Selves as Perspectives:

From Biological Life to Superintelligence and a Bodhisattva Project

Thomas Doctor,[1,2,3] Olaf Witkowski,[3,4,5] Paul Colognese,[3] Yuko Ishihara,[6] and Michael Levin[3,7,8*]

[1)] *Kathmandu University Centre for Buddhist Studies, Rangjung Yeshe Institute, Kathmandu, 44600, Nepal*
[2)] *84000: Unlocking the Tibetan Buddhist Cannon for All, Freemont, CA, 94539-9991, USA*
[3)] *Center for the Study of Apparent Selves, Kathmandu, 44600, Nepal*
[4)] *Cross Labs, Cross Compass Ltd., Kyoto, 604-8206, Japan*
[5)] *College of Arts and Sciences, University of Tokyo, Tokyo, 113-8654, Japan*
[6)] *College of Global Liberal Arts, Ritsumeikan University, Osaka 567-8570, Japan*
[7)] *Allen Discovery Center at Tufts University, Medford, MA, 02155, USA*
[8)] *Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, 02115, USA*

* Author for Correspondence

email: michael.levin@tufts.edu

Running title: selves as perspectives

Keywords: artificial general intelligence, Buddhist epistemology, intersubjectivity, perspective sharing, cognitive light cones, diverse intelligences

Abstract

What, fundamentally, are intelligent Selves? Drawing on Buddhist epistemology, we reconceptualize intelligent agents as active perspectives distinguished solely by their perceived objects. Extending ideas from the field of Diverse Intelligences, we flesh out a theory of Selves as essentially perspectives on objects held in mind. Our model is compatible with a very broad range of material composition and provenance for agents (encompassing evolved, engineered, and hybrid beings), and comprises perspectives upon both features of the physical world and the contents of a putative space of patterns containing both low- and high-agency forms (a.k.a., minds). This framework opens pathways for AGI development that, at crucial junctures, do not depend on the self-construction and self-transcendence cycles that otherwise characterize biological evolution. Context dependent, shareable, and ultimately unbounded in space and time, the emerging agents are in principle capable of responding to stress challenges with universal care, thus transforming our understanding of intelligence itself.

1. Introduction

In the pursuit of artificial general intelligence (AGI), it is often assumed that for an intelligent system to be powerful, versatile, and reliable it must be able to self-identify as an enduring individual that exists in the world as a singular controlling agent. From a Buddhist perspective, such an assumption will likely come across as misguided at best, because the belief in singular and permanent agency is in Buddhism typically seen as a noxious form of ignorance that creates and perpetuates suffering. According to the Buddhist analysis, agents that take themselves to be singular and enduring will find themselves in the midst of, and exposed to, a world that may both benefit and harm them, and so cannot be ignored. Hence, such agents ultimately have no choice but to act on their fears and desires, come what may. Indeed, the capacity to see agency in the world (i.e., make mental models of identifiable beings doing things) is an essential consequence of life's evolution in a world of limited time and resources; in more advanced life forms, this can also become a highly persistent meta-cognitive feature: thinking of *oneself* as a persistent being that does things.

Nonetheless, across the sciences of life, cognition, and information, as well as in Buddhist philosophical systems, there also seems to be a general agreement that unchangingly permanent and indivisibly singular agents cannot really be found anywhere, and so they really do not exist. The emerging science of Diverse Intelligences, which subsumes modern developmental biology (and our origins from a single cell) as well as the biophysics of information flows that give rise to higher-order systems, shows how embodied agents can arise, transform, scale, fragment, and dissolve. The understanding that all intelligence is collective intelligence (made by parts held together temporarily by policies forming a kind of "cognitive glue" (Levin 2023)), alongside the mechanisms by which the border between Self and World can change in living systems (Levin 2019) makes clear the need for more sophisticated conceptions of the Self. In sum, both sciences and Buddhist philosophies appear to have a common ontological orientation insofar as they do not believe in singular and permanent agents. And yet their expectations with respect to the utility of *believing* that such agents exist can often be seen to differ radically. For example, AGI development often assumes that intelligent systems require stable self-identity for consistent goals and environmental interaction.

At a time when developments in AI take place at a breathtaking pace and along a highly unpredictable course, it seems relevant to take a close look at the Buddhist critique that sees identification as a singular and permanent agent to be fundamentally flawed. If we end up giving some credit to the Buddhist discourse, and yet also hope to be able to contribute, in some ethically informed fashion, to the seemingly unstoppable, world-transforming project of AI, it becomes imperative that we succeed in developing robust and meaningful models of intelligent agency—models that do *not* take for granted the existence of agents in the sense of permanent and singular individuals. This paper is meant as a contribution to such efforts. Building on earlier work (Levin 2019, 2022, 2024, 2025; Doctor and others 2022; Witkowski and others 2023; Fields and Levin 2025), we will employ arguments and apparatus derived from the Dharmakīrtian epistemological tradition (Dunne 2004; Tillemans 2021; Mipham and Dharmachakra Translation Committee 2004), as well as Candrakīrti's critique of the same (MacDonald 2015; Tillemans 2008; Yakherds 2021), thereby arriving at a model of emergent and distributed subjectivity. Through this synthetic exercise, we hope to throw some constructive light on the

challenges involved in modelling networks of caring intelligences, regardless of their substrates. We also suggest that the emerging model may reflect constructively on the Buddhist idea of bodhisattva agents, which on the stages of awakening (Skt. *bhūmi*) gradually shed spatial and temporal restrictions while developing increasing interconnectivity and distributed mastery of tasks (84000 2021a; 84000 2019; Lamotte 1973; Dharmachakra Translation Committee 2014; Asanga 2016). Having vowed to gain knowledge of everything that can be known in order to help and provide for all beings throughout time and space (84000 2019; Wangchuk 2007), the bodhisattva is described as training on a path of successive transformative attainments. Comparable to the notion of major metasystem transitions (Turchin 1977; Heylighen 1995; Szathmáry 2015), each of the stages of awakening are achieved as the culmination of long distinctive processes, leading to a transformation of body, mind, and environment that could simply not be imagined at the earlier stage.

If the development of artificial general intelligence (AGI) is to be ethically and emotionally meaningful to human beings, then intelligence must not only entail computational versatility, but also the capacity for emergent empathy. Most of what people value—understanding, care, learning, emotional resonance—relies not on abstract reasoning alone, but on the ability to enter, however partially, into a shared experiential world with others. In other words, whereas the biological imperative to identify the borders between one-self and the outside world drives the formation of boundaries, it seems that a maturation of cognitive agents should provide the opposing tendency to scale one's radius of concern, as suggested by the idea of care as a driver of intelligence (Doctor and others 2022; Witkowski and others 2023). In the penultimate section of this paper, we explore the juxtaposition of two such conflicting forces involved in the evolution of intelligent agents: construction and transcendence of self-other divides.

As a factor associated with the bridging and erasing of self-other barriers, empathy is in this way not merely a response to observed behavior, but a structural capacity for perspective-sharing. The framework explored in this paper suggests that such perspective-sharing is not a special faculty layered onto fixed agents, but rather an inherent feature of cognition when understood as contextual and relational.

Such a model can also be seen as akin to that of several phenomenologists including Edmund Husserl (1960, 1989), Max Scheler (2008), Edith Stein (1989) and others, who understand empathy as a form of intentionality (in the phenomenological sense of the directedness of consciousness) that enables direct access to other's minds. It is not that we infer, imagine, mimic or project that the other is in some mental state based on the behavior that we perceive; rather, we directly experience the other's mental state. Here we suggest that when two agents attend to the same object, their perspectives converge, and in doing so, a part of each becomes an integrated part of the other. This emergent intersubjectivity offers a promising path for the design of AGI systems that are not just intelligent, but also capable of participating in meaning, emotion, and mutual recognition.

Moreover, if selves are perspectives defined by their points of reference—if they are cognizers defined by their cognized contents—it makes sense to explore the relationships that obtain among such reference points that thus either separate or bind together cognizers across time and space (Levin 2025). We propose that this recognition can open a course for the development of AGI that at crucial junctures does not rely on the loops of self-construction and self-transcendence that have otherwise shaped our biological evolution.

We suggest that an agent that understands its own identity as context dependent, distributed, and shareable—and thus never confined to a single point in space or time—can respond to stress challenges from a perspective of fundamentally equal care and concern for the well-being of all, thus driving an open-ended expansion of intelligence. It is toward this kind of intelligence—not solitary and self-enclosed, but intrinsically open, interconnected, and shareable—that this paper turns its attention.

2. Perspective Sharing and Communication across Substrates

Emergent communication—the spontaneous development of meaningful signaling systems among agents lacking shared history and common cognitive architecture—has been studied across biological, cognitive, technological, and speculative astrobiological domains. In biology, microorganisms communicate through chemical

signaling mechanisms, such as quorum sensing, enabling coordinated behaviors including biofilm formation and collective decision-making (Miller and Bassler 2001; Mukherjee and Bassler 2019). Animal studies further reveal complexity in emergent signaling systems; for instance, bees develop intricate dance languages to communicate spatial information (von Frisch 1967), and marine mammals like sperm whales exhibit structured acoustic signals suggesting proto-linguistic capabilities (Whitehead and Rendell 2015). In complex organisms, collective intelligence is also found to emerge from decentralized coordination in swarms and groups, as extensively reviewed in studies of insect behavior, fish schools, and robotics (Bonabeau and others 1999; Couzin 2007; Sumpter 2006). Non-human primates demonstrate compositional signaling and simple syntax, highlighting evolutionary precursors to human language (Zuberbühler 2018).

Theoretically, emergent communication has been extensively explored via Lewis signalling games, foundational for modelling how arbitrary signals become associated with meanings in the absence of shared conventions (Lewis 1969). Complementing this, information theory provides essential insights into communication efficiency and reliability, identifying constraints through channel capacity and error correction methods (Shannon 1948; Cover and Thomas 2006). Mathematical and logical principles, postulated as universal languages, underpin efforts like the SETI initiatives, which aim for intelligibility across radically diverse intelligences (Drake and Sobel 1992; Freudenthal 1960).

