity, asymmetry, and spatio-temporality in action. Adapt. Behav. 17, 367–386. doi:10.1177/1059712309343819.

Barbieri, M., 2015. Code Biology: A New Science of Life. Springer, Cham. doi:10.1007/978-3-319-14535-8.

Barron, A.B., Halina, M., Klein, C., 2023. Transitions in cognitive evolution. Proc. Biol. Sci. 290, 20230671. doi:10.1098/rspb.2023.0671.

Bell, C.C., Bodznick, D., Montgomery, J.C., Bastian, J., 1997. The generation and subtraction of sensory expectations within cerebellum-like structures. Brain Behav. Evol. 50 (Suppl. 1), 17–31. doi:10.1159/000113352.

Bray, D., 2009. Wetware: A Computer in Every Living Cell. Yale University Press, New Haven.

Brunet, T., Arendt, D., 2016. From damage response to action potentials: early evolution of neural and contractile modules in stem eukaryotes. Philos. Trans. R. Soc. B 371, 20150043.

Campbell, D.T., 1960. Blind variation and selective retention in creative thought as in other knowledge processes. Psychol. Rev. 67, 380–400.

Cisek, P., 2007. Cortical mechanisms of action selection: the affordance competition hypothesis. Philos. Trans. R. Soc. B 362, 1585–1599.

Clark, A., 1997. Being There: Putting Brain, Body, and World Together Again. MIT Press, Cambridge, MA.

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

Clark, A., 2016. Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press, New York.

Colin, R., Ni, B., Laganenka, L., Sourjik, V., 2021. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. FEMS Microbiol. Rev. 45, fuab038. doi:10.1093/femsre/fuab038.

Conant, R.C., Ashby, W.R., 1970. Every good regulator of a system must be a model of that system. Int. J. Syst. Sci. 1, 89–97.

Dennett, D.C., 1995. Darwin's Dangerous Idea: Evolution and the Meanings of Life. Simon & Schuster, New York.

Dennett, D.C., 2017. From Bacteria to Bach and Back: The Evolution of Minds. W.W. Norton, New York.

Dreyfus, H.L., 1992. What Computers Still Can't Do: A Critique of Artificial Reason. MIT Press, Cambridge, MA.

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., Tenenbaum, J.B., 2021. DreamCoder: bootstrapping inductive program synthesis with wake-sleep library learning. In: Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation (PLDI 2021). ACM, New York, pp. 835–850. doi:10.1145/3453483.3454080.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138. doi:10.1038/nrn2787.

Godfrey-Smith, P., 1996. Complexity and the Function of Mind in Nature. Cambridge University Press, Cambridge.

Grant, K., von der Emde, G., Gómez-Sena, L., Mohr, C., 1999. Neural command of electromotor output in mormyrids. J. Exp. Biol. 202, 1399–1407. doi:10.1242/jeb.202.10.1399.

Hinton, G.E., Nowlan, S.J., 1987. How learning can guide evolution. Complex Syst. 1, 495–502.

Hofmann, V., Sanguinetti-Scheck, J.I., Gómez-Sena, L., Engelmann, J., 2013. From static electric images to electric flow: towards dynamic perceptual cues in active electroreception. J. Physiol. Paris 107, 95–106. doi:10.1016/j.jphysparis.2012.06.003.

Kirschner, M.W., Gerhart, J.C., 2005. The Plausibility of Life: Resolving Darwin's Dilemma. Yale University Press, New Haven.

Lotem, A., Halpern, J.Y., 2012. Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. Philos. Trans. R. Soc. B 367, 2686–2694. doi:10.1098/rstb.2012.0213.

Lyon, P., Keijzer, F., Arendt, D., Levin, M., 2021. Reframing cognition: getting down to biological basics. Philos. Trans. R. Soc. B 376, 20190750.

Maturana, H.R., Varela, F.J., 1980. Autopoiesis and Cognition: The Realization of the Living. D. Reidel, Dordrecht.

Maynard Smith, J., Szathmáry, E., 1995. The Major Transitions in Evolution. W.H. Freeman, Oxford.

Montévil, M., Mossio, M., 2015. Biological organisation as closure of constraints. J. Theor. Biol. 372, 179–191.

Moroz, L.L., Kohn, A.B., 2016. Independent origins of neurons and synapses: insights from ctenophores. Philos. Trans. R. Soc. B 371, 20150041.

