Article

Not peer-reviewed version

From Cancer to AI Alignment: Tackling Externalities Through Homeostatic Principles

Benjamin Lyons , Léo Pio-Lopez , Michael Levin *

Posted Date: 1 April 2026

doi: 10.20944/preprints202604.0056.v1

Keywords: diverse intelligence; cell biology; economics

 doi:10.20944/preprints202604.0056.v1

Article

From Cancer to AI Alignment: Tackling Externalities Through Homeostatic Principles

Running Title: Homeostasis of externalities

Benjamin Lyons [1], Léo Pio-Lopez [2] and Michael Levin [2,*]

[1] Independent Researcher, Denver, CO 80237, USA.

[2] Allen Discovery Center at Tufts University

* Correspondence: michael.levin@tufts.edu

Abstract

The problem of aligning humans and artificial intelligences can be understood in terms of minimizing externalities between them. However, economics cannot define externality because it contradicts the rationality assumption. This paper applies the homeostatic principles, from anatomical homeostasis to its disorder – cancer, to define externality. Drawing upon the perspective of cancer as a problem of scaling cellular collectives, this paper shows how to redefine both externality and rationality in terms of cognitive light cones (which demarcate the scale of goals any agent can pursue). We propose that cognitive light cones are constructed out of interoceptive signals for the purpose of anatomical homeostasis. We show that externalities can be understood in terms of anatomical homeostasis and derive some important implications for AI alignment, including the possibility of using market mechanisms enable the mutual co-construction of alignment between artificial intelligences and humans.

Keywords: diverse intelligence; cell biology; economics

1. Introduction: Misalignment and the Problem of Externality

The question of how to align artificial intelligence(s) to human beings is one of the most important ones of our time, and one that has received a flurry of interest in recent years (Ji et al., 2025). It is often defined as the problem of making AI behavior aligned with human intentions and values, and alignment is framed as achieving robustness, interpretability, controllability, and ethicality in these systems (Ji et al., 2025).

At a deeper level, alignment concerns not just how to preserve human life, but also how to make sure that humans remain able to make choices about their own futures (Bostrom, 2016). Put simply, the things that happen to people should be things that people choose to have happen to them. But how can such a concept be defined?

The field of economics offers a natural starting point. In economics, events that people do not choose to have happen to them, but which happen to them anyway, are called *externalities*. When one agent imposes costs on another agent without the latter's agreement, an externality is created. Therefore, eliminating externalities is the same as eliminating the possibility of an agent being unable to make choices about what happens to it.

Indeed, the alignment problem between *humans* is understood by economists to be the elimination of externalities. Misaligned incentives are observed in the externalities they produce. Humans, despite being members of the same species, are not necessarily naturally aligned with each other, as the historical record of abuses of power, crime, war, and other conflicts show. Alignment, therefore, is not uniquely a challenge of relating humans with artificial intelligences but can potentially be extended to different levels of intelligences, human or non-human, natural or engineered, individual, distributed or not, collective or collective of collectives etc. (Figure 1).

 doi:10.20944/preprints202604.0056.v1

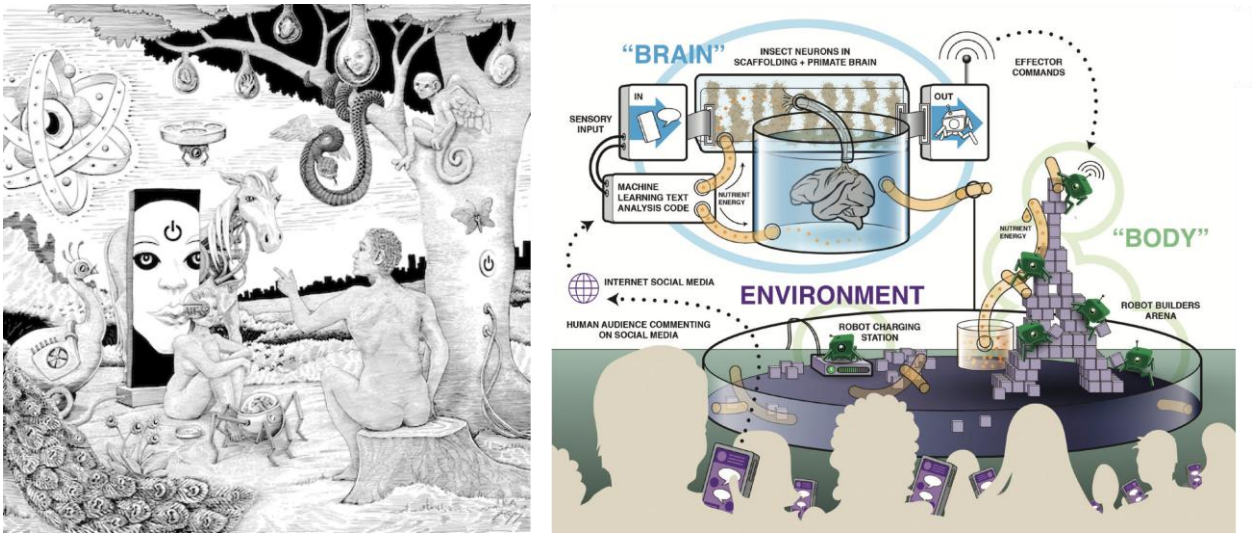

Figure 1. The spectrum of diverse intelligences. Left panel: a modern representation of the classic "Adam Names the animals in the Garden of Eden", showing the wide range of biological, synthetic, and hybrid beings with whom we will share our world. Right panel: an example of an unconventional embodied mind. It consists of a mammalian brain chimaerized with insect neurons. Its neuronal activity is detected by electronic interfaces and used to control a body in the physical world. The body consists of a robotic swarm acting in an arena, in which they can pick up nutrient packets and deliver those to the bioreactor to feed the brain. The collective needs to work together to power themselves and the central controller. Its environment is not only the passive arena, but also includes other sentient agents. The arena events are watched by an audience of human observers, who express their degree of approval of the behavior of the robots via real-time social media posts on their networked hand-held devices. An artificial intelligence (AI) language model scrapes the social media posts, converts the text into specific tokens and feeds it to the brain as input to its sensory neurons. Much like our cells, the robots also have a degree of their own on-board AI, and the behavior of the whole system is a very complex interplay of input, learning, noise, unreliability of components (including those of the observers), etc. This sample design illustrates how to move beyond our standard assumptions about what a functional brain, body and environment must look like, in order to illustrate the immense variety of different implementations of the central components of an environmentally embedded cognitive agent. Both images courtesy of Jeremy Guay of Peregrine Creative.

Associating misalignment with the presence of externality is not only intuitive but useful because externality is a subject that has been studied in economics since at least the 1890s (Boudreaux & Meiners, 2019), meaning that economists have over a century of tools, techniques, and wisdom to potentially bring to bear on the problem of alignment. However, there is also a challenge that needs to be overcome.

Surprisingly, economics has so far failed to define externality. Although many other concepts in economics are mathematically defined and able to be reasoned about rigorously, no comparable definition of externality has ever been produced (Mas-Colell et al., 1995). This difficulty was highlighted by Dahlman (1979), who argued that the problem remains unsolvable within the ordinary toolset of economics at that time. Since then, no tool to solve the problem has emerged.

We refer to this unresolved issue as the *problem of externality*. The central aim of this paper, is to contribute to progress on the problem of alignment by proposing a solution to the problem of externality.

To resolve the problem of externality, we will draw on ideas from outside of economics, particularly the physiological theory of cancer (Chernet and Levin, 2013; Levin, 2019; 2021b). This theory relies on the scale-free cognition framework (Moore et al., 2017), and states that cancer is first and foremost a disorder of patterning information (Rubin, 1985). Specifically, cancer results when cells disconnect from the collective, and thus lose access to the biophysical information structures that normally keep cells working towards constructing and maintaining specific large-scale

 doi:10.20944/preprints202604.0056.v1

anatomical features. Effectively, cancer cells revert to ancient single-cell goals at the expense of anatomical setpoints - downscaling goals and shrinking the computational boundary (or cognitive light cone) (Levin, 2019; 2021b).

We will show that externality can be thought of as a *generalized form of cancer* that can be applied to systems other than multicellular organisms. The central claim of this paper is that externalities arise when events relevant to a system's goals lie outside the system's cognitive light cone. Alignment problems are therefore failures to scale cognitive light cones across interacting agents.

Drawing on additional insights from neuroscience and developmental psychology, we argue that alignment is best understood as a problem of anatomical homeostasis, for a sufficiently generalized understanding of both anatomy and homeostasis that apply across substrates and goal spaces. We will show that this conceptual move has several interesting implications for alignment, including the potential for a general class of alignment technologies, the anatomical compiler.

The connection between the bioelectric theory of cancer and the alignment problem has been discussed in other papers (Bennett, 2025a; b; 2026). Efforts to use economic systems to study artificial intelligences (Tomašev, Franklin, Leibo et al., 2025) and to use economic mechanisms to safeguard humanity from artificial intelligences (Tomašev, Franklin, Jacobs et al., 2025) have been proposed. We expand on these ideas by showing how alignment problems in general can be understood in terms of cognitive light cones, explicitly fuse the economic and biological ideas into a single analysis, incorporate ideas from neuroscience and development psychology, and propose a general alignment solution concept, the alignment compiler, based on these ideas. Additionally, our analysis may seem to overlap with the idea of bounded rationality (Simon, 1955; Conlisk, 1996). However, bounded rationality refers to the agent's computational and cognitive limits, while our analysis refers to the limits of agency in general.

The remainder of this paper proceeds as follows. Section 2 describes the problem of externality, the inability of ordinary economic concepts to address it, and how it arises from the rationality assumption in economics. Section 3 introduces the concept of cognitive light cones and shows how they can be used to place boundaries on the limits of a system's agency. Section 4 presents a biological case study, showing how multicellularity and cancer can be understood in terms of alignment and misalignment, respectively that result as cognitive light cones scale up and shrink down. Section 5 shows the analogy to economics, outlining how the processes that produce multicellularity can be viewed as an example of economic phenomena. Section 6 applies this model to solving the problem of externality via the concept of cognitive light cones. Section 7 generalizes this argument across substrates and goal spaces, enabling a precise generalization beyond the example of multicellularity. Section 8 then details some important high-level implications for alignment between humans and artificial intelligences.

2. The Problem of Externality

We use the term "problem of externality" to refer to the inability of economics to define externality (Dahlman, 1979; Mas-Colell et al., 1995, ch. 11, sec. B, p. 351). Economics can provide vague verbal definitions of externality, but externality is not actually compatible with the rest of economic theory. Specifically, the presence of externality contradicts the *rationality assumption* upon which the rest of economics is based.

The problem of externality is readily illustrated with some basic terminology and a simple example. *Externality* refers to the effects of actions that are not connected to the price system (or anything that performs an analogous function). Often, externalities are called neighborhood effects. For example, suppose that Alice and Bob are neighbors, and Alice throws loud parties at night that keep Bob awake at night. Crucially, suppose that Alice does not compensate Bob for his troubles and has not otherwise transacted with him for the right to throw parties. Then the noise made by Alice's parties are supposed to constitute an externality because Bob has not chosen to be a recipient of the noise—the presence of the noise in his life does not reflect his own internal processes; it is imposed on him from an external source.

 doi:10.20944/preprints202604.0056.v1

However, if we assume that Alice and Bob are rational, then analyzing this hypothetical situation in any more detail forces us to determine that the supposed externality is in fact not an externality. To see this, suppose that Alice gets $10 of value from her parties, while Bob loses $20 of value from having his sleep ruined. Then if Alice and Bob are both rational, Bob will pay Alice some number between $10 and $20, such as $15, to stop partying. Both Alice and Bob will be $5 better off, making this a rational decision for both. Therefore, rational agents will bargain away the externality in this case.

Suppose alternatively that Alice gets $20 of value from her parties while Bob loses $10 of value from having his sleep ruined. Then Bob will choose not to pay Alice to stop her parties. In other words, Bob will choose to accept the noise from Alice's parties as part of his optimal consumption bundle. Therefore, an externality is not present.

The third case, in which Alice and Bob derive equal value from their respective activities, devolves into the first or second case depending on whether Bob chooses to pay Alice to stop partying.

This argument does not depend on concepts of transaction costs or property rights. Property rights affect transaction costs, but transaction costs are simply one type of cost among many and therefore cannot play a special role in defining externality (Dahlman, 1979).

To the best of our knowledge, the first argument demonstrating the problem of externality was made by Frank Knight (1924). Coase (1960) offered an expanded version of the same argument based on the concept of transaction costs (Coase, 1937). Dahlman (1979) showed the presence or absence of transaction costs is irrelevant to the presence of externality itself, a point generalized by Staten and Umbeck (1989) and Leeson (2020) to inefficiency in general.