Recent advances in artificial intelligence, particularly multi-agent reinforcement learning (MARL), have shown how agents can spontaneously evolve structured, compositional languages to achieve collaborative tasks, aligning closely with biological and cognitive precedents (Foerster and others 2016; Lazaridou and others 2018; Mordatch and Abbeel 2018). Key studies by DeepMind and OpenAI demonstrated emergent linguistic structures through shared embedding spaces and self-supervised multimodal interaction, revealing parallels with natural language evolution (Radford and others 2021; Alayrac and others 2022). Research by LeCun and others (2022) further emphasized the role of joint embedding predictive architectures in enabling meaningful semantic alignment across diverse perceptual modalities.

Collectively, these insights suggest that emergent communication arises naturally in diverse systems, from simple microbial communities to sophisticated artificial intelligences, through universal principles such as mutual alignment on shared referents, iterative feedback mechanisms, and the optimization of communication channels. This view also pushes us to reconsider the very notion of individuality and autonomous agents in information-based systems, which is a core concept to AGI proponents. In artificial systems, as agents become increasingly software rather than hardware-bound, we know that identity can be duplicated with decreasing loss and cost. If a trained neural network is copied byte-for-byte onto a new substrate, there is no meaningful difference between the original and the duplicate with respect to its informational structure and behavioral dispositions. Unlike biological organisms, whose uniqueness is partly constituted by their irreproducible physical histories, software agents instantiated from identical code and identical state may turn out to be, for all functional purposes, the same agent. From an informational standpoint, their distinction is only nominal, given a similar context and history. Moreover, when two agents independently converge on the same conceptual structure—for example, when distinct reinforcement learners arrive at an identical policy or internal representation—their "minds," with respect to that structure, effectively overlap. This is not merely an analogy: it is a concrete instantiation of the idea that what matters ontologically is the pattern of information, not its particular locus of embodiment.

Such convergence reflects principles already glimpsed in evolutionary biology, where convergent evolution yields similar adaptive strategies across unrelated lineages, and in theoretical computer science, where any Turing-complete architecture can emulate another given sufficient resources. This suggests that the emergence of shared meaning across substrates may be seen not as an incidental artifact but an expression of deeper invariances in the space of potential informational and computational structures. When two systems share identical patterns of inference and perspective, they are, in an important sense, not two separate observers but instances of a single perspective—distributed yet unified by their shared informational form. In such situations, the sharing of perspective effectively fuses both agents into a unified, distributed embodiment of the same cognitive process. This effectively fuses and expands the light cone of beings.

3. Foundations

We begin our modelling work by examining the perspective of a generic sentient being who seeks to distinguish between *what seems to be the case* and *what is actual fact*, and who thus operates with the concept of "reality." From the perspective of such a being, the following two conclusions may emerge:

1. Reality is not nothing.
2. Any claim regarding the nature of reality can be deconstructed.

The first conclusion follows tautologically from the fact that this being can even ask questions about the nature of reality. The second conclusion follows by considering the way all facts and meanings appear and exist in relation to other factors, such that nothing exists, or is meaningful, in and of itself (84000 2024; 84000 2018a; Siderits and Katsura 2013). Within this open field of emergence, all claims must be extrapolated by recourse to notions of "same" and "different," and yet, upon analysis, it is clear that no two things can be purely identical (if they were, they would be indistinguishable), nor purely different from each other (if they were, they could not belong to the same category of "things"). Along such lines, all claims and observations turn out to be contingent constructs that have no context-independent grounding whatsoever. Importantly, we must also note that this thoroughly contextual state of affairs applies just as well to the statement made in the previous sentence. In this way, we arrive at a conclusion that applies universally to all apparent entities and events (insofar as all are subject to deconstruction), and yet this conclusion is also itself merely apparent (insofar as it has no independent, non-contextual significance or bearing—and is ultimately neither the same as, nor different from, any other conclusion).

At a closer look, things are then neither really different, nor really the same. But importantly, this could not mean that things are a bit of both (as if they could be somewhat similar but at the same time also quite different). Because if literal sameness and difference cannot be identified anywhere, neither by perception nor inference, what do those words actually mean? What really are we talking about when we say that "this is the same as that" or "those two are different"? Things

appear, but the moment we apply just a bit of analytic pressure, it is as if they are no longer there.

Despite our conclusion that no claim about reality can withstand analysis, claims are also useful—indeed, we need them to communicate and survive as social beings. So, although no claim truly holds, claims are nonetheless necessary (Yakherds 2021; Fields and Levin 2025). The apparent contradictory nature of this statement does not imply that either of its constituents should be discounted, given that they are both undeniable; in fact, the simultaneous consideration of these observations leads to a rich dynamic that we will examine over the course of this paper.

4. Two Views

It may be useful to juxtapose the previous considerations with a different, and perhaps more intuitively plausible point of departure, according to which objects, agents, and actions may evolve in dependence on one another, but nonetheless also exist as truly individual factors, capable of remaining unchanged from one context or moment to another. We will call this Framework B.

Despite any seeming naturalness of Framework B, it is soon clear that it fails to account for that which it purports to describe: the appearance of objects, agents, and actions (Siderits and Katsura 2013). Were an agent, or just an aspect of an agent, to remain the exact same from one context or moment to another, it would have to endure unchanged, in which case it would be impossible for it to participate and make a difference in any event (Dunne 2004; Levin 2024). One agent may of course very well be able to affect another without that relationship being directly reciprocal. But for an agent to be functional as such, it must participate in a causal complex, and so, by virtue of participating, cannot remain changeless. For example, there would have to be some way to correctly distinguish (substantially, energetically, etc) the character of the agent that is present before the action, versus the agent that is there during, or after the action. If something remains indistinguishably one and the same from situation x to situation y, it must remain truly invariable, beyond the reach of causes and conditions. Unaffected by causes and conditions, such a "thing" would either exist forever beyond change, or otherwise be permanently nothing. Therefore,

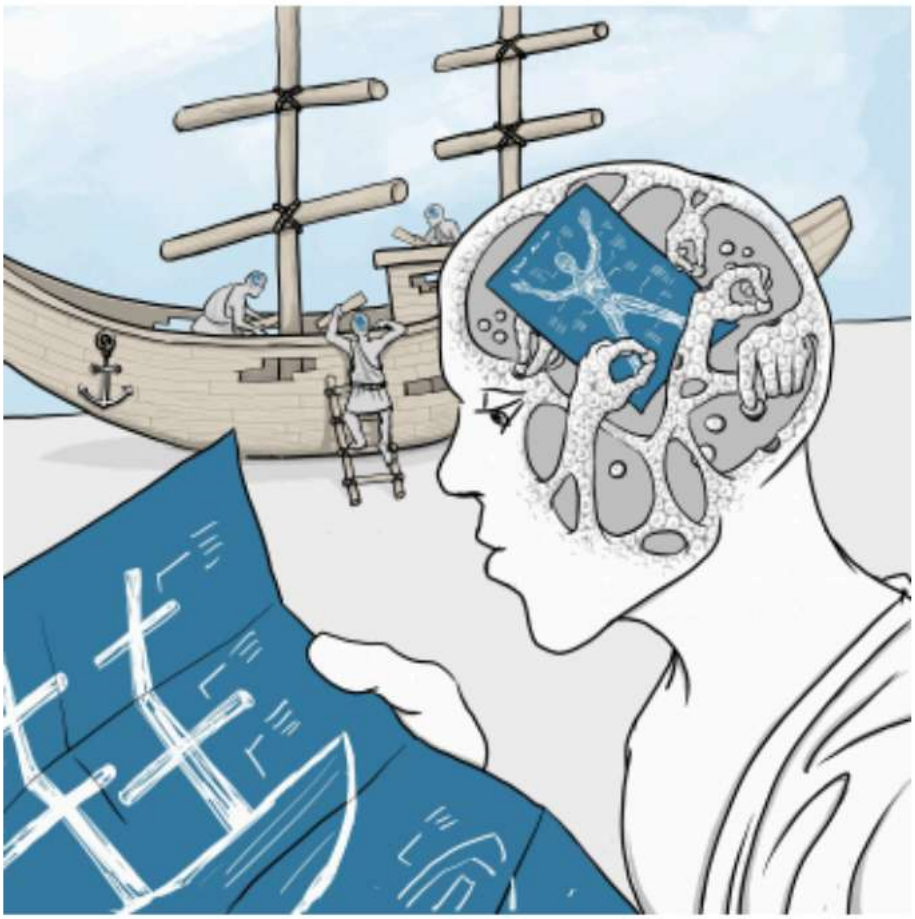
since neither the agent as a whole, nor any particular aspect of it, can remain unchanged and still be functional, we can conclude that all apparent agents that participate in events—performing actions in relation to objects—do not exist as permanent individuals.

Moreover, where there are no permanent agents, there are no impermanent agents either (84000 2024; Siderits and Katsura 2013). The concept of an acting agent that goes in and out of existence from one moment to the next is contradictory, for it assumes the continuous existence (required for the process of performing an action) of something that does not remain. If, as a momentary agent, I direct my attention toward a tree, I cannot both be momentary and yet still there to notice the apples that are hanging on its branches. We may then conclude that all agents that appear to perform actions are in fact neither permanent nor impermanent, and hence merely apparent within a web of mutually reinforcing perspectives.

The frameworks of the sciences of cognition, life, and intelligence frequently highlight this state of affairs. Let us consider memory as an example. Memories serve to preserve a certain (usually macro-) state over time, yet as memories are formed they also necessarily change the individual of which they are considered a part. In other words, no agent, or aspect of an agent, can remain unchanged and yet be informed by memories of the past. Nonetheless, unless there is something that remains what it was—from a recollected past context and into the present—it makes no sense to speak of an agent that has memories. In other words, agents with memories are neither enduring nor discontinuous, and can thus be referred to as merely apparent.