Odling-Smee, F.J., Laland, K.N., Feldman, M.W., 2003. Niche Construction: The Neglected Process in Evolution. Princeton University Press, Princeton.

Otto, S.P., Lenormand, T., 2002. Resolving the paradox of sex and recombination. Nat. Rev. Genet. 3, 252–261. doi:10.1038/nrg761.

Pessoa, L., 2017. A network model of the emotional brain. Trends Cogn. Sci. 21, 357–371.

Piaget, J., 1971. Biology and Knowledge: An Essay on the Relations between Organic Regulations and Cognitive Processes. University of Chicago Press, Chicago. Original work published 1967.

Piaget, J., 2001. Studies in Reflecting Abstraction. Campbell, R.L. (Ed. and Trans.). Psychology Press, Hove. Original work published 1977.

Piccinini, G., 2015. Physical Computation: A Mechanistic Account. Oxford University Press, Oxford.

Popper, K.R., 1979. Objective Knowledge: An Evolutionary Approach, rev. ed. Clarendon Press, Oxford.

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2, 79–87.

Roth, G., 2015. Convergent evolution of complex brains and high intelligence. Philos. Trans. R. Soc. B 370, 20150049.

Schultz, W., 2016. Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. 17, 183–195.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423, 623–656.

Thelen, E., Smith, L.B., 1994. A Dynamic Systems Approach to the Development of Cognition and Action. MIT Press, Cambridge, MA.

Thompson, E., 2007. Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Belknap Press of Harvard University Press, Cambridge, MA.

van Duijn, M., 2017. Phylogenetic origins of biological cognition. Interface Focus 7, 20160158. doi:10.1098/rsfs.2016.0158.

van Gelder, T., 1998. The dynamical hypothesis in cognitive science. Behav. Brain Sci. 21, 615–628.

Varela, F.J., Thompson, E., Rosch, E., 1991. The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge, MA.

von Uexküll, J., 2010. A Foray into the Worlds of Animals and Humans: With a Theory of Meaning. University of Minnesota Press, Minneapolis. Original work published 1934.

Wheeler, M., 2005. Reconstructing the Cognitive World: The Next Step. MIT Press, Cambridge, MA.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1, 67–82. doi:10.1109/4235.585893.

Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. Science 269, 1880–1882.

Supplementary Material

What Predictive Brains Inherit: An Evolutionary Account of Cognition

Leonel Gómez-Sena — Sección Neurociencias, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

This document contains the full specification, evaluation design, and figures for the two-timescale model summarised in "The price of a test" in the main text, referenced there as Sections S1–S2 and Supplementary Figures S1–S2.

The complete implementation is archived at https://doi.org/10.5281/zenodo.22018805 (doi:10.5281/zenodo.22018805) and developed at https://github.com/leonelfgs/predictive-brains-inherit. It is a single Python file with no dependencies beyond NumPy and Matplotlib, runs in about one minute, and reproduces every number reported below exactly.

S1. Model specification

To test the internal coherence of the proposed mechanism, we implemented a minimal two-timescale model. The model is not offered as a novel learning architecture, nor as evidence that biological evolution followed this particular implementation. Its structure is familiar from the Baldwin-effect and meta-learning traditions: a slow selective process adjusts an inherited initial condition over a family of related tasks, whereas a faster process adapts that initial condition to the task encountered within an individual lifetime (Baldwin, 1896; Hinton and Nowlan, 1987). Its purpose is narrower: to determine whether competence acquired on a slower clock can be encapsulated as a prior for a faster one, and whether that transfer reduces the number of population-level selective evaluations required to attain a common level of competence.

The environment is a family of linear prediction tasks with stable statistical structure and variable lifetime-specific realisations. A fixed vector mu* defines the mean of the task family. At the beginning of each life, a task vector w is sampled around that mean. The agent then receives a sequence of signals x and must predict the corresponding output y.

```
w ~ N(mu*, sigma_task^2 I)

x_t ~ N(0, I)

y_t = w' x_t + e_t,   e_t ~ N(0, sigma_noise^2)
```

The slow clock is an evolutionary loop. Each inherited genome contains an initial parameter vector m and a learning-rate parameter log eta. With d = 8, the model therefore has d + 1 = 9 inherited parameters. The fast clock operates within a life. The operative parameter theta begins at the inherited initialisation m and is updated online from prediction error:

```
theta_0 = m

theta_(t+1) = theta_t - eta (theta_t' x_t - y_t) x_t
```

Fitness is the negative mean squared prediction error, averaged across tasks and observations within a life. Selection uses truncation and Gaussian mutation. The slow process is rewarded for initial conditions that perform well across the family of tasks, whereas the fast process can accommodate the component of each task that departs from mu*. This division is deliberate: the stable family structure is available to the slower clock, while the realisation-specific variation must be resolved within a life.