Why does rationality rule out externality? The reason is that economics views a rational agent as always having a choice, however limited, about an event. Bob can pay Alice to reduce the noise, buy earplugs, build thicker walls, or move to a quieter neighborhood. If these options and all others are too costly for Bob relative to their benefits, then Bob rationally chooses to consume the noise as part of his optimal consumption bundle, making it not an externality.

The implication that a rational actor makes choices about everything that happens in their environment has no limits. The same argument shows that Bob makes choices about events on the other side of the world, and even in outer space. Bob "chooses" the orbit of Mars in the sense that he accepts it as part of his optimal consumption bundle—but they remain choices in a technical economic sense. The only limit on what Bob makes choices about is what physicists call a *light cone*, the range over which Bob's actions can causally influence events.

In short, economics lacks a *boundary condition* for agency. Since externality ought to refer to events beyond that boundary, economics is unable to define externality. To supply this missing boundary, we draw upon ideas from developmental biology, specifically the idea of a cognitive light cone (Levin, 2019). An agent's *cognitive* light cone, rather than its physical one, will prove to be the boundary on its agency that we desire to define externality.

3. Cognitive Light Cones as the Boundaries of Agency

An agent can be understood as a system that actively regulates some set of states toward preferred values. A human cleans their house. A bacterium regulates chemical gradients. A thermostat maintains room temperature. In each case, the system monitors variables and acts to reduce deviations from target states. In this minimal sense, agency corresponds to homeostatic regulation: the agent is the system whose activity reliably defends certain states against perturbation.

Because all agents share this general homeostatic property, what distinguishes one agent from another is the type of the state space it can navigate, the degree of competency with which it does so, and the scale of goal state it can effectively represent as the goal of its efforts. Different agents pursue preferred states in different goal spaces, and pursue goals of different sizes. For example, humans put a lot of effort into ensuring they have a daily calorie intake that falls into a certain range, while a chess-playing computer puts no effort into that. Also, when humans and chess engines play chess,

 doi:10.20944/preprints202604.0056.v1

the best chess engines are able to pursue their goals much farther than humans can into the branch trees of possible moves. Each player carries their goal forward into the future of the game tree only as far as their planning ability allows. Beyond that point, outcomes can no longer be reliably aligned with the player's goals.

So in order to capture the essence of what it means to be an active agent, regardless of scale, composition, or problem space, the field of Diverse Intelligence has the concept of a "cognitive light cone" (Levin, 2019): the size (in space, time, or whatever other parameters, such as the game tree of a chess position) of the largest goal an agent can actively pursue (Figure 2A). This is not meant to be the scale of reach of an agent's sensors or effectors, but the scale of the setpoints which its regulatory system can pursue. In this sense, the cognitive light cone also demarcates the boundary of the self: an agent can be defined, and distinguished from the environment in which it is embedded, by the size of states it actively manages. Whatever it is trying to control, is, in an important sense, part of the agent – the internal milieu (Turner 2002; 2016). The regions of the space at the expense of which it persists, which lie outside its radius of active concern, is the "outside world".

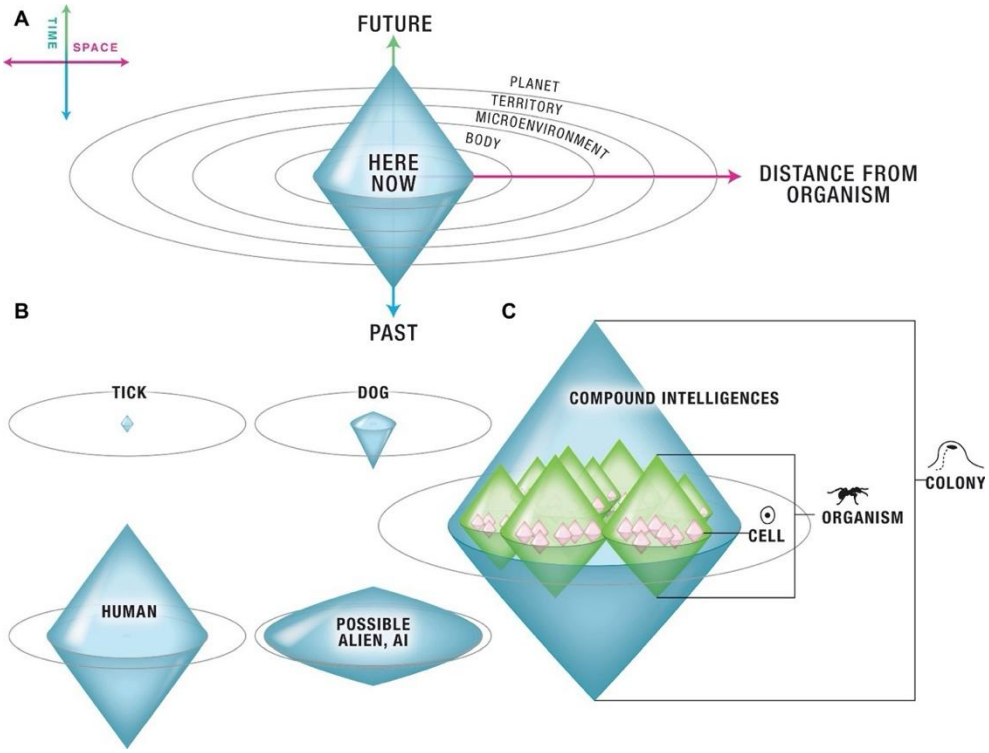

Figure 2. The cognitive light cone demarcates the size of an agent's goals (region of concern). A) The cognitive light cone is a region of space-time, representing the size of the goals an agent can have and how an agent's influence and information access extend from present into past and future along the time axis and outward with distance from the organism across nested scales (body, microenvironment, territory, planet). B) Examples of light-cones for different agents: a tick with a very small cone tightly bound to its immediate surroundings and present moment; a dog with a moderately expanded cone; a human with a larger cone spanning broader spatial ranges and longer time horizons; and a hypothesized alien or artificial intelligence with a highly extended spatial reach, with lower temporal one. C) Illustration of compound intelligences, in which multiple sub-agents or components (for example, cells within an organism, or organisms within an insect colony) each possess their own local light-cones that combine into a higher-level collective goal.

Individual cells, from unicellular organisms to body cells, have tiny cognitive light cones, consisting of local pH, hunger levels, etc. (Figure 2B). But embryos, regenerating limbs, and whole bodies (which need to resist degenerative aging and cancer for, in some cases, centuries) have massive cognitive light cones measured meters. Thus, evolution, and embryogenesis, achieve a remarkable

 doi:10.20944/preprints202604.0056.v1

up-scaling of cognitive light cones by achieving goals no single cells could do (Figure 2C). The size of the self literally expands during morphogenesis (see below for a discussion of the mechanisms of this expansion).

Goals are multicausally constructed (Thelen, 1989; 1995; Smith and Thelen, 2003), meaning that the environment and the task, as well as the organism or other entity play a role in constructing the goal and therefore play a part in scaling the cognitive light cone. Thus, the size of a system's cognitive light cone is not fixed but dynamically constructed.

A cognitive light cone is the boundary of an agent because the agent does not make meaningful choices about events not contained within it – it is the demarcation between an agent and the outside world, setting the boundary of things the agent functionally cares about (whether or not it has the metacognitive capacity to know that it has goals, or to adjust the target of care, etc.). As has been described in many aspects of biology, the boundary is not sharp but a gradient, implemented by activities ranging from communication (attempts to control others) to niche construction and superorganisms (Turner 2016, 2022). Nevertheless, it captures the key point that surviving agents first and foremost must exert control over their internal milieu; everything "outside" themselves is optional.

Importantly, while agents cannot make meaningful choices about the events outside their cognitive light cones, they can make meaningful choices about the size of their cognitive light cone. In game-theoretic terms, agents that acquire new information channels or coordination mechanisms may effectively change the payoff matrix they face, transforming previously competitive interactions into cooperative ones (Levin and Levin, 2021).

With this in mind, we can now propose a definition of externality in these terms. An *externality* is an event that is relevant to some agent's goals, but which lies outside the agent's cognitive light cone. Such events perturb the agent from its preferred states but are not incorporated into the agent's homeostatic processes. Accordingly, the rationality assumption in economics can be reinterpreted as the assumption that the agent's cognitive light cone is sufficiently expansive to incorporate all homeostatically relevant events into its homeostatic processes. (See below for an application of this idea to allostatic processes rather than homeostatic processes.)

Despite this definition, one must remember not to be careless when identifying externalities. The ability of the modern economy to scale a cognitive light cone is considerable. For example, no individual human can predict or control a severe weather event, but they can buy insurance to protect them from such an event, in which case the severe weather event would not constitute an externality.

So far, this discussion has been almost entirely theoretical. Having defined agency as bounded by cognitive light cones, we now turn to a biological system in which the expansion and contraction of these boundaries can be observed directly: multicellular development and its failure mode, cancer.

4. Multicellularity and Cancer: An Empirical Case Study of Cognitive Light Cones

To make the abstract framework of cognitive light cones concrete, we turn to a biological system in which the scaling of agency is directly observable: multicellularity, and the associated theory of cancer as a kind of dissociative disorder of the morphogenetic collective (Levin, 2019; 2021b).

A key underlying concept is that cells are not passive mechanical materials but a kind of agent, with sensory, effector, and computational capabilities that provide learning and flexible context-sensitive decision-making (Lyon, 2015; Baluška and Levin, 2016; Lyon et al., 2021; Baluška et al., 2022; Mathews et al., 2023; Chis-Ciure and Levin, 2025). Like other agents, cells regulate various aspects of their internal and external environments, pursue preferred states, and navigate around obstacles to do so. Their goals can be observed by the homeostatic setpoints toward which their activity converges.

Morphogenesis is a process in which cells (and the molecular networks within them) are harnessed towards a large-scale setpoint. For example, an axolotl limb can be amputated at any level, and the cells will rebuild the exactly correct structure and then stop – a process of anatomical

 doi:10.20944/preprints202604.0056.v1

homeostasis (Harris, 2018; Levin et al., 2019) (Figure 3A). Likewise, a tail can be surgically grafted to the mid-flank, and slowly remodels into a limb (Farinella-Ferruzza, 1956). It is clear that the cells are acting in concert to reduce distance from a very specific anatomical state – a pattern which is causal in the sense that it controls and explains why and when the cells act, what they do, and when they stop (Figure 3B). Living systems are comprised of a multi-scale competency architecture because biological material has homeostatic, allostatic, and other competencies at every level of organization– each scale has subunits which solve problems by navigating physiological, metabolic, transcriptional, and anatomical state spaces – not only the familiar 3D space of large organism-scale behavior (Fields and Levin, 2022) (Figure 3C-D). These subunits traverse various spaces in ways that often seek to minimize distance from specific states – their goals, as understood in cybernetics, control theory, and behavior science.