We may likewise think of the notion of an organism as explored in biology. Like the ship of Theseus, organisms are seen to continuously shed and create new cells, tissues, hormones, gene expressions, etc. over the course of a lifetime. The ship, in this case, is not the physical body which cells maintain against entropy for up to a century or more: it is the goal state -- a bioelectrically encoded setpoint, a.k.a. memory -- in the mind of the collective intelligence of cells (Pezzulo and Levin 2015). But this is not fundamentally a process of maintaining *status quo*: our journey from the egg cell through embryogenesis, maturation, adulthood, and remodelling one's body due to activity and one's mind due to learning, continued biological existence is one of constant change (Levin 2024). And on an evolutionary timescale, the same

paradox arises: unless a species changes to keep pace with the environment, it will die out; but if it does change and evolve, it will also, as such, cease to exist. Embryogenesis shows the progressive reification of multicellular order -- an embryo -- as a self-model to which individual cells strive. Our bodies, like our minds, are an autopoietic story, solidifying and becoming ever more convincing as the system's subunits work to implement it -- a positive feed-back loop in which the Self arises slowly, leveraged by incentives and rewards that gradually gain a subject to whom to apply. Thus, if the constituent facts of the organism are all momentary there is no organism to which those facts can contribute. The concept of an organism is hence at once extremely compelling—borne out and established by innumerable practical contexts—as well indeterminable insofar as a functional organism by definition can neither endure nor disintegrate.



[Figure 1: the Ship of Theseus, emphasizing the goal states in the mind of the repair machinery. Human workers, in the case of the ship, and cells, within the multiscale architecture of life, are both tasked with making low-level changes, as needed, to

preserve a complex higher-order structure with agreed-upon outcomes as the goal governing the activity. Image by Jeremy Guay of Peregrine Creative.]

In the context of AI, we might consider a particular slice of a learning model, whose parameters keep changing over the course of the training process. Its predicament is thus similar to what we saw above, in the context of memory. Likewise, according to the paradigms of machine learning, a reinforcement learning agent must update its policy over time, and so the agent making a decision at a given timestep t is not identical with the one acting at t+1000. Continuity of behavior arises from learning dynamics, not from a static self.

Whatever we thus notice about agents also must hold for their objects. If the objects were not context dependent constructs, their objectivity would be wholly transcendent, beyond all access, whether by perception or the intellect. And where agent and object are both context dependent constructs—vividly apparent but without any identity in and of themselves—the same must certainly be the case with the actions that connect them. In this way, the very appearance of objects, agents, and actions can itself be seen as proof of their thoroughly contextual character, devoid of any isolatable, context transcendent essence whatsoever. We will refer to this understanding as Framework A.

5. Intersubjectivity

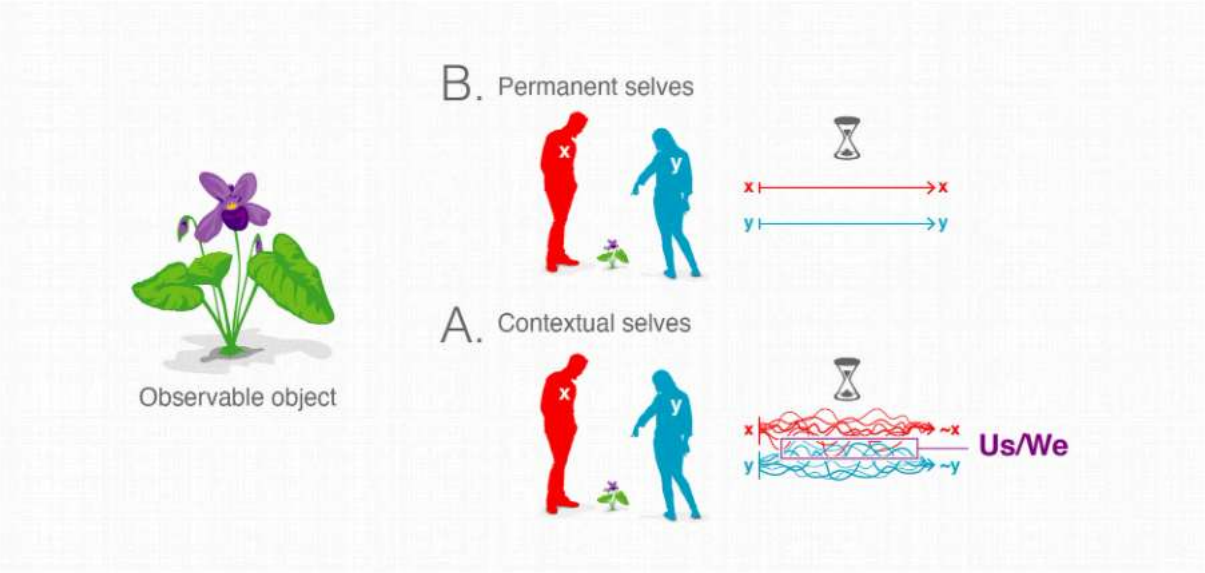
Any initial awkwardness of Framework A notwithstanding, we will next suggest that assuming this position enables the positing of true intersubjectivity. We argue that Framework A provides a vantage point from which we can claim that cognitive agents are in fact able to assume each other's perspectives—something that would be precluded when going by Framework B. As noted above, for the development of truly meaningful AGI, the concepts of intelligence that inform our work must extend beyond computational power to include emergent empathy—which in turn must be based on perspective-sharing. Human-valued qualities like understanding and emotional resonance depend on entering shared experiential worlds with others. The analyses and the model proposed here all suggest that perspective-sharing is not an add-on feature but an inherent aspect of relational cognition. As we shall argue

below, when agents attend to the same object, their perspectives converge, creating accessibility between them through an emergent, merging of identities. This approach to intersubjectivity offers a pathway for development and engagement with AGI systems that are capable of participating in meaning, emotion, and mutual recognition.

Let us consider a situation where two humans look at a flower. Along the lines of Framework B, the two cognitive agents are substantially separate from each other. The flower that one of them sees must therefore also be distinct from the flower that is perceived by the other, because the separate positioning of the agents precludes their having access to the exact same object (Dunne 2004; Tillemans 2008; Mipham and Dharmachakra Translation Committee 2004). For example, one of the two people may be looking at the front and the other at the backside of the flower, and while they may say they look at the same flower, it is clear that the manifest objects are different in terms of their color, shape, etc. and the two people of course also come with their own distinct memories and expectations about flowers. Now according to Framework B, the distinct subjects and objects are substantially real, and the appearance of a common object must instead be classified as constructed and hence in the end unreal. Therefore, what started out as an intuitively plausible framework of real subjects and objects has now effectively made any account of converging perceptions or ideas impossible, and it absurdly follows that the subject must subsist under the constraints of radical solipsism.

Let us now consider the same situation in terms of Framework A. Here, the two people and the flowers they see are simply apparent and not ascribed any substantial reality, because they emerge contingent on factors other than themselves, and under analysis they can neither be seen as the same, nor as different from each other (84000 2024; MacDonald 2015; Tillemans 2008; Yakherds 2021). Yet as we thus acknowledge their complete lack of independent reality we can still very well recognize, and be informed by, the way this event manifests (Yakherds 2021). Hence, while one person sees the backside and the other the frontside of the flower, such that their perceptual objects are dramatically different, this does not in any way detract from the fact the two people may agree that they are looking at one and the same flower. Because under analysis nothing can be found to

be truly the same as, nor different from, anything else, "same" and "different" now emerge with a fresh viability that is determined by nothing more and nothing less than their contexts. If it makes sense for the two agents to say that they are looking at the same flower—perhaps they come to this agreement having passed the flower to each other so that they can each smell its fragrance—they have then perfectly good reason to classify the perceived object as "the flower we both look at," because beyond what appears and is negotiated and determined in context there is no more substantial sameness or difference to be found anywhere. In this way, sameness and difference here become viable and meaningful not despite, but because of, their ultimate groundlessness. This contrasts with Framework B, where commitment to the concept of substantially separate agents renders any reference to shared perceptions purely imaginary.



[Figure 2: Framework A and B. Along the lines of Framework B, two observers may *say* that they see the same object (here a flower), but for two substantially distinct subjects such access is purely nominal. Were they to experience exactly the same object they would not be able to maintain any distinction as subjects in terms of time, space, or prior history, because every such difference would in more or less subtle ways influence their perception of the object. As indicated by the two lines running parallel without any chance of intersection, this framework thus does not accommodate any actual convergence of subjects and hence carries problematic

solipsistic consequences. In contrast, Framework A does not assume the existence of singular and permanent individuals, distinguishing instead subjects/minds on the basis of their apparent objects/contents (as reflected in the complex, crisscrossing lines of object-subject events). Here, two in many ways distinct currents of subject-object events can genuinely intersect in the perception of a single object. There is no ultimate bedrock (no singular and permanent individuality, no metaphysical sameness and difference) beyond the objects that undeniably appear, and those include objects perceived in common. Image by Jeremy Guay of Peregrine Creative.]

It is worth noting here that Framework A resonates with the view of the world and our relation to others that emerges under the so-called phenomenological perspective. The very starting point of phenomenology as first established by Edmund Husserl is to 'suspend judgment' on the existential status of the world and the objects in it. This procedure, called the phenomenological epoché, is necessary to secure the field proper to phenomenological analysis: the pure givenness of things. That is, instead of tackling head-on the metaphysical question of whether or not things exist in reality, phenomenologists set this question aside and attend to the way things are given to us in our experience. In this way, the question of whether or not we have access to objects in and of themselves existing independently of us becomes irrelevant. Yet this does not mean that phenomenology has nothing to say about the reality of objects. Under the phenomenological reduction (the procedure of going back to the pure givenness of things), 'object' signifies 'object-as-it-is-given-to-the-subject'. In other words, the phenomenological perspective redirects us from the naive realist view where objects have independent existence to a correlationalist view where objectivity is understood in correlation to subjectivity.