S2. Evaluation design and thresholds

The comparisons were designed to prevent advantages that arise from unequal data or parameter tuning. Every life is driven by a generator seeded from an explicit task seed, so paired conditions receive identical tasks, signals, and observation noise; the only difference between arms is the one under test. Learning rates were tuned separately for each condition by grid search over eight values from 0.0025 to 0.4, using seeds disjoint from the evaluation seeds. This matters because the learning rate coevolves with the prior: comparing an evolved prior against a zero initialisation at the same rate would impose an illegitimate handicap on the control. Paired bootstrap confidence intervals were used for the within-life advantage of the evolved prior, and bootstrap intervals for the 2x2 ablation, in both cases 4000 percentile resamples.

The champion of every evolutionary generation was evaluated on a fixed validation set that never drives selection; truncation selection operates on a separate per-generation task batch. The genome reported for each run is the one attaining the highest validation fitness across generations rather than the final generation's champion. This introduces a mild optimistic bias, applied identically to every condition, and so cannot manufacture a difference between arms.

Two thresholds are used, for different purposes. The first is the *competence criterion*, which fixes the level of performance a system must reach and is used in the ablation, selective-cost and multiple-implementation analyses. It was defined before the runs as removal of 70% of task-specific variance. If rho denotes the proportion of that variance removed, the criterion is MSE_criterion = sigma_noise^2 + (1 - rho) d sigma_task^2, with rho = 0.70. With sigma_noise = 0.3, sigma_task = 0.7, and d = 8, this gives MSE <= 1.266, or equivalently fitness >= -1.266. Because it is calculable in advance from the parameters of the world, it is independent of the observed evolutionary trajectories.

The second is the *within-life threshold*, used only when measuring how quickly a single learning trajectory converges on its task. It is set at twice the irreducible observation noise, MSE <= 2 sigma_noise^2 = 0.18, and marks the point at which within-life learning has essentially resolved the task at hand. References below to "the within-life criterion" denote this second threshold, not the competence criterion.

Table S1. Parameters of the completed run.

Parameter	Value	Role
Task dimension, d	8	Dimensionality of signals, tasks, and inherited prior
Task dispersion, sigma_task	0.7	Lifetime-specific variation around mu*

Parameter	Value	Role
Observation noise, sigma_noise	0.3	Irreducible within-life noise
Lifetime	40 steps	Observations available for within-life learning
Tasks per life	2	Independent task realisations per fitness evaluation
Population	30	Genomes evaluated per generation
Generations	50	Selective budget per replicate
Elite fraction	0.30	Fraction retained before mutation
Mutation SD	0.15	Gaussian mutation scale
Validation lives	8	Fixed independent validation set per generation
Selection batches	2	Task draws shared by the population each generation
Evolutionary replicates	6	Replicates in the selective-cost comparison
Runs per update rule	4	Independent runs for SGD and for normalized LMS
Learning-rate grid	8 values, 0.0025–0.4	Per-condition tuning by grid search
Tuning lives	40	Learning-rate tuning; seeds disjoint from evaluation
Evaluation lives	300 / 500	Reported statistics: encapsulation / ablation
Bootstrap resamples	4000	Percentile bootstrap
Competence criterion, rho	0.70	Proportion of task-specific variance removed