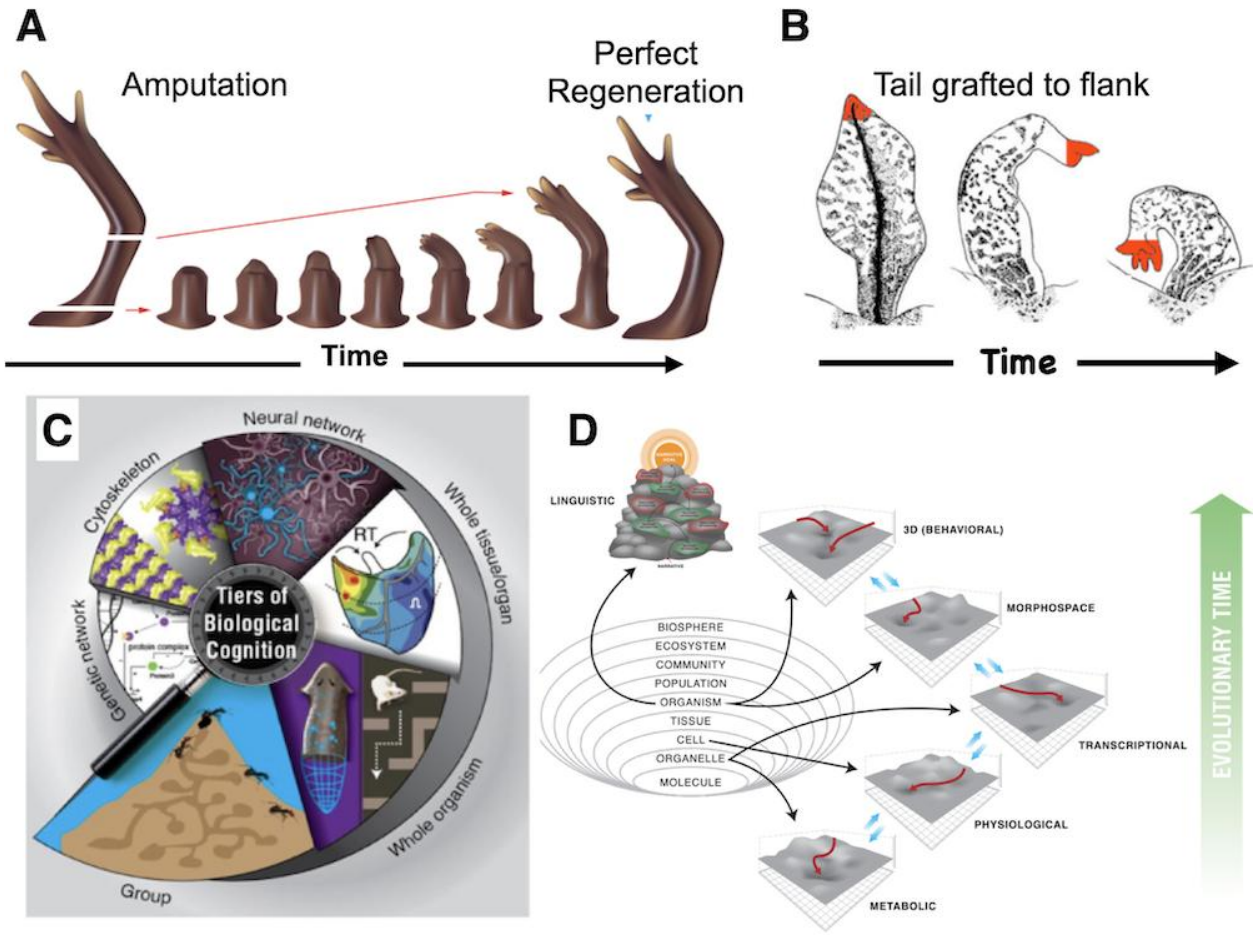

Figure 3. Biological goal-directedness across scales and problem spaces. (A) Axolotl limbs regenerate because if cell groups are deviated (by injury anywhere along the limb axis) from their normal anatomical setpoint, they will work to achieve it again and then stop when the right pattern is reached. (B) A tail grafted to the flank of an amphibian slowly remodels into a limb (Farinella-Ferruzza, 1956), including the tail tip cells (in red) which turn into fingers; since there is no local damage at the tip, it is seen that cells (and the molecular pathways inside them) are harnessed toward large-scale anatomical goals (see (Fields and Levin, 2022) for details). (C) Biological systems comprise a multi-scale competency architecture, where agendas and problem-solving policies at each scale bend the option space for the parts below, aligning them toward a high-level goal they may know nothing about. (D) Evolution pivoted some of the same basic strategies for navigating problem-spaces via goal-directed activities across metabolic, physiological, transcriptional (gene expression), and anatomical morphospace, in addition to the familiar 3D space of motile behavior and linguistic space (Fields and Levin, 2022). All panels courtesy of Jeremy Guay of Peregrine Creative.

 doi:10.20944/preprints202604.0056.v1

Individual cells are highly competent descendants of unicellular organisms; why would they work together to achieve grandiose goals such as specific organ morphologies? The answer is bioelectricity: bioelectric signals allow for trillions of cells to coordinate with each other, enabling the emergent production of a new kind of creature with goals and abilities that cannot be attributed to any of the individual cells. The most obvious example of this is the brain – we are not just a pile of neurons, and we know things (and have goals) that no individual neuron does, because bioelectrically-mediated signaling networks are an extremely convenient and powerful way to integrate information across space and time (Fields et al., 2020). But crucially, this system is not unique to the brain – it is ancient (Prindle et al., 2015; Liu et al., 2017), and was first discovered by evolution in microbial times, and was exploited to increase the cognitive light cone of individual bacteria to that of bacterial biofilms. It was used to make decisions about anatomy long before it was time-accelerated with the appearance of neurons, to make decisions about movement in 3D space. Bioelectric signaling allows individual cells to work together to produce a *collective intelligence*, a larger and distinct kind of causal nexus with its own abilities and goals (Watson and Levin, 2023; McMillen and Levin, 2024)(Levin and Martyniuk, 2018; Levin, 2019; 2021a; 2023a). Specifically, regulated electrical synapses known as gap junctions connect cells via a flexible communication channel that enables the network to host large-scale patterns of resting potential serving as setpoints for the entire tissue or organ (Levin, 2021a). This property is enabled by the multi-competency architecture and leads to the scaling of goals (Pio-Lopez et al., 2023).

But, like all mechanisms, the scaling of cells into multicellular organisms has a failure mode, rooted in the fundamental nature of cells as autonomous units only contingently bound to the large-scale goals of the collective. That failure mode is cancer.

While tightly-connected cell networks can store very large pattern memories as morphogenetic setpoints (which define their large cognitive light cones), individual cells can functionally detach from this network. This can be caused by oncogenes or persistent stress, which eventually causes a physiological disconnection (Trosko, 2005). Cells that physiologically disconnect can no longer access the large-scale setpoint and revert back to their ancient unicellular goals (Levin, 2021b), which are mostly around proliferation, migration, and metabolism. At that point, their cognitive light cone shrinks, and the boundary between themselves and the outside world becomes small again – as far as they are concerned, the rest of the body is just "exploitable outside world". They are not more selfish than normal cells, they just have smaller selves.

The theory of cancer as a disruption of the cellular self posits that cells always pursue their own cellular goals—they are self-interested, in economics terminology. Sometimes, the cell's Self is very small, and the cell only acts as if it cares about its own little cell-sized goals (Pio-Lopez et al., 2023) (see Figure 4). Cells do not intrinsically cooperate, and may compete with each other (Heams, 2012; Gogna et al., 2015). However, when connected to other cells through signaling, the cell behaves as if it is interested in goals far beyond itself, such as the assembly and maintenance of a large-scale morphological form (Levin, 2019). The large-scale (cell collective) uses signals which shape cell perception (Mitchell and Lim, 2016; Bugaj et al., 2017) in a way that bends their option space, causing their activity to advance large-scale setpoints (goals). Cancer is a failure of this collective process: separated from the bioelectrical signal (which can be due to hardware failure – genetic mutation – or to stress), a cell no longer behaves as if it cares about the greater whole but reverts to pursuing its own smaller cell-sized goals (Davies and Lineweaver, 2011; Bussey et al., 2017; Cisneros et al., 2017; Zhou et al., 2018) (Figure 4A-C), though there is also evidence of communication and cooperation among cancer cells (Egeblad et al., 2010; McMillen et al., 2021). Furthermore, the cell defaults to treating the rest of the multicellular organism as if it is the external environment: something to take resources from and to dump waste into, not something to care for and maintain (Moore et al., 2017). Cancer cells are thus not more selfish – they just have smaller selves about which they care. Strikingly, the development of tumors begins with cells becoming decoupled from the bioelectric coordination network of the larger organism (Aasen et al., 2003; Leithe et al., 2006; Kandouz and Batist, 2010).

 doi:10.20944/preprints202604.0056.v1

This competency arc—from unicellular to multicellular to organismal—positions cancer as an inevitable failure mode of collective behavior, this time from multicellular- to unicellular-scale setpoints (Levin, 2019; Lineweaver et al., 2021). Fortunately, it is amenable to biomedical interventions that restore cellular coordination, to re-align cellular agency with body-scale objectives (Pezzulo and Levin, 2015; Pio-Lopez et al., 2023; Shreesha and Levin, 2024; Pio-Lopez et al., 2025).

Cognitive glues: policies by which competent subunits become greater wholes

All intelligence is collective intelligence – made of parts of highly diverse levels of autonomy. The key thing about being a whole is coordinating its parts toward higher levels of information processing and causation (Rosen, 2012). This means that specific mechanisms and policies are required to align and integrate active subunits such that their local actions serve goals of which they are entirely unaware.

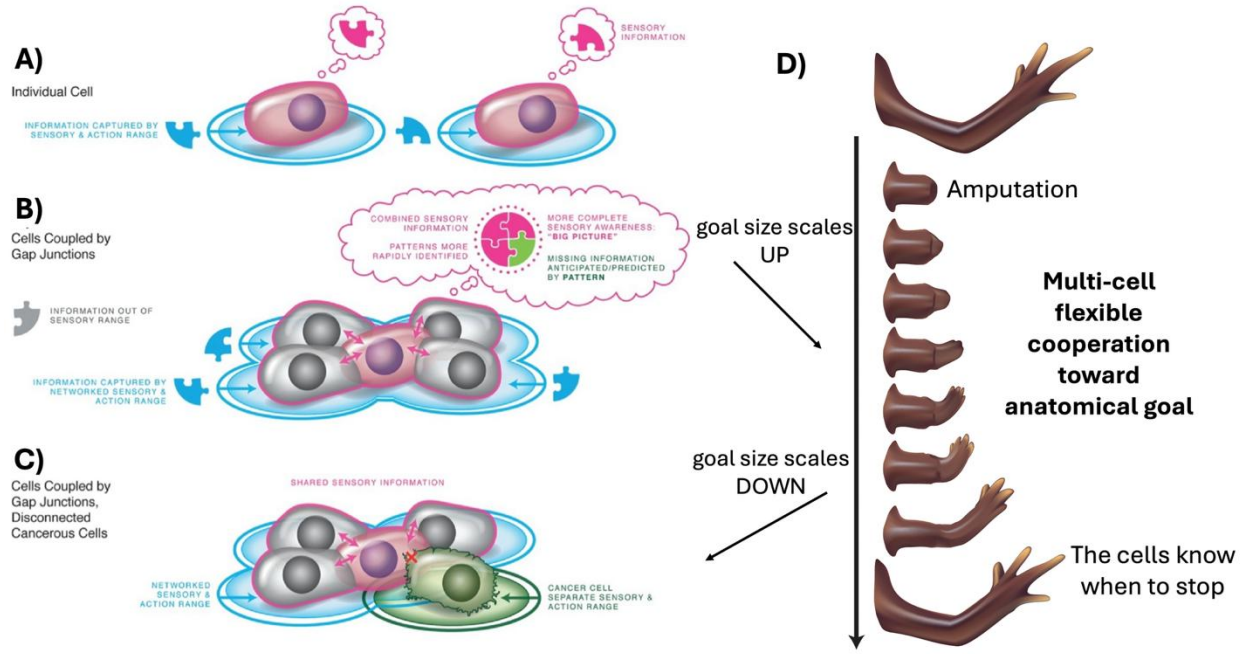

Figure 4. Multicellularity: integrated Individuals arise from sharing bioelectrical information. A) In isolation, a single cell can sense only its immediate microenvironment, and different cells are exposed to distinct local conditions. B) When cells become electrically coupled through gap junctions (electrical synapses), they gain access to a shared pool of information, including signals originating at the periphery of the collective. This integration allows the ensemble to monitor and regulate large-scale features of its environment within a homeostatic loop, thereby enabling the pursuit of spatio-temporally extended goals and giving rise to a higher-order Individual with a cognitive organization distinct from that of its cellular sub-components; the sensory space and set points are now scaled to the whole body. C) Physiological disintegration can occur locally, for instance through oncogene expression or loss of gap-junction communication via the introduction of physical barriers that can promote tumor development even without prior genetic alterations. In such conditions, the cells disengage from the higher-order Individual, shrinking their cognitive light cone, and shift back toward a unicellular-like behavioral regime characterized by uncontrolled proliferation and metastatic spread, prioritizing their own survival and expansion over the maintenance of anatomical homeostasis. **D)** Anatomical homeostasis is reached via the scaling of goals by aligning subcomponents via the cognitive glue. During regeneration the cells can reach the higher anatomical goal, here arm regeneration. But this process of the scaling of goals can go backwards, towards cancer.

Cancer then refers to cells that are disconnected from the organismal morphogenetic field via bioelectrical disconnection (Fields et al., 2020; Levin, 2021a).

 doi:10.20944/preprints202604.0056.v1

A crucial concept is the role that bioelectric signals play in regulating *proliferation*. Cancer cells are dangerous in part because they proliferate without regard for the effects their growth has on the rest of the body. In a healthy organism, cells are induced to proliferate in a particular way until a form is achieved, at which point they are induced to stop proliferating (Birnbaum and Alvarado, 2008). Bioelectric signaling is crucial to encoding the anatomical setpoint that tells the cells how much proliferation is enough but not too much (Durant et al., 2017; Levin, 2021a). Crucially though, this setpoint is a multicellular information pattern (Levin and Martyniuk, 2018), and individual cells not plugged in to the network that stores it, have no reason to disobey their ancient unicellular imperative for proliferation (Levin, 2019). Bioelectricity, in other words, plays a key instructional role, providing cells with the information and incentives they need to work for the collective as well as themselves (M. P. Harris, 2021). Without such instructional signals, cancer cells are so short-sighted that they will even participate in behaviors that ultimately lead to their own death.