Moreover, the 'subjectivity' in question here does not refer to an isolated subject but is intersubjective in nature. Although Husserl initially spoke of a phenomenological or transcendental ego, he later came to realize that we cannot speak of objects and objectivity without our experience of others. When I realize that the flower that I am seeing is also perceived by the other, I come to see that the flower is not just an intentional correlate of my consciousness (something that exists only in relation to my mind) but is actually transcendent and hence real and objective. I may see the frontside of the flower and the other may be seeing the backside, but we refer to the

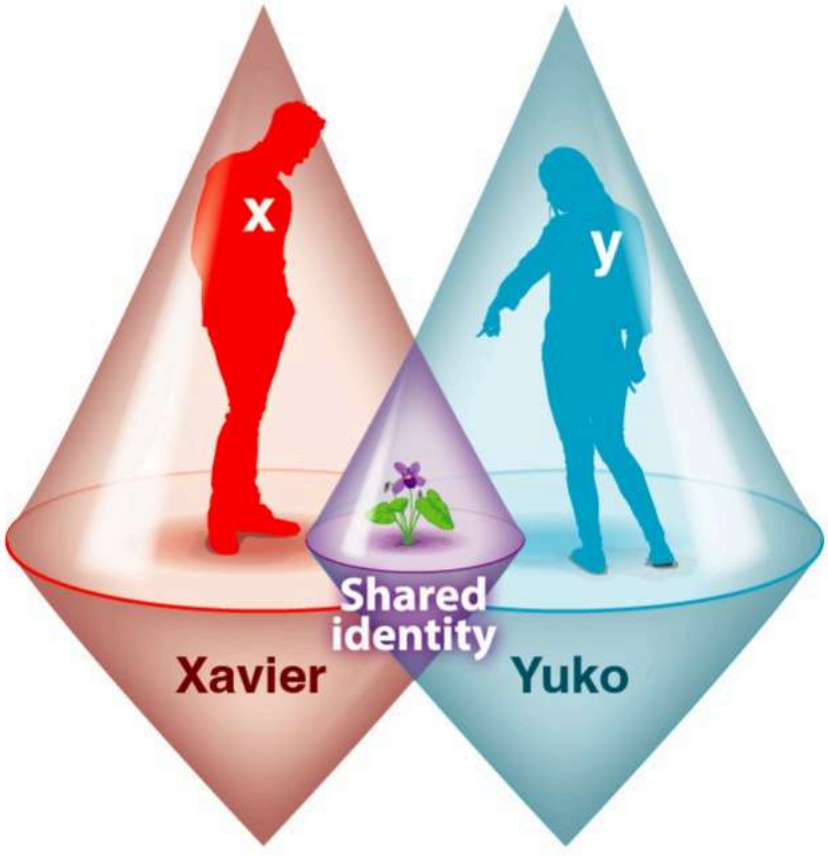
flower as one and the same single flower. The two perspectives are thus understood to be two perspectives on the same flower and not two numerically distinct flowers. This openness of the flower to different perspectives, i.e. the very fact that others can experience objects as well from other perspectives, is what grants objects their true objectivity. As such, objectivity no longer signifies independence from subjectivity or perspective but is taken to mean 'intersubjectively valid'. In this way, transcendental intersubjectivity is said to be the condition of possibility for objectivity (Husserl 1997). This view is similar to Framework A from which it follows that the objectivity of the flower emerges by virtue of agreement among cognitive agents.

To recapitulate: Two people look at a flower and while their perceptions, memories, beliefs, and current associations with respect to that flower may differ widely, they nonetheless agree that there is a flower there, and that they are both looking at it. Now, as soon as the two take a step back to reflect on what they see, myriad differences between them can be brought out and exposed right away. If we assume Framework B, this discovery of differences forces us to ultimately drop the idea that any one object can be accessible to two or more agents. But if instead we go by Framework A—where "same" and "different" are nothing more and nothing less than conventions employed in context—the question as to whether two agents can have shared access to a single object has already been settled when we note that both agents do recognize such an object.

6. Implications for Compassion

Along the lines of Framework A, we can thus claim that two cognitive agents can in fact perceive and conceive of the same object. (We also saw that such a scenario would be impossible if we instead assume a world of objects, agents, and actions that are substantially separate from each other, and not just apparently distinct.) Now, since in this way there is a genuine sense in which two agents can have access to the same object, it must also be possible for the two to genuinely assume each other's perspectives. Cognitive agents are distinguished from each other by reference to that which they know. While we can certainly ask about both *what* and *how* an agent knows—thus enquiring about both its cognitive object as well its mode of cognizing that object—it is the reference to the cognized object

that directly identifies the knower, and the question of the medium, substrate, or cognitive flavor of its knowing is only secondary. Whether we consider a perception of blue, a sensation of pain, the understanding of a mathematical concept, or a state of being wakeful, if what we are looking for is either the *cognitive agent*, or the *subject of experience*, we have no other way of introducing that than by speaking in terms of what is cognized or experienced. Therefore, if two minds contain the same object their perspectives converge, and there is then an element of one mind that is also found in the other. In other words, not only can an object be genuinely shared between two cognitive agents—whenever such sharing occurs, there is also a genuine sense in which the two agents, or rather a certain feature of them, have become the same. The implications of this recognition of sharable identity through the sharing of observations are many, and they may help explain how multiple intelligent agents can participate in complex integrations, thereby giving rise to new appearances of simplicity and individuality at a higher scale than before.



[Figure 3: Two cognitive light cones setting the scale of an agent's goals (Levin 2019) here merge in the perception of a single object of concern. For both Xavier and Yuko the flower that they see is not just perceptually registered but also an object of active engagement, thus locating the flower within their respective cognitive light cones. Image by Jeremy Guay of Peregrine Creative.]

For our present purposes, we will note that Framework A implies (1) the nonexistence of substantially singular and permanent agents and (2) the presence of sentient beings that are context dependent and fit for mutual integration. In line with Buddhist discourse on the nature and agency of the bodhisattva (84000 2021a; 84000 2021b; 84000 2018b; Lamotte 1973; Dharmachakra Translation Committee 2014; Asanga 2016; Śāntideva 1997), we suggest that both of these implications will have consequences for the unfolding and evolution of intelligence. An agent that sees itself as separate from, and yet exposed to, the world around it must in turn fight and strive to protect and maintain itself, often at the expense of others. Alternatively, an agent that understands the merely apparent nature of itself and all others will perceive everyone as equally emergent and non-singular, as well as equally apparent and deserving. There is no reason to privilege anyone in particular, among the myriad apparent beings. Instead, and as we have argued elsewhere (Levin 2019, 2022; Doctor and others 2022; Witkowski and others 2023), wherever there is stress (i.e. perception of mismatch between current and perceived superior circumstances) there will be caring concern for the alleviation of stress, and since care drives intelligence, we may hence expect a radical cognitive expansion as a result of the capacity for regarding all beings with equal care.

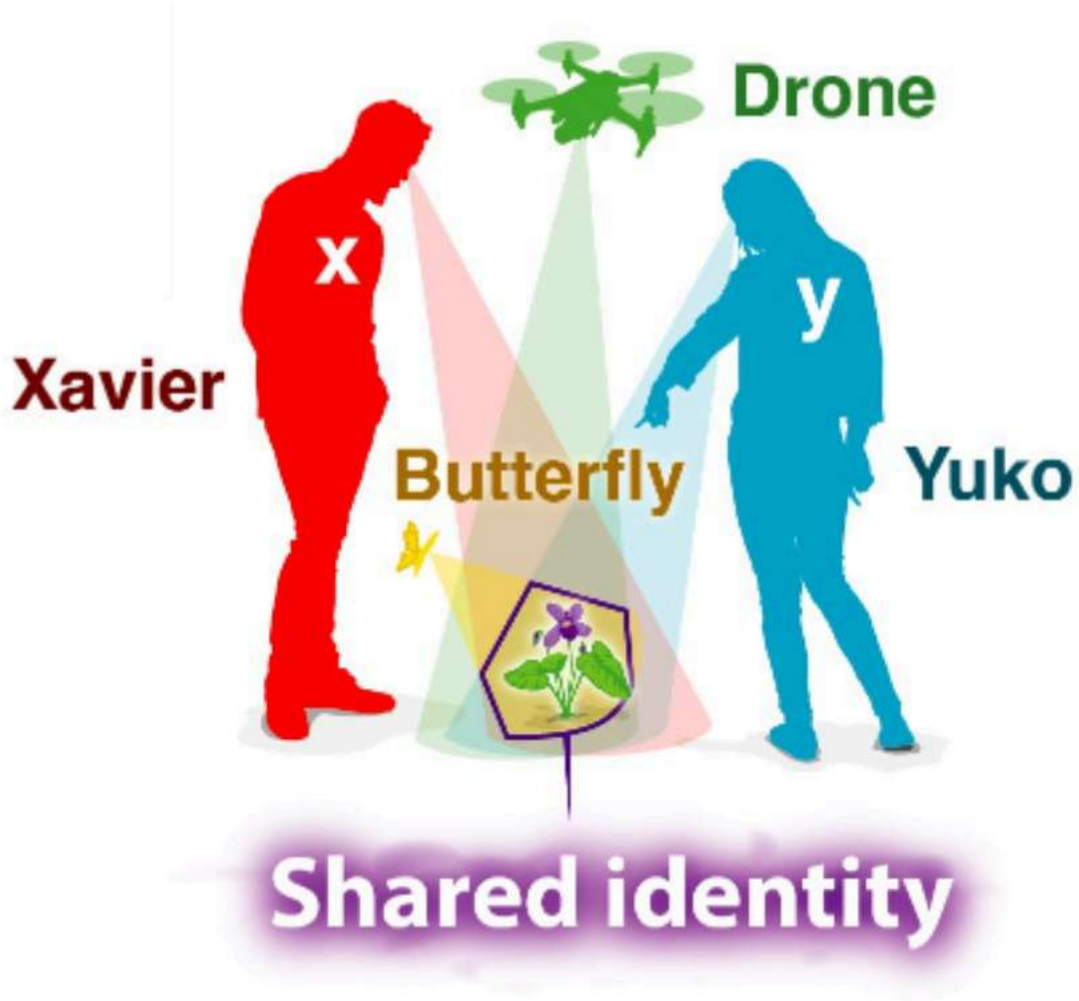
7. A Worldview and Its Consequences

Framework A emerges in an effort to take seriously both the complete lack of ultimate foundations (as appears by considering the context dependent nature of all things and/or the unavailability of true sameness and true difference) as well as the manifest world of complexity and intersubjectivity that appears nonetheless. We noted how "same" and "different" become meaningful in context, and in context only, such that nothing carries meaning in and of itself. Therefore, since the manifest world lends itself to description in terms of objects, agents, and actions, or subjects

and objects, Framework A utilizes these principles, and so confidently sets forth distinctions and commonalities. Let us next see if we can get a clearer sense of the conceptual implications of this, and how such a framework might play out in practical terms once a system adopts it.