Supplementary Figure S1

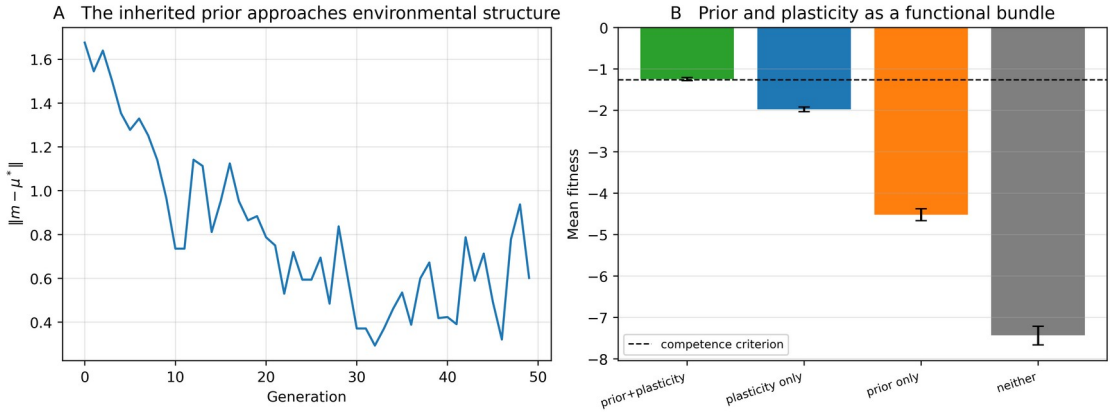

Figure S1. Encapsulation and functional complementarity in the two-timescale model. **(A)** Distance between the inherited initialisation m and the stable task-family mean mu* over evolutionary generations. The downward trend shows that selection installs shared environmental structure in the inherited starting point of lifetime learning. **(B)** Mean validation fitness in the 2x2 ablation of evolved prior and lifetime plasticity. Error bars are 95% bootstrap confidence intervals; the dashed line is the externally specified competence criterion (fitness >= -1.266). Higher, less negative fitness indicates lower prediction error. The combined condition is the only observed mean above criterion, while each ablated condition remains below it.

Supplementary Figure S2

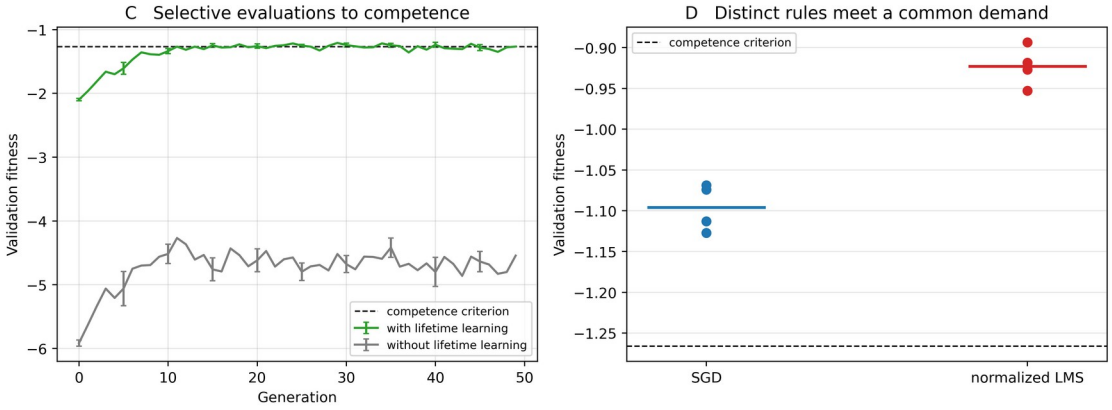

Figure S2. Selective cost and multiple implementations. **(C)** Mean validation fitness across six evolutionary replicates with and without lifetime learning. Error bars are standard errors of the mean and are displayed every five generations for legibility. The dashed line is the competence criterion. Every run with lifetime learning crossed criterion within 5–13 generations; no run without lifetime learning crossed it within the 50-generation budget. **(D)** Validation fitness for two distinct lifetime-update rules across four independent runs per rule. Points are individual runs, horizontal segments are group means, and the dashed line is the competence criterion. Both SGD and normalized LMS satisfy the functional demand despite

differing update equations and performance levels. Dispersions quoted in the text for this panel are population standard deviations over the four runs.

Note on nomenclature

"Normalized LMS" is retained with US spelling throughout, in both this document and the main text, because *normalized least mean squares* is the established name of the algorithm rather than an ordinary adjective. This is a deliberate exception to the manuscript's UK convention.

Supplementary references

Baldwin, J.M., 1896. A new factor in evolution. Am. Nat. 30, 441–451, 536–553.

Hinton, G.E., Nowlan, S.J., 1987. How learning can guide evolution. Complex Syst. 1, 495–502.