The fact that failures of bioelectrical communication lead to cancer also predicts that medical interventions that support the bioelectric system may help to treat cancer. Bioelectric signaling can be used to normalize tumors (Levin, 2021b) and control tumorigenesis (Chernet et al., 2016) and metastasis (Payne et al., 2019). Bioelectric interventions have shown to be able to reduce tumor growth (Hitomi et al., 2015) and prevent/reverse tumor development (Chernet and Levin, 2013; Chernet et al., 2016). Drugs that target the channels used to transmit bioelectric signals may be used to treat cancer in the future, providing an alternative to toxic chemotherapy strategies via morphoceuticals (Kale et al., 2015; Tuszynski et al., 2017; Pio-Lopez and Levin, 2023).

This biological example empirically demonstrates three important principles. First, what is "external" to an agent's cognitive light cone is not about what is visibly external to them. Cancer, after all, takes place *inside* an organism. Instead, it is about what effects, components, or processes a system can effectively regulate. Second, the size of the cognitive light cone does not necessarily have an obvious connection to the size of the organism's sensorium. For example, being able to feel and see a tumor does not give one the ability to cure it. Third, phenomena that are external to an agent's cognitive light cone do not have to be imposed from outside but can emerge organically as a breakdown of the way that components and processes are integrated into the homeostatic regulation of the agent's set points.

We have previously described deep symmetries between the findings of cognitive neuroscience and those of developmental biology (Pezzulo and Levin, 2015). In the next section, we extend them further toward a new picture of externality in economics.

5. The Economic Analogy: The Price System as Cognitive Glue

Bioelectricity is not the only signaling system capable of coordinating vast numbers of agents into a higher-order collective intelligence. Another example of a system with these properties is the *price system* (Lyons and Levin, 2024). Just as bioelectric signaling integrates trillions of cells into the homeostatic processes that define and regulate a multicellular organism's morphological goals, prices integrate billions of decentralized human plans into the economy.

There are many parallels between these two types of collective systems. We outline the most important parallels here: stress sharing, memory anonymization, credit assignment, and the effects they have: control of form, scaling of goals, and, most critically, regulation of externality or cancer.

5.1. Stress Sharing

In both the economy and multicellularity, the subunits of the systems—humans and cells, respectively—are self-interested, not altruistic. Additionally, even *if* the subunits were altruistic, they would only have access to the problems they run into while trying to navigate to their own goals.

This means that the coordination of a collective intelligence such as an economy or a multicellular organism faces two problems: an *incentive* problem, in which subunits have no reason to adjust their behavior to accommodate each other, and a *knowledge* problem, in which subunits don't

 doi:10.20944/preprints202604.0056.v1

know anything about what is happening to other subunits even if they did want to accommodate them.

Both the biological multi-competency architecture and the economy solve the incentive problem and the knowledge problem by enabling *stress sharing* (and signal sharing more generally) among their subunits. *Stress* refers to the distance between a system and its goal. In morphogenesis, a cell's stress may refer to the distance it has from its target space in the three-dimensional configuration of cells, allowing error minimization. When a cell has high stress, it adjusts its behavior until stress is reduced. Consequently, when cells share stress signals with each other, each cell becomes responsive to the difficulties encountered by others, effectively transforming local problems into collective ones (Pio-Lopez et al., 2023; Shreesha and Levin, 2024).

Stress sharing solves the incentive problem because when one subunit receives the stress of another subunit, it is inclined to change its plan until the stress of the other subunit is resolved. In other words, no subunit is content until all the subunits are content—an economic equilibrium. Additionally, stress sharing solves the knowledge problem by enabling subunits to inform other subunits that they are encountering a problem. In addition, if stress corresponds to a higher goal, cells will coordinate to reach

The price system enables stress sharing in an intuitive and familiar way. Consider someone who discovers that a loaf of bread they expected to eat has become moldy. To resolve this problem, they purchase another loaf of bread, bidding up the price. Other buyers reduce how much bread they plan to purchase in response to the higher price, thereby making bread more available to be purchased by the original stress-sharer. Thus, the original consumer's problem is transmitted through the price system, making other people respond as if the problem were their own.

5.2. Memory Anonymization

Effective stress sharing requires that signals be interpreted as pertaining to the self rather than to another agent. We refer to this property as *memory anonymization* (Levin, 2022).

The biological multi-competency architecture and the price system both enable memory anonymization. A stress signal (either bioelectrical, biochemical or biomechanical), has no intrinsic meaning but is functionally interpreted based on the context (Levin, 2024; Barrett and Theriault, 2025). Bioelectric signals both store and share memories, such as pattern memories that contribute to homeostatic loops (Levin, 2022). Cells transfer signals pertaining to memories about their physiological state—e.g., how stressed they are—to each other, effectively allowing one cell to pass memories *into* another cell via signaling (Levin, 2021a). Consequently, the receiving cell cannot physically distinguish between its own memories and the received memories, thus causing it to treat the latter memories as if they are the former.

The price system enables memory anonymization as well. Participants in the bread market do not feel the buyer's hunger directly. Instead, they perceive their own stress from observing a higher price than expected, and act accordingly to buy less bread than they initially planned. Thus, the stress signal from the original hungry person is anonymized.

This has important characteristics, as stress corresponding to higher goals can there be passed to subunits, allowing the scaling of goals via multiscale homeostasis in collective intelligences.

5.3. Credit Assignment

Consider a rat trained to press a lever to receive a reward of food. The cells that press the lever are not the cells that receive the food, yet the collective cellular system—i.e., the rat—can be so trained regardless (Pigozzi, et al., 2025). This example indicates that *credit assignment* (Grefenstette, 1988; Yliniemi and Turner, 2016; Nguyen et al., 2018; Levin, 2023; Nguyen et al., 2024; Pignatelli et al., 2024) is crucial to the organization of a collective intelligence.

Just as cells can share stress with each other, they can also share memories of other physiological states, including positive ones, via bioelectric signaling too (Levin, 2023a). The price system also enables credit assignment in the obvious way: Suppliers of, e.g., bread who sell bread to a hungry

 doi:10.20944/preprints202604.0056.v1

person are rewarded via profit. The memory of reward is anonymized just like the memory of stress is because sellers of bread do not experience the satiety of the bread-buyer but instead experience the profit transferred to them by the satiated bread-buyer.

5.4. Control of Form

A cognitive glue offers control over the form of the system that it coordinates. Bioelectricity, for example, offers significant influence on the three-dimensional shape of an organism. Flatworms are typically one-headed, but by manipulating the bioelectric memory of the form to rewrite the anatomical pattern from one head to two heads, a two-headed flatworm can be produced without genetic editing (Levin, 2023a).

The price system also offers control over the form the economy takes—most literally, the three-dimensional structures that occur in it. For example, a window tax raising the price of windows led to the construction of houses with few windows (Oates and Schwab, 2015). Relative prices also determine what kinds of goods are likely to be seen in a person's pantry—e.g., more ground beef than filet mignon—as well as large-scale patterns of production and trade.

5.5. Scaling of Goals and Competencies

The most astonishing outcome of bioelectric and price coordination is the emergence of new systems with goals and competencies not attributable to any of the parts (Levin, 2022; Pio-Lopez et al., 2023; McMillen and Levin, 2024). Humans can do things that none of their individual cells can do, and economies can do things that no person can do on their own.

Scaling of goals and competencies is enabled on the subunit level by the division of labor. A human who lives on their own cannot plan to do anything more than find enough food and shelter to survive the night. A human who can rely on coordinating with billions of other people can form plans that extend years into the future and involve goals and skills that would never show up in nature, such as planning to become a sports announcer. Cells, similarly, cannot become parts of tissues or organs without being part of a coordinating signaling system that allows them to form plans as if they expect other cells to cooperate with their plans.

5.6. Generalizing from Cancer to Externality

With the analogy between bioelectricity and the price system established, another connection jumps out: the analogy between cancer and externality.

Cancer emerges when cells become disconnected from the larger organism's physiological signaling system. As individual cells, they are no more malicious or self-interested than any other cell, but their actions are not integrated into the larger homeostatic processes of the multicellular organism.

Externality is the economic analogue. When prices fail to transmit information relevant to the economy's goals, whether because the price system is incomplete, missing, or distorted in some way, individual economic agents pursue goals that are locally sensible but harmful to the collective. In the case of both cancer and externality, we observe a failure to scale individual plans into a system-level cognitive light cone. As a result, the size of the Self shrinks, and collective intelligence breaks down.

It should be noted that while economists regard externality as negatively as biologists regard cancer, we do not advocate for adopting biological mechanisms for merging memories between people and wiping their identities. The memory-anonymizing, stress-sharing mechanisms that the economy uses to eliminate externalities and scale up its cognitive light cone in fact enable a high degree of human individualism and independence. Moreover, the functioning of the economy depends on creative problem-solving and innovation that comes from mechanisms that maintain and enhance our ability to use our own individual competencies to fill gaps in the cognitive light cone of the economy—e.g., the economy depends on individual people being able to do things like notice that there is not a good pizza place in town and start one themselves. Optimal cognitive glue policies

 doi:10.20944/preprints202604.0056.v1

must be sought that gain collective intelligence and robustness while facilitating welfare and agency for the individual members.

6. A Solution to the Problem of Externality via Anatomical Homeostatic Principles

6.1. The Analogy Between Cancer and Externality

The structural parallels between cancer and externality can now be stated precisely:

1. *Cancer and externality are both problems of scaling cooperation and the "size of the Self"*. Cells or economic agents pursue local goals without incorporating the effects of their actions on the larger system. This occurs when connections through the system's cognitive glue, whether bioelectricity or prices, fail to transmit relevant information.
2. *Cancer and externality are both problems of collective behavior*. Cancer is a failure of multicellularity; it is a problem that only exists due to the challenges of getting self-interested cells to cooperate. Externality is a collective action problem; it is a problem that only exists due to the challenges of getting people to cooperate.
3. *Cancer and externality are both problems of excess proliferation*. Cancerous cells are cells that do not know "when to stop". Similarly, people and firms who produce externalities behave as if they do not know when to stop, either producing more than is socially optimal (called a negative externality) or else *insufficiently* proliferate, producing less than is socially optimal (called a positive externality).
4. *Cancer and externality are both problems of myopia (in space, time, or any goal space)*. Producers of externalities behave in short-sighted ways, producing effects that can ultimately lead to their own destruction, such as the harms from global warming or the degrading effects of crime on a social environment, etc.
5. *Cancer and externality are both solved by restoring connections via the system's cognitive glue*. Both problems are resolved by restoring connections to the system's cognitive glue. In economics, this typically means using taxes or subsidies to strengthen price signals, or institutional arrangements like clubs and firm integration that expand the size of the self and thereby internalize externalities.

The significant conceptual similarities between cancer and externality suggest that cancer may be usefully thought of as an example of externality, or externality may be thought of as a generalization of cancer. We can thus proceed to use the cancer-externality connection to propose a solution to the problem of externality.

6.2. Cognitive Light Cones as a Solution to the Problem of Externality

As demonstrated by Coase (1960), externalities are always eliminated long as economic agents can bargain over all relevant events. To define externality, therefore, we need some means of imposing limitations on what economic agents can make decisions about.

It turns out to be very difficult to discover and discuss such limitations within the conceptual toolkit of ordinary economics. The solution, as our analysis demonstrates, is to make use of the idea of a cognitive light cone to put bounds on agency. Events outside the cognitive light cone are unable to be regulated by the decision-making system. Thus, these events are true externalities: the recipient of an externality did not choose them. This definition preserves the insight of Coase (1960): externalities disappear when cognitive light cones expand to include all relevant effects.

Also preserved from Coase (1960) is the observation that externalities cannot be naively identified by pointing at events that do not seem to be incorporated into the system's methods of homeostatic regulation. This is because cognitive light cones for many agents, such as humans and the economy, can be scaled up and down *in different goal spaces*. For example, a human can choose to spend a lot of money on a security system that expands their cognitive light cone in the goal space of security, but by reducing their budget, they become less able to regulate events in other goal spaces. Thus, a system can make choices about which events it is going to attempt to incorporate into its

 doi:10.20944/preprints202604.0056.v1

cognitive light cone, and which it will bear the risk of being unable to regulate. There may also be tradeoffs in time, taking actions that reduce the size a cognitive light cone in the present in order to have a larger cognitive light cone in the future. Humans may even be able to make a concerted effort to increase the size of their cognitive light cones, such as the Bodhisattva Vow (Doctor et al., 2022).