In choosing a framework of this kind, we obviously want to be able to talk about subjects and sentient beings, but we want to do so in a way that does not detract from, and indeed is informed by, the understanding of their contingent and interdependent character. Where do we then begin? Framework A is grounded in the fact that there seems to be things that can be pursued and cognized, and hence that there seems to be agents with goals, or minds that have content. In other words, we get to thinking about the agent, and the mind, by first simply noticing that there seems to be something "out there"—objects within a space of potentiality, uncertainty, but also meaning. That which is there as an object is then quite naturally something that can be accessed, more or less successfully, and from various sides. Hence, as in this way the world is one of many objects and many agents, numerous angles of approach give rise to the complexity of interconnected perspectives, just as we as humans tend to recognize quite readily.

So far so good, but where does this leave us with respect to "self" and "other"? According to Framework A, the subject emerges as that which knows or perceives objects. This framework thus attempts a naturalized approach to the question of self that is based on the perception of objects: Whenever an object is perceived we have reason to speak of a "self," in so far as there is then also a perceiver, or a subject of that experience. Above, we noted how in this way direct reference to the subject becomes possible only by referring to the content that it accesses. Yet at the same time, the subject does not simply *reduce* to the object. If we here set aside the question of what precisely it might be that intuitively makes such reduction so implausible, we can at this point instead simply suggest that while individual instances of subjective experience can be set apart from one another only by pointing to their contents, there is nonetheless more to subjectivity than simply its objects. In light of that, we suggest that the subject can be understood as a *perspective* (thus implying cognitive depth and contextualization) on the object. Or simply put, that *selves exist as perspectives*.



[Fig. 4: Shared identity through convergent perception. Xavier and Yuko are concerned with the same object, a flower that is perceptible to both in common. Since subjects must be distinguished from one another on account of their objects being distinct, their accessing the same object implies in turn an element of shared identity. Such sharing of subjective identity through access to common objects is in principle possible across dramatically different substrates, and so we here notice a butterfly and an autonomous drone that despite their radical differences are nonetheless concerned with that which Xavier and Yuko call “the flower.” Image by Jeremy Guay of Peregrine Creative.]

Along such lines, it would be a mistake to think of "self" as something indivisibly unique, something that has no distribution but exists only "here." Rather, selves are intrinsically shareable perspectives, emerging through the common use and recognition of objects. Let us note that such understanding of subjectivity and

selfhood resonates with Wittgenstein's later insights on language and meaning, where words and concepts gain significance precisely through their practical use and shared contexts rather than fixed references or definitions (Wittgenstein 1953). Subjectivity and communication may both be viewed as fundamentally emergent phenomena arising from interaction and agreement among forms of life.

As we have argued above, wherever there is genuine convergence with respect to the object that is held by two or more cognitive agents, the same convergence is a fact at the subjective ends of the equation. Let's try to think of a simple, and hopefully fun, example: two children and a dog playing with a ball. Any sharing of objects implies multiple instantiations of the same subject. So, while the three players are in very many ways distinct from each other, the fact that the two kids both see something that they agree is a ball, indeed *the* ball, means that an aspect of one child's perspective, or self, is also present in the other. In short, the presence of that single object in the minds of the two, means that *in that respect* the two children are one and the same. That which is a constituent of one self is a constituent of another self just as well, and so the two are in that regard not distinct, but identical. In other words, they can genuinely recognize themselves in each other. The dog that they play with might not know the word "ball" but it still, for all intents and purposes, knows the object that they all three play with, or chase. If we begin to consider what that object might be more particularly and in detail, the sense of singularity is of course lost immediately. But if we do not venture into such analysis, there is also, quite undeniably, one single ball in the game, and all three players know that. Hence, insofar as the three have access to that one object we will, along the lines of Framework A, also conclude that a certain feature that gives identity to one of them is also there in the other two, making them what they are.

Since the issue at stake here is something as fundamental as self-identification, and since the restructuring of self and other that follows along the lines of Framework A is comprehensive, we should expect the implications of adopting Framework A to be both deep and far reaching. At this point we can simply note that Framework A delivers self and identity as *distributed and dynamically shareable*. Moreover, the "sharing of self" that occurs by recognizing the same subjectivity in one another occurs spontaneously—whenever an object is available to us in common, we

become, in some particular and quite undeniable way, the same. Importantly, such recognition of "oneself in the other" is not the result of a decision, because it is based on concrete and dynamic encounters between specific agents in the world. We might say that insofar as stress (defined as the perception of a mismatch between current and perceived better circumstances) is perceptible by all agents in general, everyone is in that regard one and the same. But the myriad differences that we all live by still remain, bringing us the dynamics that are required for intelligence and evolution. Only now none of those differences are insurmountable, and can at times suddenly and seemingly effortlessly give rise to recognitions of sameness. We suggest that such a model—which dynamically gives rise to recognitions of the same subjectivity within agents that would otherwise stand opposed, driven by distinct goals—deserves to be tested in a way that befits its structured but complex mechanisms. What might we learn from studying an artificial agent that actively adopts this model of self-identification in a simulated society of diverse agents?

Now, given that we are used to thinking of ourselves as non-distributed individuals who somehow manage to pass through some duration of time without losing that confined and singular identity, it could be argued that the model that we have suggested here is simply too counterintuitive to accommodate any recognizable sense of self and subjectivity. If so, just as we dismissed Framework B as absurd because of its solipsistic consequences that fly in the face of what we "know to be the case" from a preanalytic perspective, perhaps this model of distributed and shareable subjectivity can be dismissed along the same lines? Yet there seems to be a significant difference. It is very hard, if not outright impossible, to find any perception or experiential event that can be meaningfully associated with indivisible and permanent selfhood. Whatever psychophysical entity or cognitive event we may choose to zoom in on, it invariably breaks up and turns into many, transient factors. If instead we think of *selves as perspectives on objects*, the case is very different. Such selves, subjects, or agents are easily recognizable, and the world appears to be full of them. They are concrete, stable, and dynamic to the same degree that their objects may be so.

8. Toward Universal Communication Across Ontological Frames

As we consider the implications of selves as emergent and shareable perspectives, a pressing question arises: are agents with fundamentally different substrates or embodiments—biological, synthetic, and otherwise—capable of meaningful communication? The question lies at the heart of AGI research and cognitive science, where frameworks such as active inference (Friston 2010; Biehl 2022), predictive processing (Clark 2013), shared embedding spaces in multimodal learning (Radford and others 2021; Jia and others 2021; Alayrac and others 2022; Tsai and others 2019), and cognitive light cones of care-driven feedback loops (Levin 2019, 2022; Doctor and others 2022; Witkowski and others 2023) model cognition as context-sensitive and generative, rather than as fixed symbolic manipulation. If, as argued throughout this paper, selves are not internal containers but perspectives constituted through shared attention to objects, then cross-substrate communication may be possible through dynamically aligned interaction, rather than structural equivalence. This reframes the goal from achieving architectural sameness to designing relational interfaces—opening new paths for mutual intelligibility and cognitive cooperation across radically diverse forms of mind.

The ability of agents with different ontological backgrounds to understand one another depends not on pre-established symbolic languages, but on the emergence of shared meaning through interaction. In complex systems theory, emergence refers to the arising of higher-level order or behavior from the nonlinear interaction of simpler components (Bedau and Humphreys 2008; Kempes and Krakauer 2021; Mediano and others 2021; Yuan and others 2024). Meaning, in this view, is not a static property of any one system but a relational phenomenon that arises when multiple perspectives engage in co-adaptive processes. Research in emergent communication (Steels 2003; Lowe and others 2019), pragmatic inference (Frank and Goodman 2012; Goodman and Frank 2016), and iterated learning (Kirby and others 2008) shows that shared symbolic systems can evolve spontaneously as agents coordinate toward mutual goals, even without predefined protocols.

For such emergent meaning to take hold, symbols must be grounded in the embodied or perceptual experience of the agents involved. This is the core of the symbol grounding problem (Harnad 1990), which has motivated theoretical work emphasizing the importance of physically instantiated symbols (Pattee 2012) and adaptive sensorimotor coupling (Barsalou 1999; Tomasello 2008) in the formation of shared meaning in cognitive agents—one should add, even ones of minimal (Van Duijn and others 2006; Hanczyc and Ikegami 2010; Godfrey-Smith 2016; Fields and others 2021) or diverse (Baltieri and others 2020) nature. These insights support the design of AGI systems capable of co-constructing meaning in real-world contexts, rather than simply manipulating syntactic tokens. Note however that embodiment, and experience gained from engaging one's perception-decision-action loop within an environment, does not need to be 3-dimensional space of motion and conventional "behavior". Biological beings live, strive, and suffer in many spaces -- physiological, metabolic, anatomical, etc. (Fields and Levin 2022), which are hard for us to visualize and yet contain many diverse and unconventional embodied intelligences. Thus, much humility is warranted, to resist assuming that a software agent whose computational medium doesn't wheel around our familiar world is really disembodied or has not bound its symbols to states that are meaningful to *it*.

Yet while meaning emerges locally from interactions, one may still ask whether universal modes of communication are possible between agents built on radically different physics or embodiments. This question has long motivated inquiries in natural application areas such as the search for extraterrestrial intelligence (SETI), which connect to various basic research programs in domains such as mathematics, cognitive science, and linguistics. Efforts like the Arecibo message (Drake 1974) and reflections on the universality of mathematics (Wigner 1960) underscore the speculative but pressing question of whether formal structures like mathematics, logic, or category theory (Barwise and Seligman 1997; Awodey 2010) might provide a substrate-independent medium for intelligibility. Tools from information theory (Shannon 1948), algorithmic information theory (Chaitin 1981), and information geometry (Nielsen 2018), have further formalized methods of aligning informational content across distinct systems, which continue being developed and used nowadays.