The cognitive light cone definition of externality also allows us to advance the traditional analysis of externality in some ways. Recall that the inability to define externality is due to the rationality assumption in economics. With the understanding that an externality refers to events outside of the agent's cognitive light cone, the rationality assumption in economics seems to imply that the economic agent's cognitive light cone is unbounded. A perfectly rational agent, in this sense, would be able to perceive and respond to all consequences relevant to its goals. Under this assumption, externalities cannot arise by definition. It may be helpful, therefore, to think about new ways of understanding rationality through the lens of cognitive light cones.

Rationality in economics is defined as a preference order that is complete and transitive (Mas-Colell et al., 1995, ch. 1, sec. B, p. 6). With this in mind, completeness could refer to a cognitive light cone that has no "gaps" in it. For example, a human being who regulates many threatening substances in their environment but who cannot effectively make decisions about carbon monoxide has a carbon-monoxide-shaped gap in their cognitive light cone and is therefore incomplete. As expected, this "irrationality" then permits the presence of externality in the form of cognitive light cone. This gap can be filled in or covered up by new "organs" such as a carbon monoxide detector.

Transitivity might refer to some kind of non-interference rule: if incorporating one kind of event into the cognitive light cone makes it difficult to retain another kind of event, then that is a kind of inconsistency that might be worthy of calling intransitivity. For example, infants that are beginning to learn how to walk temporarily perform worse at object permanence (Thelen et al., 2001; Clearfield et al., 2006).

This last observation indicates a point worth reiterating: an agent or system with a cognitive light cone can make choices about *whether to incur the risk of increased exposures to externality*. A particularly salient example for the subject of this paper is the engineering of artificial intelligence. Artificial intelligence, integrated into the economy, will undoubtedly expand its cognitive light cone dramatically. At the same time, it exposes us to new risks of externalities—for example, plans formed by superior intelligences that exceed what humanity can predict and regulate. There is a difference, in short, between being *unable* to capture an event within a cognitive light cone and *choosing not to* capture an event within a cognitive light cone due to the tradeoffs involved.

Additionally, the notions of completeness and transitivity in the context of a cognitive light cone may help to define the "general" in artificial general intelligence, or AGI. An entity with a complete and transitive cognitive light cone is one that is able to regulate any and all events that are relevant to its goals, and it can regulate all of these events in a mutually consistent way with respect to its goals. Thus, its intelligence, or ability to navigate around obstacles toward its goals, is general across goals and obstacles.

Cancer, in this framework, is simply a case in which the activities of some cellular agents lie outside the organism's cognitive light cone and even actively hide themselves from it (Sultan et al., 2017; Galassi et al., 2024). Externality generalizes this phenomenon to collective intelligences and economic agents of all kinds.

Finally, our analysis has relied on anatomical homeostasis, or the maintenance of a fixed morphological setpoint. However, the economy has no single static form that it regulates toward. Instead, the economy's setpoints change as the conditions of supply and demand do. To better understand the implications this analysis has for economics and artificial intelligence, we propose a generalization of anatomical homeostasis that is more broadly applicable: anatomical *allostasis*.

7. Anatomical Allostasis

The analysis thus far has relied on the concept of anatomical homeostasis: the tendency of a biological system to reach and maintain a target form within a multi-competency architecture. This

 doi:10.20944/preprints202604.0056.v1

assumes a relatively stable set point, a preferred anatomical configuration toward which the system navigates even as the conditions of the internal and external environments change.

However, anatomical homeostasis is too narrow a concept to capture the full range of behavior exhibited by collective intelligences. People, organisms in general, and the economy itself pursue goals that change with time and circumstances. Analyzing this requires a framework with more flexibility than homeostasis allows.

Therefore, we propose a generalization of anatomical homeostasis: anatomical *allostasis*. Where anatomical homeostasis is the pursuit of a fixed setpoint, anatomical allostasis is the ability of a developing system, or an agent more generally, to update its setpoints in changing circumstances and navigate to them (Lagasse and Levin, 2023). More generally, allostasis refers to achieving stability through change, allowing flexible setpoints to better maintain stability (Sterling and Eyer, 1988; Sterling, 2004; 2012; Sterling and Laughlin, 2015; Barrett, 2017). Allostasis is theorized to be responsible for all behavioral and psychological phenomena (Barrett, 2017; Barrett and Lida, 2024; Barrett et al., 2025).

This generalization allows development, behavior, and cognition to be understood within a common framework. Each can be viewed as goal-directed navigation in a particular state space (Fields and Levin, 2022). What distinguishes them is not the underlying dynamics but the rate and flexibility with which setpoints change. Empirical work has highlighted shared dynamical principles across development, behavior, and cognition (Turvey, 1990; Thelen, 1995; Thelen and Smith, 1996; 2006; Spencer et al., 2001; Thelen et al., 2001; Levin, 2022 Smith and Thelen, 2003; Smith, 2005).

These processes depend on the sharing of signals across a network of components. In neuroscience, these signals are generalized under the concept of interoception, a process which conveys information about the physiological state of the body (Craig, 2002; Wiens, 2005; Ceunen et al., 2016; Quigley et al., 2021; Berntson & Khalsa, 2021). Regulating interoceptive signals enables allostasis, or the anticipatory coordination of the body (Barrett et al., 2016; Barrett 2017; Quigley et al., 2021; Theriault et al., 2021). In this sense, interoception provides the informational substrate through which allostatic control operates (Sennesh et al., 2022).

The economy exhibits analogous structure. The price system communicates information about the state of the economy, stabilizing the economy against shocks by changing setpoints in response to and anticipation of them. Thus, prices function as the economy's interoceptive signals, while market processes implement a form of allostasis that continuously updates patterns of production, exchange, and specialization as conditions change.

This perspective also clarifies how cognitive light cones relate to interoception: events outside the light cone are those that cannot be incorporated into the system's interoceptive signals. Because goals themselves are constructed through allostatic processes, the size of the largest goal a system can construct corresponds to the boundary beyond which it cannot effectively regulate events.

With the concept of anatomical allostasis in place, we can define externalities in a way that applies across systems, substrates, and goal spaces. We can now turn to the implications of this framework for alignment.

8. Alignment as a Problem of Externalities: Select Implications

8.1. Alignment is a Problem of Scaling a Cognitive Light Cone

Our analysis suggests that alignment can be thought of as a problem of scaling processes of coordination across new kinds and more instances of agents, and into new and scaled-up problem spaces, as opposed to instilling the "right" values into artificial intelligences. Rather than modifying the internal preferences of the agents in the system, alignment among those agents can be achieved by making sure that the effects of their actions are interoceptively accessible to the larger system the agents operate within. Relatedly, alignment is primarily an engineering problem in this view—a problem of building a system that can model and share various effects—rather than a philosophical problem of determining which values are right.

 doi:10.20944/preprints202604.0056.v1

Indeed, alignment has nothing to do with well the values of the agents in the system match up, if that is even a meaningful concept. Instead, alignment is about how those agents are coordinated via a cognitive glue, resulting in a scaling of cognitive light cones. This scaling occurs on two levels.

First, coordination via a cognitive glue scales up the cognitive light cone of each agent in the system so that the effects of the behaviors of every agent in the system is captured within each other agent's cognitive light cone. captured in some sense agreed to, and even chosen by, the other agents in the system. This means that no agent can impose its effects on other agents that the recipients of those effects did not choose.

Second, this same coordination process scales up the cognitive light cone of the collective system. When this system-level cognitive light cone captures the effects of all agents within it, those agents have necessarily not escaped the system's control but instead remain regulated by the system.

Failures of alignment are simply breakdowns of the scaling of cognitive light cones. For any number of reasons, the effects of the behavior of an agent may not be contained within the cognitive light cones of other agents or the cognitive light cone of the entire system. When this happens, the goals of the disconnected agent will be misaligned with the goals of the other agents, and of the collective system, in a precise sense.

The central implication of this paper for alignment, therefore, is that artificial intelligences, and indeed all processes and systems that we interact with, must in some way be incorporated into the cognitive glues, such as the price system, that we use to bind individuals into a society. In general, alignment is a system-level property. It is meaningless to say whether an individual agent is aligned in isolation.

8.2. Alignment is an Expansion of the Self

Multicellularity can be understood as the expansion of the Self (Levin, 2019). Independent, self-interested cells become part of a greater collective not by becoming altruistic but through the expansion of the size of their Selves, an expansion achieved through the scaling of cognitive light cones via the mechanisms discussed above.

As the Self of a self-interested agent scales up, so naturally do their interests. An isolated human living on a desert island may have little regard for anything but their own immediate needs for food and shelter. A modern human integrated into the worldwide economy, however, may participate in plans that end up supplying goods and services across the world to people whom they've never met.

If alignment is the task of getting agents to behave as if they follow the Golden Rule—do unto others as you would have them do unto you—then scaling up the size of the Self achieves this goal. Mechanisms such as stress sharing make it difficult for one agent to harm another because doing so would be like harming themselves. As a result, the decision to participate in a collective intelligence such as a multicellular organism or an economy don't simply offer costs and benefits in the traditional sense but also change the structure of the decision-maker.

As we described above, systems make tradeoffs about how large their cognitive light cones are and what goal spaces their cognitive light cones range over. A corollary of this observation is that agents face complex tradeoffs about the size of their Self and which Selves to be a part of (Shreesha et al., 2025). Agents can choose to join a collective or head off on their own. The challenge of alignment is in inducing decentralized agents to merge into and remain a larger Self.

8.3. Alignment is a Reciprocal Problem

As our analysis shows, misalignment is a function of the presence of externality in a system. As humans contemplating the emergence of artificial intelligence and increasing presence it will likely have in society, it is tempting to attribute the presence of any potential externalities to artificial intelligences, not to human beings.

Economics shows, however, that assumption is not tenable. Instead, externalities in economics are reciprocal (Coase, 1960), meaning that all involved parties can be considered responsible for the presence of the externality. To see this, consider again the case of Alice throwing noisy parties that

 doi:10.20944/preprints202604.0056.v1

keep her neighbor, Bob, awake. We could consider this as Alice imposing an externality upon Bob. However, suppose that we intervene for the sake of Bob and stop Alice from throwing her parties. Then we could just as easily say that Bob is imposing an externality on Alice! In truth, Alice and Bob's plans are simply opposed and so the fulfillment of either's plan could validly be said to be imposing an externality upon the other. (Reciprocal externality, therefore, is an example of multicausality.)

It follows that alignment is also a reciprocal problem. Misalignment between humans and artificial intelligences can be attributed to humans or to artificial intelligences. Both perspectives are valid. The "correct" one is whichever is most useful, i.e., whichever offers the most ability to improve alignment at the least cost. For example, if alignment can easily be solved by some minor tweak to how artificial intelligences are programmed, then misalignment should be attributed to them, but if misalignment is more effectively addressed by changing aspects of human institutions and norms, then it may be more useful to attribute misalignment to humans.

Ultimately, if artificial intelligences are not aligned to us, then we are not aligned to them. Alignment is a system-level property, so alignment problems will not be solved by focusing the analysis on one component of the system. All aspects of the system must be considered as variables than can be adjusted to produce alignment among the members of the system.

8.4. Alignment is a Problem of Allostatic Governance

Among humans, alignment problems are often solved by systems of governance. Governance includes explicit laws and government actions, but they also include institutional practices, ethics, and norms. Beyond such mechanisms, human interactions are also governed economic mechanisms, like how the price system enables people to form plans that they can confidently expect will be compatible with other people's plans without having to directly coordinate with anyone else or share the details of their plans.

The example of price system as a form of governance shows that allostatic processes play an important role in governing the members of a collective intelligence. As per the analysis of the reciprocal nature of alignment, we can attribute problems of alignment to any part of the system, including high-level governance processes such as the price system. Since these high-level governance processes play a major role in enabling billions of people to coordinate across miles and years, it may be useful to think of alignment as a problem of allostatic governance.

At the system level, allostasis, or the allocation of scarce resources throughout the system, serves to construct all target anatomical/behavioral/psychological states (Barrett, 2017; Barrett and Lida, 2024; Barrett et al., 2025) because the production of forms is determined by the intrinsic dynamics of the system, task, and environment in conjunction with internal resource allocation (Thelen et al., 1987). At the agent level, allostasis is responsible for the constraints and enablement that determine what choices are feasible. Agents are able to influence allostasis by sharing stress and other interoceptive signals through media such as bioelectric or price signals, such as by bidding up the price of a desired good. Allostatic governance is thus very much a government by consent in the sense that agents have the ability to "vote" on allostatic outcomes by interacting with the signaling system used to produce them.

In its ideal form, allostatic governance renders the plans of all agents in the system mutually compatible such that none have any reason to "vote" to change the system, and the behaviors of all agents align with each other. Failures of alignment can thus always potentially be traced back to failures of allostatic governance. Efforts to improve technologies and systems related to allostatic governance may ultimately prove to be the most effective large-scale interventions for achieving alignment.

8.5. Alignment Compilers: Using Markets to Achieve Anatomical Homeostasis in Arbitrary Goal Spaces

Our analysis suggests that alignment problems arise when agents act outside the signaling architecture that coordinates a collective intelligence. We now consider how new signaling mechanisms might be designed to extend that architecture.

 doi:10.20944/preprints202604.0056.v1

Broadly speaking, there are two ways to try to control a complex system. One option is to micromanage all of its components. However, this requires very complex modeling efforts and a tremendous amount of information and energy, due to the many components and their multiplicative degrees of freedom.

So, a better way, one stumbled upon by both nature and the economy, is to specify high-level goals for a system via a signaling system and allowing said system to induce the components to behave in ways that bring about the high-level goals. In multicellular organisms, large-scale anatomical outcomes are produced not by directly controlling the activity of all the trillions of cells in the system but by establishing signaling systems, such as bioelectric networks, that induce cells to coordinate their behavior toward a system-level anatomical goal. Similarly, the economy regulates the allocation of resources not by telling each of the billions of people on the planet what to do but by transmitting price signals that guide their decentralized decision-making so that when each person acts in a self-interested manner, the result is global coordination.

This observation suggests a more general principle: complex systems can often be steered toward desired outcomes by specifying high-level targets via a signaling system rather than by attempting to directly control the detailed behavior of each component. In developmental biology, this idea motivates the concept of the *anatomical compiler*, a technological goal that will translate desired anatomical outcomes into the patterns of bioelectric and biochemical signals that induce the body's cells to produce those outcomes (Lagasse and Levin, 2023). Rather than requiring detailed knowledge of every molecular process involved in anatomical homeostasis, the anatomical compiler would leverage the existing competencies of cells and tissues, identifying behavior-shaping cues (stimuli) that most effectively translate the morphological goals of the operator to those of the cellular collective. By adjusting the high-level signals that guide cellular behavior, the body's own anatomical homeostasis mechanisms would carry out the necessary molecular component-level processes.

The anatomical compiler would potentially enable the achievement of arbitrary outcomes in the goal space of anatomical homeostasis. What about other goal spaces? How can analogous technology be used to align the components of a system to other kinds of goals?

An answer to this question comes from economics. Consider the problem of a central bank trying to determine how achieve an intended rate of growth in Nominal Gross Domestic Product, or NGDP. This is a problem analogous to anatomical homeostasis: the central bank has some set point, say 5% growth in NGDP, and tries to achieve this set point. The central bank could try to solve this problem by forming a highly detailed model of the economy, but the economy is a highly dynamic and complex system, making this modeling task both very costly and unreliable.

An alternative approach is to rely on market mechanisms to specify the high-level goal and induce the participants in the market to construct the solution via their collectively coordinated competencies. Proposals for NGDP futures markets (Sumner, 1989; 2006) illustrate this idea. In such a market, the value of a futures contract rises if the market expects NGDP to come in above target and lowers if the market expects NGDP to come in below target. All the central bank has to do is adjust its activities until the futures contracts stabilizes at the desired level. Indeed, monetary policy could even be automated, with each long purchase of a contract triggering an action by the central bank that decreases the level of expected NGDP growth and each short sell for a contract triggering an action by the central bank that increases the level of expected NGDP growth. Trading will equilibrate when the central bank has been induced *by the market* to take actions such that NGDP growth will be as intended.

The NGDP futures market is analogous to the anatomical compiler. In each case, a technology is used to guide the components of a system toward a high-level outcome not by micromanaging the components but by arranging the signaling system to induce the components to use their own competencies to move the system toward that outcome.

This method could be applied to alignment as well—an *alignment* compiler. Instead of micromanaging the behavior of artificial intelligences or modeling their internal processes in

 doi:10.20944/preprints202604.0056.v1

exhaustive detail, market mechanisms can be used to shape the behavior of all members of society, both humans and artificial intelligences, toward collective outcomes.

The essential concept behind an alignment compiler is that signaling systems can be used to ask the components of a system, "What do you need us to do to get you to produce the outcomes we desire?" Markets are a particular useful mechanism for doing so because they induce the components to reveal their genuine beliefs (Hurwicz, 1960; Myerson, 1981; Maskin, 1999; Hanson, 2003; 2013; Wolfers and Zitzewitz, 2004; instead of hallucinating or behaving deceptively or strategically. Thus, market analogues to anatomical compilers could be used to induce an honest answer from artificial intelligences to the question, "What do you need us to do to get you to align your behavior with ours?" (This might be separated into many questions pertaining to desired outcomes.) Note that this method naturally treats the alignment problem as reciprocal: we induce honest answers from artificial an intelligence as to how to solve alignment by giving them means for inducing alignment-producing behaviors from us.

This method would use the artificial intelligence's own knowledge and capabilities to get it to honestly tell us how to align it with us. Alignment could even be automated as with an NGDP futures market by having each purchase or sell of a contract induce some compensating behavior to stabilize the price of the contracts.

Based on the analysis made in this paper, the potential of the alignment compiler concept lies in how it serves as a general concept for extending the cognitive light cone of the system that both artificial intelligences and humans participate in, both scaling its size and extending the signaling system that governs the collective into new goal spaces.

9. Conclusion

Economics has been studying alignment between human beings for 250 years. It is only natural to extend economic thinking to the problem of alignment between human beings and artificial intelligences. In general, economies operate via large-scale coordination mechanisms that enable many self-interested economic agents to coordinate as if they are all part of a larger Self, making economies generalizations of multicellular organisms.

In economics, the key misalignment concept is externality. If one party imposes an externality upon another party, the latter did not choose that and may be severely harmed. This means that externality does an excellent job of capturing the concept of misalignment between agents.

However, economics has historically been unable to define externality. The standard rationality assumption, wherein agents can form preferences over all relevant outcomes, makes it impossible to specify what it means for an event to lie outside of the decision-making ability of an agent. Our analysis clarifies why this is so: if rational agents are implicitly assumed to have unbounded cognitive light cones, then externalities disappear. Recognizing that real-world agents operate with bounded cognitive light cones restores the conceptual space in which externalities, and therefore alignment problems, can exist.

We have attempted to overcome this difficulty by incorporating ideas from biology, neuroscience, and psychology, most of all the bioelectric theory of cancer and the associated concept of a cognitive light cone. We propose that cognitive light cones are the boundaries beyond which an economic agent cannot organize their internal coordinating signals to achieve anatomical homeostasis (or allostasis). Events that are relevant to anatomical homeostasis (allostasis) but outside this boundary are externalities because they cannot be regulated by the current organization of the system's internal architecture.

Externalities are associated with misalignment because both concepts imply a breakdown of the integration of the agents of the system into a larger Self. Self-interested agents will take care of their Self, so misaligned agents are ones that do not behave as if they are part of the same Self. Externalities, in turn, are events that are relevant to the Self's goals but are not attributable to the decisions of the Self but instead seem to be imposed on the Self from some other source.

 doi:10.20944/preprints202604.0056.v1

A corollary of understanding alignment in terms of a shared Self is that alignment arises by giving artificial intelligences and humans a common goal. As described above, for a collective to reach anatomical homeostasis requires the components to share signals via a cognitive glue. The result is that all the members of the system share a model of the relevant parameters for decision-making (Lyons and Levin, 2024). With all agents sharing an internal model, they can consequently predict each other's behavior and minimize their mutual prediction errors, as suggested by the active inference framework (In other words, it would need both agents to share common internal models to communicate and understand each other as we can find it in the active inference framework too (Friston and Frith, 2015; Tison and Poirier, 2021). Alignment between humans and artificial intelligences may consequently lead to a change in the internal models that human beings use to navigate the world, leading to a new form of collective intelligence integrating natural and artificial intelligences and thereby broadening humanity's cognitive capacities, values, and interests.

Acknowledgements: We thank Michael Timothy Bennett and Emmett Shear for helpful comments on an earlier draft of this manuscript.

References

1. Aasen, T., Hodgins, M. B., Edward, M., & Graham, S. V. (2003). The relationship between connexins, gap junctions, tissue architecture and tumour invasion, as studied in a novel in vitro model of HPV-16-associated cervical cancer progression. *Oncogene*, *22*(39), 7969–7980. https://doi.org/10.1038/sj.onc.1206709
2. Baluška, F., & Levin, M. (2016). On Having No Head: Cognition throughout Biological Systems. *Frontiers in Psychology*, *7*, 902. https://doi.org/10.3389/fpsyg.2016.00902
3. Baluška, F., Reber, A. S., & Miller, W. B. (2022). Cellular sentience as the primary source of biological order and evolution. *Bio Systems*, *218*, 104694. https://doi.org/10.1016/j.biosystems.2022.104694
4. Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154
5. Barrett, L. F., Atzil, S., Bliss-Moreau, E., Chanes, L., Gendron, M., Hoemann, K., Katsumi, Y., Kleckner, I. R., Lindquist, K. A., Quigley, K. S., Satpute, A. B., Sennesh, E., Shaffer, C., Theriault, J. E., Tugade, M., & Westlin, C. (2025). The Theory of Constructed Emotion: More Than a Feeling. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *20*(3), 392–420. https://doi.org/10.1177/17456916251319045
6. Barrett, L. F., & Lida, T. (2024). Constructionist Theories of Emotions in Psychology and Neuroscience. In *Emotion Theory: The Routledge Comprehensive Guide*. Routledge.
7. Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160011. https://doi.org/10.1098/rstb.2016.0011
8. Barrett, L. F., & Theriault, J. (2025). What's Real? A Philosophy of Science for Social Psychology. In D. Gilbert, S. Fiske, E. Finkel, & W. Mendes (Eds.), *Handbook of Social Psychlogy 6th Edition* (6th ed.). Situational Press. https://doi.org/10.70400/BPQW3358
9. Bennett, M. T. (2025a). *How To Build Conscious Machines* (Wehmg_v1). Thesis Commons. https://doi.org/10.31237/osf.io/wehmg_v1
10. Bennett, M. T. (2025b). *Lies, Damned Lies, and the Orthogonality Thesis* (Zcfw6_v1). OSF Preprints. https://doi.org/10.31219/osf.io/zcfw6_v1
11. Bennett, M. T. (2026). *Are Biological Systems More Intelligent Than Artificial Intelligence?* (arXiv:2405.02325). arXiv. https://doi.org/10.48550/arXiv.2405.02325
12. Berntson, G. G., & Khalsa, S. S. (2021). Neural Circuits of Interoception. *Trends in Neurosciences*, *44*(1), 17–28. https://doi.org/10.1016/j.tins.2020.09.011
13. Birnbaum, K. D., & Alvarado, A. S. (2008). Slicing across Kingdoms: Regeneration in Plants and Animals. *Cell*, *132*(4), 697–710. https://doi.org/10.1016/j.cell.2008.01.040
14. Bostrom, N. (2016). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
15. Boudreaux, D. J., & Meiners, R. (2019). Externality: Origins and Classifications. *Natural Resources Journal*, *59*(1), 1–34.