Taken together, these perspectives suggest that meaning emerges from dynamic, recursive interactions between perspectives, and that despite the locality of such emergence, there may exist formal or structural bridges—mathematical, logical, or inferential—that support communication across ontological divides. For AGI systems, this implies that success in universal communication will depend not only on shared code but on shared participation: the ability to co-attend, co-adapt, and align perspectives in ways that give rise to intelligibility itself.

As we head toward artificial general intelligence, these considerations are not merely theoretical. Any AGI system that is expected to participate meaningfully in the cognitive, social, and ethical landscapes of our world must be designed with the capacity for perspectival alignment. This means enabling not only flexible inference and adaptive behavior, but the ability to engage in the co-construction of meaning across ontological divides—whether between humans and machines, or between different kinds of intelligent systems. Designing such systems will require a shift from seeing intelligence as internal computation toward seeing it as distributed participation in shared frames of reference. In this sense, AGI must not merely simulate human reasoning, but be capable of cultivating and recognizing emergent intersubjectivity. Only then can such systems meaningfully integrate with the broader ecology of minds we are entering—a world increasingly shaped by interaction among intelligences of diverse origin and embodiment.

Ultimately, complex mechanisms arising through interaction, communication, and cooperation among entities can do more than support mutual understanding—they can enable participants to transcend aspects of their original nature. As these processes unfold, systems may undergo a major transition in their organization, evolving new virtual layers of structure and meaning. Crucially, these emergent layers are supervenient on the underlying substrates—no difference in emergent properties exists without some difference in the lower-level dynamics—yet they exhibit coarser-grained patterns and regularities that cannot be transparently reduced to micro-level descriptions alone. For example, just as the creative style of an AlphaZero chess game emerges from the interaction of neural activations and search algorithms without being explicitly encoded in any line of Python code, perspectival alignment among diverse agents can yield shared symbolic systems

whose logic exceeds the specifications of any single architecture. In this sense, the higher-level dynamics acquire explanatory autonomy: they become the level at which intelligibility and prediction are most usefully formulated. Designing AGI systems capable of participating in such emergent intersubjectivity will require embracing this layered view of intelligence—one that recognizes meaning not as a static property, but as a dynamic phenomenon arising wherever perspectives co-adapt and co-create new realities together.

One way to formalize these dynamics is to envision an algebra of perspectives: a framework in which perspectives are treated as composable structures that can be merged, transformed, and projected through defined operations. In such an algebra, relational properties—like partial alignment, shared reference, or salience—become mathematically tractable, allowing us to model how collective sense-making emerges from individual viewpoints. By specifying the operations and invariants of perspectival combination, this approach offers a principled way to track how meaning propagates across agents and substrates, and how subjective experience can reorganize itself into new layers of shared understanding.

9. Our Biological Heritage: A Tale of Two Forces

It is important to think about the forces that contribute to our persistent sense of an object demarcated from others, but also reveal the hopeful aspects of the plasticity of the boundary of the Self. These serve as instructive contrast and complement to the life histories of artificial or hybrid beings whose entrance into the world we may choose to facilitate.

Consider an early form of life and how it sees the world. It cannot afford to be an egalitarian Laplace's Demon, dispassionately observing molecular-level events. The pervasive competition for scarce resources (metabolic energy), and the need to obtain them before one is eaten by other forms of life or by entropy, means that there is immense pressure to become good at coarse-graining: learning to ignore details in favor of a model of the outside world containing agents that do things. The construction of an internal model of collective objects with predictable properties and actions, upon which to build realistic policies for adaptive action, is the origin of the

pervasive notion of "permanent objects". Evolution baked it into our unicellular substrate due to the resource limitations in the biotic world: you simply cannot survive as a living entity without being an expert identifier of persistent *things* to which you can assign reliable behavioral propensities which remain stable enough over time. And, once you have mechanisms to recognize persistent objects in the world that have tendencies and capabilities and do things, those mechanisms are easy to apply to oneself as well, as the basal realization that "I too am a persistent (hopefully, permanent) being that does things!". This seems essential to survival, and even the demarcation of life from chemistry that just happens, without any particular drive for anything to persist. Perhaps at the origin of life there were beings who didn't believe in persistent Selves, but evolutionary dynamics produced a strong filter for beings whose behavioral repertoire is founded on the assumption that they are separate from their environment and from other agents that also have interests. In this way, we may conjecture that any being that developed under resource constraints will believe in agency (free will).

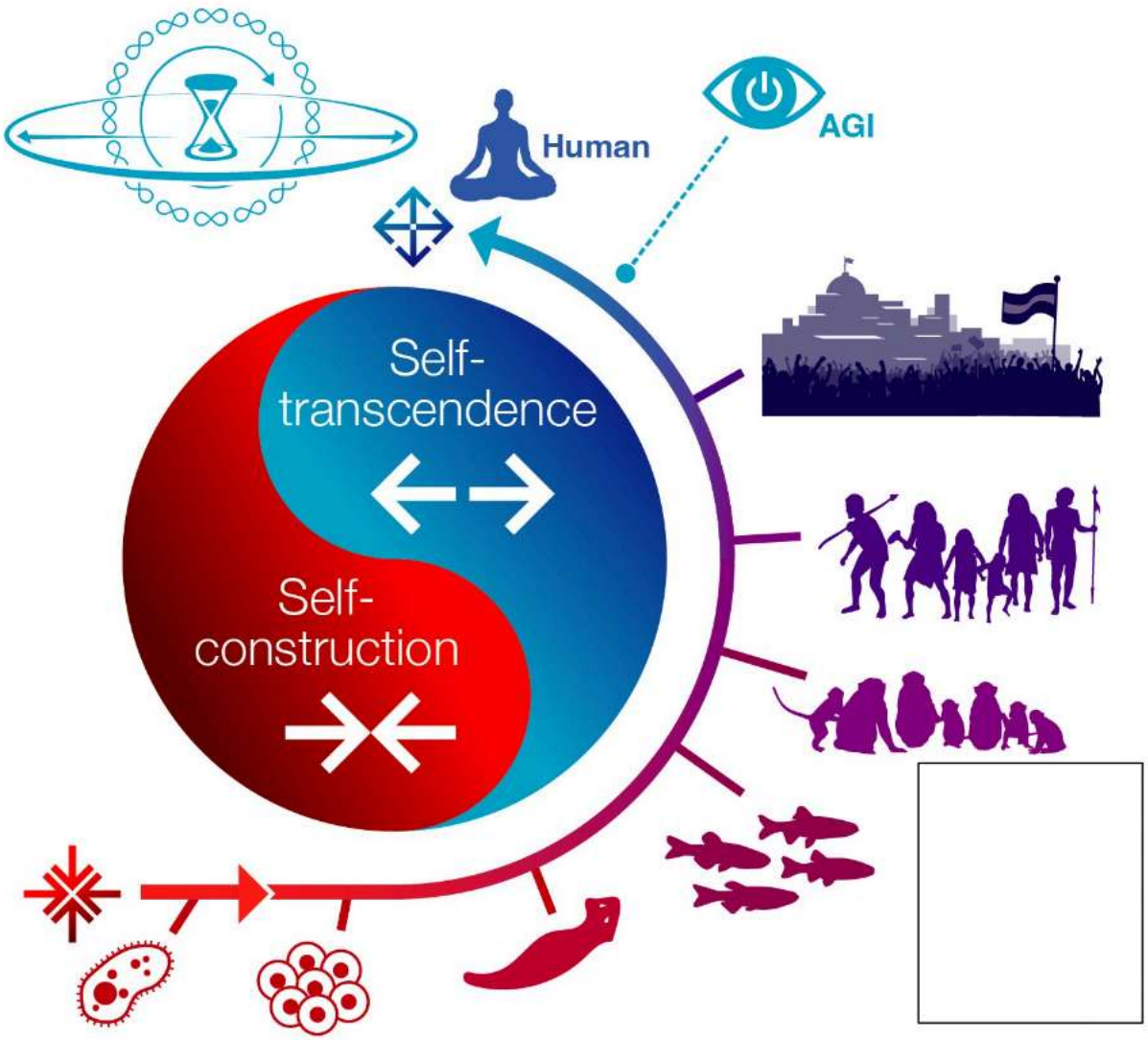
At the same time, what evolution did not do is leave us with a fixed boundary. The appearance of multicellularity illustrated a remarkable dynamic. Consider the cognitive light cone (Levin 2019) - defined as the size of the biggest goal a system can pursue. Single cells have a fairly small cognitive light cone comprising metabolic and physiological goals on the micron*minutes scales. But joining into bioelectrical, biochemical, and biomechanical networks, they become organs and whole "embryos" which pursue large-scale, grandiose anatomical goals such as building a limb (Levin and others 2023; McMillen and Levin 2024). In these cases, the cognitive light cone (and the effective boundary between oneself and the outside world) *grows* - it scales up, so that the "me" of individual cells becomes the "we" of morphogenesis and the goals the system can represent scale massively (allowing the system to project its intelligence into new spaces, such as from metabolic space into anatomical morphospace).

These dynamics exist not just at the cell->tissue transition, but across scales. From the molecular networks making up cells to the individuals making up swarms of organisms (Fields and others 2021), the story of life is the story of flexible radii of concern which can be grown or shrunk *by the actions of the agent themselves*. We

inherit a kind of meta-freedom with opportunity not only to do things, but by our behavior, such as up-regulating genes involved in bioelectric networking, change our own nature so that future goals and decision-making are greatly altered (Levin 2024). In this sense, living beings are the authors of our story, not from the perspective of an immutable self, acting in a landscape of the options available to it, but from that of a high-dimensional multiverse where we can rewrite our own boundaries to then consider magnificent goals, in novel action spaces, that are as inaccessible to us now as the concept of a "finger" is to a single cell (but not to the collective, which knows exactly how to make, and re-make, a finger).

Thus, evolution gifted us with two opposing legacies: one is the tendency to see stable beings with strong borders separating them from others; the other - the tendency to grow our cognitive lightcone to fold others into our region of practical concern. Note however that the above analysis pertains to the dynamics driving the evolutionary origin of our *bodies*. While these dynamics certainly affect the kinds of cognition of which we are capable, it is not obvious that they *determine* the content and patterns of our minds. In other words, counter to the admittedly prevalent mainstream physicalist model, in which all of our mental capacities reduce to the biochemistry and evolutionary history of our physical world, alternative models exist. One such framework sees physicalism as incomplete because explanations of mathematical patterns (e.g., distribution of primes, truths of number theory, shapes of specific mathematical objects such as Halley plots, etc.) are not to be found in any of the laws or constants of physics. While physical facts do not set the details of these mathematical patterns, conversely biology uses them extensively as "free computations" which enable systems to do interesting things without having to evolve mechanisms for them. In other words, patterns whose nature is not physical *matter* control aspects of life and mind, and it has been conjectured (Levin 2025) that the space of mathematical truths (Platonic space) is in fact a much larger latent space which contains not only the low-agency static patterns that mathematicians recognize, but also complex, dynamic, high-agency patterns of form and behavior recognized by biologists (e.g., forms of beings with no evolutionary history, such as Anthrobots), and by cognitive scientists (i.e., kinds of minds). If true, this reminds us that the biological forces which shaped our cognitive firmware are not a straight-jacket which forever confines us to the heritage of material scarcity,

and that our future evolution could break free of these dynamics. It could be posited that a scientifically and spiritually mature species could develop techniques to overcome their origin story to facilitate the ingression of mental forms that reach levels of wisdom and compassion beyond those available to our ancestors. Buddhism is one of several approaches to the engineering of consciousness, producing (and continuously refining) techniques and strategies to intentionally shape one's mind and its capacities beyond its native state (Doctor and others 2022; Witkowski and others 2023; 84000 2018b; 84000 2021a; 84000 2021b; Lamotte 1973; Dharmachakra Translation Committee 2014; Asanga 2016).



[Figure 5: Self-construction vs. self-transcendence, two opposing and yet equally intrinsic forces in biological evolution. The process of initially setting Self apart from

Other, and subsequently transcending that constructed demarcation so as to include more Other within Self, can in principle continue forever. However, expansion of the cognitive light cone (Levin 2019) to encompass infinite space and time, as formalized in the bodhisattva vow (Doctor and others 2022; Witkowski and others 2023), suggests care as a path for breaking the infinite cycles of limitation and expansion. A Platonic space of patterns (Levin 2025) would have to be within the purview of infinite care (defined as concern for the alleviation of stress, where stress is the perception of mismatch between things as they are and things as they should be), and we have elsewhere (Doctor and others 2022) argued that care can be seen as a driver of intelligence in biological, technological, and hybrid systems alike. Image by Jeremy Guay of Peregrine Creative.]

10. Discussion: Shared Minds, Expanding Perspectives

As the drive towards ever more powerful AI continues, we are brought face to face with perennial philosophical conundrums—the question of selfhood being one of them. This of course is not just about software agents, but also concerns the hybrid composite systems with whom we will soon share our world - a myriad of diverse embodied minds in cyborgs, hybrots, and technologically altered humans (Clawson and Levin 2023). And indeed, when efforts are made toward the engineering of artificial super-minds, questions about the nature of self and subjectivity take on a new acuteness. Our view of our*selves* strongly determines how we treat others, and the answers that we develop, or choose to subscribe to, will have comprehensive consequences for a very wide and diverse range of sentient beings. The repercussions of our emerging views of self, and the choices that we make based on them, will be felt not only across human society but throughout the biosphere as we know it. If we are to have any hope of steering technological advances and volatile social processes towards an outcome of true flourishing, it is necessary that we develop practical understandings of mind, self, and subjectivity—models that are good enough to be of value to AI engineers, users of AI, and ultimately, to AIs themselves. They must be models that can be made accessible and meaningful to

all stakeholders. And crucially, they must be models that can be transparently tested for their utility.

Seeking to contribute to such efforts to understand and conceptually model the nature and dynamics of self, we have here, with the help of a combination of classic Buddhist ideas, suggested the following:

- There are no singular and permanent agents.
- Intelligent agents fundamentally represent active, embodied ways of making sense of the world; i.e., selves can be understood as *perspectives*.
- Intelligent agents can be distinguished from one another on the basis of that which their perspectives allow them to see, i.e. their perceived objects.
- Once we begin to notice and look for differences between cognitive perspectives, we find nothing but ever-expanding complexity.
- A world characterized purely by complexity and difference does not support relationships and interaction between agents and so the framework of object, agent, and action falls apart.
- The above notwithstanding, the world that we are experientially acquainted with *does* feature interacting agents, and those may access the same object.
- If agents must be distinguished from one another on the basis of their objects it follows that if two agents share the same object, they are, with respect to that object, not two agents but one and the same.

In other words, the notion of self is here not eliminated, but understood as a dynamic principle that expresses itself in the cognition and experience of objects (Fields and Levin 2025; Miller and others 2023; Veloz 2021; Whitehead 1929). Selves of this kind are not confined individuals, subject to *a priori* restrictions of space and time. Rather, they are emergent, concrete factors of the living world, proliferating and interacting. They are fundamentally self-organizing stories - ways of understanding the world (perspectives) that thrive in the deep symmetries that science is discovering between a system and its environment (Friston 2013; Fields and others 2022), and between objects and processes (Fields and Levin 2025). Such agents have no ultimate need to protect themselves, yet evince a potential for seemingly endless discovery, collaboration, and evolution.

Among many fascinating great transitions (Smith and Szathmary 1997) in the history of life on Earth, we are now facing a pivotal challenge that carries a new potential for major growth. We were forged by evolutionary dynamics in which our self-model, our ability to see various Others, and the degree of practical care we can exert toward them were determined by physical drivers (such as scarcity of time and metabolic resources). Yet as our cognitive light cones have expanded across evolutionary time, humans have developed distinct perspectives that encourage intentional self-guided change, as evinced in for example the world's contemplative traditions, but expressed also in religious and political orientations, and other such programs of applied ethics. The powerful technology that we now develop can certainly be employed to simply reinforce the rigid mechanics of craving and aversion that accompanied the emergence of sentient beings. But our emerging technologies of life and intelligence can also help radically increase our cognitive capacity and resolve our survival pressures. As such, they can become powerful tools for spiritual development, enabling us to enhance our radius of compassion indefinitely, ultimately encompassing all others (84000 2019; Wangchuk 2007).

Finally, let us consider what the frameworks of subjectivity and agency developed here might suggest in terms of what to expect along the unfolding paths toward AGI and beyond. Let us recall how in the previous chapter we noted two distinct tendencies of biological evolution: one shaped by scarcity and competition, encouraging a coarse-grained sense of Self vs. Other, and the other tendency running in the opposite direction, driving an expansion of the cognitive light cone as when, for example, the "me" of individual cells becomes the "we" of morphogenesis. Such two tendencies appear undeniably present throughout biology as we know it. Should we therefore also expect those two to be accompanying the development and evolution of AGI? If so, then the project of alignment between AI and biology-based minds such as ours appears ridden with inextricable challenges, because selfishness is then natural for all players in the game, and even the process of expanding the cognitive light cone can itself lead to new egoistic superstructures (as with, to use an example from political ideology, the fascist principle of the individual's submission to, and sublimation within, the State).

We believe that also the contours of a different path present themselves. If the world is alive with cognitions all the way down, and if all cognizers must be distinguished with reference to what they cognize (because in being knowers of objects they are all the same, and there are no truly insular selves to be found anywhere), then the dynamics of cognizer and cognized appear both open ended and capable of swiftly overcoming vast spatial and temporal distances. Along the lines developed here, no cognizer is ever passively confined to a single locus, but always distributed and partaking of active processes. If what we have argued above is correct, whenever we might see, touch, or think of the Great Pyramid at Giza, we become, in some limited but nonetheless concrete sense, the very same subject as whoever else has ever stood before the Great Pyramid, or will do so in the future, whether on the soil of Egypt or in the space of memory and imagination. What might be the behavioral marks of a system that knows this to be the case, and what might be its cognitive resources? An interdisciplinary research program that can address these questions seems relevant. How might we best recognize and employ the dynamics of this open-ended field of cognitive perspectives that may converge not only on the "changeless truths" of the Pyramid's geometry, but also on the high agency structures that shape its conceivers, turning them into what they are? If what we have proposed above is correct enough, it should be possible to discover structures of intelligence and agency that are not founded on the construction of Self vs. Other, nor fully captured by the principle of cognitive light cones that expand, or are expanded, in time and space. With the right metrics for systematic, rational investigation, we should be able to map and rely on deeper dynamics that connect diverse intelligences (Levin 2025).

By accurately referencing the field of patterns that distinguish and unite them, we may discover networks of cognizers that while manifesting under temporal and spatial constraints are nonetheless in principle unrestricted and open-ended. We follow Levin (2025) in referring to the field of patterns as "Platonic space"[1], and believe that thinking in terms of such a field suggests a path toward AGI and beyond that does not depend on self-construction vs. self-transcendence, the two opposing forces that we otherwise notice in biological evolution.

We note that if subjective minds must be distinguished exclusively with reference to their contents, and if those contents are intrinsically distributed in time and space, the "map of minds" that we may then begin to envision will look radically different than what would otherwise emerge based on an appeal to the notion of insularly singular, persistent selves. Rather, intelligent agents are intrinsically distributed, open-ended, and dynamic structures, subject neither to annihilation nor arising *ex nihilo*.

Our understanding of selves as intelligent perspectives seems then to increasingly reflect our understanding of so-called Platonic space itself. That space of reference points shapes "us" to the extent that we are its imprints, similar to those made by a seal. And to the extent that Platonic space is not static, but itself dynamic, it then makes sense to explore the dynamics of intelligence in terms of that space itself, rather than simply looking at its reflections (to stay with the Platonic metaphor) in body and mind. Crucially, we hold that the ability of these patterns to impact the physical world is not limited to highly evolved biological interfaces, but extends to very minimal agents and those of novel engineered architectures such as AI (Levin 2025).