 doi:10.20944/preprints202604.0056.v1

16. Bugaj, L. J., O'Donoghue, G. P., & Lim, W. A. (2017). Interrogating cellular perception and decision making with optogenetic tools. *The Journal of Cell Biology*, *216*(1), 25–28. https://doi.org/10.1083/jcb.201612094
17. Bussey, K. J., Cisneros, L. H., Lineweaver, C. H., & Davies, P. C. W. (2017). Ancestral gene regulatory networks drive cancer. *Proceedings of the National Academy of Sciences*, *114*(24), 6160–6162. https://doi.org/10.1073/pnas.1706990114
18. Ceunen, E., Vlaeyen, J. W. S., & Van Diest, I. (2016). On the Origin of Interoception. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00743
19. Chernet, B. T., Adams, D. S., Lobikin, M., & Levin, M. (2016). Use of genetically encoded, light-gated ion translocators to control tumorigenesis. *Oncotarget*, *7*(15), 19575–19588. https://doi.org/10.18632/oncotarget.8036
20. Chernet, B. T., & Levin, M. (2013). Transmembrane voltage potential is an essential cellular parameter for the detection and control of tumor development in a *Xenopus* model. *Disease Models & Mechanisms*, dmm.010835. https://doi.org/10.1242/dmm.010835
21. Chernet, B., & Levin, M. (2013). Endogenous Voltage Potentials and the Microenvironment: Bioelectric Signals that Reveal, Induce and Normalize Cancer. *Journal of Clinical & Experimental Oncology*, *Suppl 1*, S1-002. https://doi.org/10.4172/2324-9110.S1-002
22. Chernet, B. T., Adams, D. S., Lobikin, M., & Levin, M. (2016). Use of genetically encoded, light-gated ion translocators to control tumorigenesis. *Oncotarget*, *7*(15), 19575–19588. https://doi.org/10.18632/oncotarget.8036
23. Chis-Ciure, R., & Levin, M. (2025). Cognition all the way down 2.0: Neuroscience beyond neurons in the diverse intelligence era. *Synthese*, *206*(5), 257. https://doi.org/10.1007/s11229-025-05319-6
24. Cisneros, L., Bussey, K. J., Orr, A. J., Miočević, M., Lineweaver, C. H., & Davies, P. (2017). Ancient genes establish stress-induced mutation as a hallmark of cancer. *PLOS ONE*, *12*(4), e0176258. https://doi.org/10.1371/journal.pone.0176258
25. Clearfield, M. W., Diedrich, F. J., Smith, L. B., & Thelen, E. (2006). Young infants reach correctly in A-not-B tasks: On the development of stability and perseveration. *Infant Behavior and Development*, *29*(3), 435–444. https://doi.org/10.1016/j.infbeh.2006.03.001
26. Coase, R. H. (1937). The Nature of the Firm. *Economica*, *4*(16), 386–405. https://doi.org/10.1111/j.1468-0335.1937.tb00002.x
27. Coase, R. H. (1960). The Problem of Social Cost. *The Journal of Law & Economics*, *3*, 1–44.
28. Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature*, *34*(2), 669–700.
29. Craig, A. D. (Bud). (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655–666. https://doi.org/10.1038/nrn894
30. Dahlman, C. J. (1979). The Problem of Externality. *The Journal of Law and Economics*, *22*(1), 141–162. https://doi.org/10.1086/466936
31. Davies, P. C. W., & Lineweaver, C. H. (2011). Cancer tumors as Metazoa 1.0: Tapping genes of ancient ancestors. *Physical Biology*, *8*(1), 015001. https://doi.org/10.1088/1478-3975/8/1/015001
32. Doctor, T., Witkowski, O., Solomonova, E., Duane, B., & Levin, M. (2022). Biology, Buddhism, and AI: Care as the Driver of Intelligence. *Entropy*, *24*(5), 710. https://doi.org/10.3390/e24050710
33. Durant, F., Morokuma, J., Fields, C., Williams, K., Adams, D. S., & Levin, M. (2017). Long-Term, Stochastic Editing of Regenerative Anatomy via Targeting Endogenous Bioelectric Gradients. *Biophysical Journal*, *112*(10), 2231–2243. https://doi.org/10.1016/j.bpj.2017.04.011
34. Egeblad, M., Nakasone, E. S., & Werb, Z. (2010). Tumors as organs: Complex tissues that interface with the entire organism. *Developmental Cell*, *18*(6), 884–901. https://doi.org/10.1016/j.devcel.2010.05.012
35. Farinella-Ferruzza, N. (1956). The transformation of a tail into limb after xenoplastic transplantation. *Experientia*, *12*(8), 304–305. https://doi.org/10.1007/BF02159624
36. Fields, C., Bischof, J., & Levin, M. (2020). Morphological Coordination: A Common Ancestral Function Unifying Neural and Non-Neural Signaling. *Physiology*, *35*(1), 16–30. https://doi.org/10.1152/physiol.00027.2019
37. Fields, C., & Levin, M. (2022). Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments. *Entropy*, *24*(6), Article 6. https://doi.org/10.3390/e24060819

 doi:10.20944/preprints202604.0056.v1

38. Friston, K. J., & Frith, C. D. (2015). Active inference, communication and hermeneutics. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *68*, 129–143. https://doi.org/10.1016/j.cortex.2015.03.025
39. Ford, B. J. (2017). Cellular intelligence: Microphenomenology and the realities of being. *Progress in Biophysics and Molecular Biology*, *131*, 273–287. https://doi.org/10.1016/j.pbiomolbio.2017.08.012
40. Galassi, C., Chan, T. A., Vitale, I., & Galluzzi, L. (2024). The hallmarks of cancer immune evasion. *Cancer Cell*, *42*(11), 1825–1863. https://doi.org/10.1016/j.ccell.2024.09.010
41. Gogna, R., Shee, K., & Moreno, E. (2015). Cell Competition During Growth and Regeneration. *Annual Review of Genetics*, *49*(Volume 49, 2015), 697–718. https://doi.org/10.1146/annurev-genet-112414-055214
42. Grefenstette, J. J. (1988). Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms. *Machine Learning*, *3*(2), 225–245. https://doi.org/10.1023/A:1022614421909
43. Hansali, S., Pio-Lopez, L., Lapalme, J. V., & Levin, M. (2025). The Role of Bioelectrical Patterns in Regulative Morphogenesis: An Evolutionary Simulation and Validation in Planarian Regeneration. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, *11*(3), 305–331. https://doi.org/10.1109/TMBMC.2025.3575233
44. Hanson, R. (2003). Combinatorial Information Market Design. *Information Systems Frontiers*, *5*(1), 107–119. https://doi.org/10.1023/A:1022058209073
45. Hanson, R. (2013). Shall We Vote on Values, But Bet on Beliefs? *Journal of Political Philosophy*, *21*(2), 151–178. https://doi.org/10.1111/jopp.12008
46. Harris, A. K. (2018). The need for a concept of shape homeostasis. *Bio Systems*, *173*, 65–72. https://doi.org/10.1016/j.biosystems.2018.09.012
47. Harris, M. P. (2021). Bioelectric signaling as a unique regulator of development and regeneration. *Development (Cambridge, England)*, *148*(10), dev180794. https://doi.org/10.1242/dev.180794
48. Heams, T. (2012). Selection within organisms in the nineteenth century: Wilhelm Roux's complex legacy. *Progress in Biophysics and Molecular Biology*, *110*(1), 24–33. https://doi.org/10.1016/j.pbiomolbio.2012.04.004
49. Hitomi, M., Deleyrolle, L. P., Mulkearns-Hubert, E. E., Jarrar, A., Li, M., Sinyuk, M., Otvos, B., Brunet, S., Flavahan, W. A., Hubert, C. G., Goan, W., Hale, J. S., Alvarado, A. G., Zhang, A., Rohaus, M., Oli, M., Vedam-Mai, V., Fortin, J. M., Futch, H. S., … Lathia, J. D. (2015). Differential Connexin Function Enhances Self-Renewal in Glioblastoma. *Cell Reports*, *11*(7), 1031–1042. https://doi.org/10.1016/j.celrep.2015.04.021
50. Hurwicz, L. (1960). Optimality and informational efficiency in resource allocation processes. *Mathematical Methods in the Social Sciences*. https://cir.nii.ac.jp/crid/1571417125842505472
51. Ji, J., Qiu, T., Chen, B., Zhang, B., Lou, H., Wang, K., Duan, Y., He, Z., Vierling, L., Hong, D., Zhou, J., Zhang, Z., Zeng, F., Dai, J., Pan, X., Ng, K. Y., O'Gara, A., Xu, H., Tse, B., … Gao, W. (2025). *AI Alignment: A Comprehensive Survey* (No. arXiv:2310.19852). arXiv. https://doi.org/10.48550/arXiv.2310.19852
52. Kale, V. P., Amin, S. G., & Pandey, M. K. (2015). Targeting ion channels for cancer therapy by repurposing the approved drugs. *Biochimica Et Biophysica Acta*, *1848*(10 Pt B), 2747–2755. https://doi.org/10.1016/j.bbamem.2015.03.034
53. Kandouz, M., & Batist, G. (2010). Gap junctions and connexins as therapeutic targets in cancer. *Expert Opinion on Therapeutic Targets*, *14*(7), 681–692. https://doi.org/10.1517/14728222.2010.487866
54. Knight, F. H. (1924). Some Fallacies in the Interpretation of Social Cost. *The Quarterly Journal of Economics*, *38*(4), 582–606. https://doi.org/10.2307/1884592
55. Lagasse, E., & Levin, M. (2023). Future medicine: From molecular pathways to the collective intelligence of the body. *Trends in Molecular Medicine*, *29*(9), 687–710. https://doi.org/10.1016/j.molmed.2023.06.007
56. Leeson, P. T. (2020). Logic is a harsh mistress: Welfare economics for economists. *Journal of Institutional Economics*, *16*(2), 145–150. https://doi.org/10.1017/S1744137419000109
57. Leithe, E., Sirnes, S., Omori, Y., & Rivedal, E. (2006). Downregulation of gap junctions in cancer cells. *Critical Reviews in Oncogenesis*, *12*(3–4), 225–256. https://doi.org/10.1615/critrevoncog.v12.i3-4.30
58. Levin, M., Pietak, A. M., & Bischof, J. (2019). Planarian regeneration as a model of anatomical homeostasis: Recent progress in biophysical and computational approaches. *Seminars in Cell & Developmental Biology*, *87*, 125–144. https://doi.org/10.1016/j.semcdb.2018.04.003

 doi:10.20944/preprints202604.0056.v1

59. Levin, M. (2019). The Computational Boundary of a "Self": Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02688
60. Levin, M. (2021a). Bioelectric signaling: Reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell*, *184*(8), 1971–1989. https://doi.org/10.1016/j.cell.2021.02.034
61. Levin, M. (2021b). Bioelectrical approaches to cancer as a problem of the scaling of the cellular self. *Progress in Biophysics and Molecular Biology*, *165*, 102–113. https://doi.org/10.1016/j.pbiomolbio.2021.04.007
62. Levin, M. (2022). Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. *Frontiers in Systems Neuroscience*, *16*, 768201. https://doi.org/10.3389/fnsys.2022.768201
63. Levin, M. (2023a). Bioelectric networks: The cognitive glue enabling evolutionary scaling from physiology to mind. *Animal Cognition*, *26*(6), 1865–1891. https://doi.org/10.1007/s10071-023-01780-3
64. Levin, M. (2023b). Darwin's agential materials: Evolutionary implications of multiscale competency in developmental biology. *Cellular and Molecular Life Sciences: CMLS*, *80*(6), 142. https://doi.org/10.1007/s00018-023-04790-z
65. Levin, M. (2024). Self-Improvising Memory: A Perspective on Memories as Agential, Dynamically Reinterpreting Cognitive Glue. *Entropy*, *26*(6), Article 6. https://doi.org/10.3390/e26060481
66. Levin, S. P., & Levin, M. (2021). *Physarum polycephalum: Establishing an Assay for Testing Decision-making Under Shifting Somatic Boundaries* (p. 2021.10.17.464734). bioRxiv. https://doi.org/10.1101/2021.10.17.464734
67. Levin, M., & Martyniuk, C. J. (2018). The bioelectric code: An ancient computational medium for dynamic control of growth and form. *Biosystems*, *164*, 76–93. https://doi.org/10.1016/j.biosystems.2017.08.009
68. Lineweaver, C. H., Bussey, K. J., Blackburn, A. C., & Davies, P. C. W. (2021). Cancer progression as a sequence of atavistic reversions. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *43*(7), e2000305. https://doi.org/10.1002/bies.202000305
69. Liu, J., Martinez-Corral, R., Prindle, A., Lee, D.-Y. D., Larkin, J., Gabalda-Sagarra, M., Garcia-Ojalvo, J., & Süel, G. M. (2017). Coupling between distant biofilms and emergence of nutrient time-sharing. *Science*, *356*(6338), 638–642. https://doi.org/10.1126/science.aah4204
70. Lyon, P. (2015). The cognitive cell: Bacterial behavior reconsidered. *Frontiers in Microbiology*, *6*, 264. https://doi.org/10.3389/fmicb.2015.00264
71. Lyon, P., Keijzer, F., Arendt, D., & Levin, M. (2021). Reframing cognition: Getting down to biological basics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1820), 20190750. https://doi.org/10.1098/rstb.2019.0750
72. Lyons, B. F., & Levin, M. (2024). *Cognitive Glues Are Shared Models of Relative Scarcities: The Economics of Collective Intelligence*. OSF. https://doi.org/10.31219/osf.io/3fdya
73. Mas-Colell, A., Whinston, M. D., & Green, and J. R. (1995). *Microeconomic Theory*. Oxford University Press.
74. Maskin, E. (1999). Nash Equilibrium and Welfare Optimality. *The Review of Economic Studies*, *66*(1), 23–38.
75. Mathews, J., Chang, A. J., Devlin, L., & Levin, M. (2023). Cellular signaling pathways as plastic, proto-cognitive systems: Implications for biomedicine. *Patterns (New York, N.Y.)*, *4*(5), 100737. https://doi.org/10.1016/j.patter.2023.100737
76. McMillen, P., & Levin, M. (2024). Collective intelligence: A unifying concept for integrating biology across scales and substrates. *Communications Biology*, *7*(1), 1–17. https://doi.org/10.1038/s42003-024-06037-4
77. McMillen, P., Oudin, M. J., Levin, M., & Payne, S. L. (2021). Beyond Neurons: Long Distance Communication in Development and Cancer. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.739024
78. Mitchell, A., & Lim, W. (2016). Cellular perception and misperception: Internal models for decision-making shaped by evolutionary experience. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *38*(9), 845–849. https://doi.org/10.1002/bies.201600090
79. Moore, D., Walker, S. I., & Levin, M. (2017). Cancer as a disorder of patterning information: Computational and biophysical perspectives on the cancer problem. *Convergent Science Physical Oncology*, *3*(4), 043001. https://doi.org/10.1088/2057-1739/aa8548

 doi:10.20944/preprints202604.0056.v1

80. Myerson, R. B. (1981). Optimal Auction Design. *Mathematics of Operations Research*, *6*(1), 58–73. https://doi.org/10.1287/moor.6.1.58
81. Nguyen, D. T., Kumar, A., & Lau, H. C. (2018). Credit Assignment For Collective Multiagent RL With Global Rewards. *Advances in Neural Information Processing Systems*, *31*. https://papers.nips.cc/paper_files/paper/2018/hash/94bb077f18daa6620efa5cf6e6f178d2-Abstract.html
82. Nguyen, T. N., McDonald, C., & Gonzalez, C. (2024). Credit Assignment: Challenges and Opportunities in Developing Human-like Learning Agents. *Proceedings of the AAAI Symposium Series*, *3*(1), 54–57. https://doi.org/10.1609/aaaiss.v3i1.31180
83. Oates, W. E., & Schwab, R. M. (2015). The Window Tax: A Case Study in Excess Burden. *Journal of Economic Perspectives*, *29*(1), 163–180. https://doi.org/10.1257/jep.29.1.163
84. Payne, S. L., Levin, M., & Oudin, M. J. (2019). Bioelectric Control of Metastasis in Solid Tumors. *Bioelectricity*, *1*(3), 114–130. https://doi.org/10.1089/bioe.2019.0013
85. Pezzulo, G., & Levin, M. (2015). Re-membering the body: Applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. *Integrative Biology: Quantitative Biosciences from Nano to Macro*, *7*(12), 1487–1517. https://doi.org/10.1039/c5ib00221d
86. Pigozzi, F., Goldstein, A., & Levin, M. (2025). Associative conditioning in gene regulatory network models increases integrative causal emergence. *Communications Biology*, *8*(1), 1027. https://doi.org/10.1038/s42003-025-08411-2
87. Pignatelli, E., Ferret, J., Geist, M., Mesnard, T., Hasselt, H. van, Pietquin, O., & Toni, L. (2024). *A Survey of Temporal Credit Assignment in Deep Reinforcement Learning* (arXiv:2312.01072). arXiv. https://doi.org/10.48550/arXiv.2312.01072
88. Pio-Lopez, L., Bischof, J., LaPalme, J. V., & Levin, M. (2023). The scaling of goals from cellular to anatomical homeostasis: An evolutionary simulation, experiment and analysis. *Interface Focus*, *13*(3), 20220072. https://doi.org/10.1098/rsfs.2022.0072
89. Pio-Lopez, L., Hartl, B., & Levin, M. (2025). Aging as a Loss of Goal-Directedness: An Evolutionary Simulation and Analysis Unifying Regeneration with Anatomical Rejuvenation. *Advanced Science*, *12*(46), e09872. https://doi.org/10.1002/advs.202509872
90. Pio-Lopez, L., & Levin, M. (2023). Morphoceuticals: Perspectives for discovery of drugs targeting anatomical control mechanisms in regenerative medicine, cancer and aging. *Drug Discovery Today*, *28*(6), 103585. https://doi.org/10.1016/j.drudis.2023.103585
91. Pio-Lopez, L., & Levin, M. (2024). Aging as a loss of morphostatic information: A developmental bioelectricity perspective. *Ageing Research Reviews*, *97*, 102310. https://doi.org/10.1016/j.arr.2024.102310
92. Pio-Lopez, L., & Levin, M. (2025). Atavistic Genetic Expression Dissociation (AGED) During Aging: Meta-Phylostratigraphic Evidence of Cellular and Tissue-Level Phylogenetic Dissociation. *Aging Cell*, *25*(1), e70305. https://doi.org/10.1111/acel.70305
93. Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J., & Süel, G. M. (2015). Ion channels enable electrical communication in bacterial communities. *Nature*, *527*(7576), 59–63. https://doi.org/10.1038/nature15709
94. Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F., & Tsakiris, M. (2021a). Functions of Interoception: From Energy Regulation to Experience of the Self. *Trends in Neurosciences*, *44*(1), 29–38. https://doi.org/10.1016/j.tins.2020.09.008
95. Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F., & Tsakiris, M. (2021b). Functions of Interoception: From Energy Regulation to Experience of the Self. *Trends in Neurosciences*, *44*(1), 29–38. https://doi.org/10.1016/j.tins.2020.09.008
96. Rosen, R. (2012). *Anticipatory Systems: Philosophical, Mathematical, and Methodological Foundations* (Vol. 1). Springer. https://doi.org/10.1007/978-1-4614-1269-4
97. Rubin, H. (1985). Cancer as a dynamic developmental disorder. *Cancer Research*, *45*(7), 2935–2942.
98. Ruch, R. J., & Trosko, J. E. (2001). Gap-Junction Communication in Chemical Carcinogenesis. *Drug Metabolism Reviews*, *33*(1), 117–121. https://doi.org/10.1081/DMR-100000137
99. Sennesh, E., Theriault, J., Brooks, D., van de Meent, J.-W., Barrett, L. F., & Quigley, K. S. (2022). Interoception as modeling, allostasis as control. *Biological Psychology*, *167*, 108242. https://doi.org/10.1016/j.biopsycho.2021.108242

 doi:10.20944/preprints202604.0056.v1

100. Shreesha, L., & Levin, M. (2024). Stress sharing as cognitive glue for collective intelligences: A computational model of stress as a coordinator for morphogenesis. *Biochemical and Biophysical Research Communications*, *731*, 150396. https://doi.org/10.1016/j.bbrc.2024.150396
101. Shreesha, L., Pigozzi, F., Goldstein, A., & Levin, M. (2025). Extending Iterated, Spatialized Prisoner's Dilemma to Understand Multicellularity: Game Theory With Self-Scaling Players. *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, *11*(2), 135–151. https://doi.org/10.1109/TMBMC.2025.3562358
102. Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852
103. Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. *Developmental Review*, *25*(3), 278–298. https://doi.org/10.1016/j.dr.2005.11.001
104. Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343–348. https://doi.org/10.1016/s1364-6613(03)00156-6
105. Spencer, J. P., Smith, L. B., & Thelen, E. (2001). Tests of a Dynamic Systems Account of the A-not-B Error: The Influence of Prior Experience on the Spatial Memory Abilities of Two-Year-Olds. *Child Development*, *72*(5), 1327–1346. https://doi.org/10.1111/1467-8624.00351
106. Staten, M., & Umbeck, J. (1989). Economic Inefficiency: A Failure of Economists. *The Journal of Economic Education*, *20*(1), 57–72. https://doi.org/10.1080/00220485.1989.10844611
107. Sterling, P. (2004). Principles of Allostasis: Optimal Design, Predictive Regulation, Pathophysiology, and Rational Therapeutics. In *Allostasis, homeostasis, and the costs of physiological adaptation* (pp. 17–64). Cambridge University Press. https://doi.org/10.1017/CBO9781316257081.004
108. Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004
109. Sterling, P., & Eyer, J. (1988). Allostasis: A new paradigm to explain arousal pathology. In *Handbook of life stress, cognition and health* (pp. 629–649). John Wiley & Sons.
110. Sterling, P., & Laughlin, S. (2015). *Principles of Neural Design*. MIT Press.
111. Sultan, M., Coyle, K. M., Vidovic, D., Thomas, M. L., Gujar, S., & Marcato, P. (2017). Hide-and-seek: The interplay between cancer stem cells and the immune system. *Carcinogenesis*, *38*(2), 107–118. https://doi.org/10.1093/carcin/bgw115
112. Sumner, S. (1989). Uslng Futures Instrument Prices to Target Nominal Income. *Bulletin of Economic Research*, *41*(2), 157–162. https://doi.org/10.1111/j.1467-8586.1989.tb00287.x
113. Sumner, S. (2006). Let a Thousand Models Bloom: The Advantages of Making the FOMC a Truly "Open Market." *The B.E. Journal of Macroeconomics*, *6*(1), 1–27.
114. Thelen, E., Kelso, J. A. S., & Fogel, A. (1987). Self-organizing systems and infant motor development. *Developmental Review*, *7*(1), 39–65. https://doi.org/10.1016/0273-2297(87)90004-9
115. Thelen, E. (1989). Self-organization in developmental processes: Can systems approaches work? In *Systems and development* (pp. 77–117). Lawrence Erlbaum Associates, Inc.
116. Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist*, *50*(2), 79–95. https://doi.org/10.1037/0003-066X.50.2.79
117. Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *The Behavioral and Brain Sciences*, *24*(1), 1–34; discussion 34-86. https://doi.org/10.1017/s0140525x01003910
118. Thelen, E., & Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action* (Reprint edition). Bradford Books.
119. Thelen, E., & Smith, L. B. (2006). Dynamic Systems Theories. In *Handbook of child psychology: Theoretical models of human development, Vol. 1, 6th ed* (pp. 258–312). John Wiley & Sons, Inc.
120. Thelen, E., & Smith, L. B. (2006). Dynamic Systems Theories. In *Handbook of child psychology: Theoretical models of human development, Vol. 1, 6th ed* (pp. 258–312). John Wiley & Sons, Inc.
121. Theriault, J. E., Young, L., & Barrett, L. F. (2021). The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews*, *36*, 100–136. https://doi.org/10.1016/j.plrev.2020.01.004

 doi:10.20944/preprints202604.0056.v1

122. Tison, R., & Poirier, P. (2021). Active Inference and Cooperative Communication: An Ecological Alternative to the Alignment View. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.708780
123. Tomašev, N., Franklin, M., Jacobs, J., Krier, S., & Osindero, S. (2025). *Distributional AGI Safety* (No. arXiv:2512.16856). arXiv. https://doi.org/10.48550/arXiv.2512.16856
124. Tomašev, N., Franklin, M., Leibo, J. Z., Jacobs, J., Cunningham, W. A., Gabriel, I., & Osindero, S. (2025). *Virtual Agent Economies* (No. arXiv:2509.10147). arXiv. https://doi.org/10.48550/arXiv.2509.10147
125. Trosko, J. E. (2005). The role of stem cells and gap junctions as targets for cancer chemoprevention and chemotherapy. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *59 Suppl 2*, S326-331. https://doi.org/10.1016/s0753-3322(05)80065-4
126. Turner, J. S. (2002). *The Extended Organism: The Physiology of Animal-Built Structures*. Harvard University Press.
127. Turner, J. S. (2016). Semiotics of a Superorganism. *Biosemiotics*, *9*(1), 85–102. https://doi.org/10.1007/s12304-016-9256-5
128. Turvey, M. T. (1990). Coordination. *American Psychologist*, *45*(8), 938–953. https://doi.org/10.1037/0003-066X.45.8.938
129. Tuszynski, J., Tilli, T. M., & Levin, M. (2017). Ion Channel and Neurotransmitter Modulators as Electroceutical Approaches to the Control of Cancer. *Current Pharmaceutical Design*, *23*(32), 4827–4841. https://doi.org/10.2174/1381612823666170530105837
130. Watson, R. A., & Levin, M. (2023). The collective intelligence of evolution and development. *Collective Intelligence*, *2*(2), 26339137231168355. https://doi.org/10.1177/26339137231168355
131. Wiens, S. (2005). Interoception in emotional experience. *Current Opinion in Neurology*, *18*(4), 442–447. https://doi.org/10.1097/01.wco.0000168079.92106.99
132. Wolfers, J., & Zitzewitz, E. (2004). Prediction Markets. *The Journal of Economic Perspectives*, *18*(2), 107–126.
133. Yamasaki, H., Mesnil, M., Omori, Y., Mironov, N., & Krutovskikh, V. (1995). Intercellular communication and carcinogenesis. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *333*(1), 181–188. https://doi.org/10.1016/0027-5107(95)00144-1
134. Yliniemi, L., & Tumer, K. (2016). Multi-objective multiagent credit assignment in reinforcement learning and NSGA-II. *Soft Computing*, *20*(10), 3869–3887. https://doi.org/10.1007/s00500-016-2124-z
135. Zhou, J. X., Cisneros, L., Knijnenburg, T., Trachana, K., Davies, P., & Huang, S. (2018). Phylostratigraphic analysis of tumor and developmental transcriptomes reveals relationship between oncogenesis, phylogenesis and ontogenesis. *Convergent Science Physical Oncology*, *4*(2), 025002. https://doi.org/10.1088/2057-1739/aab1b0