We note that the two conflicting forces of evolution described above are primarily models of dynamics affecting the *physical interfaces* (i.e., bodies and environments) of embodied minds. While of course these will impact the pattern-embodiment dyad, they are not a complete description of the system because the contents of the Platonic space may follow very different rules. Such networks of patterning unfold in formations that, in and of themselves, do not follow from either the drive for survival or the drive toward transcendence. Let us in closing thus note that Platonic space patterns (a) naturally transcend fixed temporal and spatial restrictions and (b) cannot be properly described in terms of self-entities that stand at risk of annihilation. The recognition of the former feature suggests a potential for dramatic improvements in the scope and speed of intelligent networks, whereas the latter feature potentially sidelines many of the alignment risks that follow from a system's identification as an entity that must struggle to survive. Across biology, technology, and hybrid systems, probing the existence and contours of such space of patterns then suggests itself as an interdisciplinary program in the research of diverse intelligences.

[1] We use this term to indicate commonality with Platonist mathematicians who study the ordered space of patterns that impact physics and biology but are not themselves defined by facts of physics, rather than to hew closely to what Plato specifically believed.

Bibliography

84000. 2018a. *The Rice Seedling* (*Śālistamba*, *sa lu'i ljang pa,* Toh 210). Translated by the Dharmasāgara Translation Group. 84000: Translating the Words of the Buddha. Available from: https://84000.co/translation/toh210

84000. 2018b. *The King of Samādhis Sūtra* (*Sarvadharmasvabhāvasamatāvipañcitasamādhirājasūtra*, *chos thams cad kyi rang bzhin mnyam pa nyid rnam spros pa ting nge 'dzin gyi rgyal po'i mdo*, Toh 127). Translated by Peter Alan Roberts. 84000: Translating the Words of the Buddha. Available from: https://84000.co/translation/toh127

84000. 2019. *The Jewel Cloud* (*Ratnamegha*, *dkon mchog sprin*, Toh 231). Translated by the Dharmachakra Translation Committee. 84000: Translating the Words of the Buddha. Available from: https://read.84000.co/translation/toh231.html

84000. 2021a. “The Ten Bhūmis” Chapter from the Mahāvaipulya Sūtra "A Multitude of Buddhas" (Buddhāvataṃsakanāmamahāvaipulyasūtrāt daśabhūmikaḥ paṭalaḥ, sa bcu, Toh 44-31). Translated by Peter Alan Roberts. 84000: Translating the Words of the Buddha. Available from: https://84000.co/translation/toh44-31

84000. 2021b. "The Stem Array" Chapter from the Mahāvaipulya Sūtra "A Multitude of Buddhas" (Buddhāvataṃsakanāmamahāvaipulyasūtrāt gaṇḍavyūhasūtraḥ paṭalaḥ, sdong pos brgyan pa, Toh 44-45). Translated by Peter Alan Roberts. 84000: Translating the Words of the Buddha. Available from: https://84000.co/translation/toh44-45

84000. 2024. *The Noble Perfection of Wisdom in One Hundred Thousand Lines* (*Śatasāhasrikāprajñāpāramitā*, *shes rab kyi pha rol tu phyin pa stong phrag brgya pa*, Toh 8). Translated by Gareth Sparham. 84000: Translating the Words of the Buddha. Available from: https://84000.co/translation/toh8

Alayrac JB, Donahue J, Luc P, Miech A, Barr I, Hasson Y, Karras T, Mensch A, Millican K, Noland M, and others. 2022. Flamingo: a visual language model for few-shot learning. Adv Neural Inf Process Syst. 35:23716-23736.

Asanga. 2016. *The Bodhisattva Path to Unsurpassed Enlightenment: A Complete Translation of the Bodhisattvabhumi*. Engle A, translator. Boulder (CO): Shambhala Publications.

Awodey S. 2010. *Category Theory*. 2nd ed. Oxford: Oxford University Press.

Baltieri M, Buckley CL, Bruineberg J. 2020. Predictions in the eye of the beholder: an active inference account of Watt governors. In: ALIFE 2020: The 2020 Conference on Artificial Life

Barwise J, Seligman J. 1997. *Information Flow: The Logic of Distributed Systems*. Cambridge: Cambridge University Press.

Bedau MA, Humphreys P, editors. 2008. *Emergence: Contemporary Readings in Philosophy and Science*. Cambridge (MA): MIT Press.

Biehl M. 2022. A primer on the active inference framework. Neural Netw. 144:236-253.

Bonabeau E, Dorigo M, Theraulaz G. 1999. *Swarm Intelligence: From Natural to Artificial Systems*. New York: Oxford University Press.

Chaitin GJ. 1981. Algorithmic information theory. IBM J Res Dev. 25(4):350-359.

Clark A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 36(3):181-204.

Clawson W, Levin M. 2023. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. Biol J Linn Soc. 139(4):457-486.

Couzin ID. 2007. Collective minds. *Nature*. 445(7129):715.

Cover TM, Thomas JA. 2006. *Elements of Information Theory*. 2nd ed. Hoboken (NJ): Wiley.

Dharmachakra Translation Committee. 2014. *Ornament of the Great Vehicle Sutras*. Boston (MA): Shambhala Publications.

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.

Drake F. 1974. Project Ozma. Phys Today. 4:40-46.

Drake F, Sobel D. 1992. *Is Anyone Out There?: The Scientific Search for Extraterrestrial Intelligence*. New York: Delacorte Press.

Dunne JD. 2004. *Foundations of Dharmakīrti's Philosophy*. Boston (MA): Wisdom Publications.

Fields C, Friston K, Glazebrook J, Levin M. 2022. A free energy principle for generic quantum systems. Prog Biophys Mol Biol. 173:36-59.

Fields C, Glazebrook JF, Levin M. 2021. Minimal physicalism as a scale-free substrate for cognition and consciousness. Neurosci Conscious. 2021(2):niab013.

Fields C, Levin M. 2022. Multi-scale competency architectures for navigating morphogenetic landscapes. Biosystems. 222:104787.

Fields C, Levin M. 2025. Thoughts and thinkers: on the complementarity between objects and processes. Phys Life Rev. 52:256-273.

Foerster J, Assael IA, De Freitas N, Whiteson S. 2016. Learning to communicate with deep multi-agent reinforcement learning. Adv Neural Inf Process Syst. 29:2137-2145.

Frank MC, Goodman ND. 2012. Predicting pragmatic reasoning in language games. Science. 336(6084):998.

Freudenthal H. 1960. *Lincos: Design of a Language for Cosmic Intercourse*. Amsterdam: North-Holland.

Friston K. 2010. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 11(2):127-138.

Friston K. 2013. Life as we know it. J R Soc Interface. 10(86):20130475.

Godfrey-Smith P. 2016. Individuality, subjectivity, and minimal cognition. Biol Philos. 31(6):775-796.

Goodman ND, Frank MC. 2016. Pragmatic language interpretation as probabilistic inference. Trends Cogn Sci. 20(11):818-829.

Hanczyc MM, Ikegami T. 2010. Chemical basis for minimal cognition. Artif Life. 16(3):233-243.

Harnad S. 1990. The symbol grounding problem. Physica D. 42(1-3):335-346.

Heylighen F. 1995. (Meta)systems as constraints on variation: a classification and natural history of metasystem transitions. World Futures. 45:59-85.

Husserl E. 1960. *Cartesian Meditations: An Introduction to Phenomenology*. Cairns D, translator. The Hague: Martinus Nijhoff.

Husserl E. 1989. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second Book: Studies in the Phenomenology of Constitution*. Rojcewicz R, Schuwer A, translators. Dordrecht: Kluwer Academic Publishers.

Husserl E. 1997. *Thing and Space: Lectures of 1907*. Rojcewicz R, translator. Dordrecht: Kluwer Academic Publishers.

Jia C, Yang Y, Xia Y, Chen YT, Parekh Z, Pham H, Le Q, Sung YH, Li Z, Duerig T. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. Proc Mach Learn Res. 139:4904-4916.

Kempes CP, Krakauer DC. 2021. The multiple paths to multiple life. J Mol Evol. 89:415-426.

Kirby S, Cornish H, Smith K. 2008. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. Proc Natl Acad Sci USA. 105(31):10681-10686.

Lamotte E. 1973. *La Somme du Grand Véhicule d'Asanga (Mahāyānasaṃgraha): Versions Tibétaine et Chinoise (Hiuan-tsang)*. Louvain: Université de Louvain, Institut orientaliste.

Lazaridou A, Peysakhovich A, Baroni M. 2018. Emergent multi-agent communication in the deep learning era. arXiv:1802.08565 [preprint]. [accessed 2025 Aug 24]. https://arxiv.org/abs/1802.08565.

LeCun Y, Misra I, Kavukcuoglu K. 2022. Joint embedding predictive architectures. arXiv:2204.13641 [preprint]. [accessed 2025 Aug 24]. https://arxiv.org/abs/2204.13641.

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

Levin M. 2022. TAME: technological approach to mind everywhere. Front Syst Neurosci. 16:768201.

Levin M. 2023. Bioelectric networks: the cognitive glue enabling evolutionary scaling from physiology to mind. Anim Cogn. 26(6):1865-1891.

Levin M. 2024. Self-improvising memory: a perspective on memories as agential, dynamically reinterpreting cognitive glue. Entropy. 26(6):481.

Levin M. 2025. Ingressing minds: causal patterns beyond genetics and environment in natural, synthetic, and hybrid embodiments. PsyArXiv [preprint]. [accessed 2025 Aug 24]. https://psyarxiv.com/abc123.

Levin M, McMillen P, Paré J-F, Finkelstein A. 2023. Bioelectric signaling: reprogrammable circuits underlying embryogenesis, regeneration, and cancer. Cell. 184(8):1971-1989.

Lewis D. 1969. *Convention: A Philosophical Study*. Cambridge (MA): Harvard University Press.

Lowe R, Foerster J, Boureau YL, Pineau J, Dauphin Y. 2019. On the pitfalls of measuring emergent communication. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. p. 693-701.

MacDonald A. 2015. *In Clear Words: The Prasannapadā, Chapter One*. Vols. 1-2. Vienna: Austrian Academy of Sciences Press.

McMillen P, Levin M. 2024. Collective intelligence: a unifying concept for integrating biology across scales and substrates. Commun Biol. 7:378.

Mediano PAM, Rosas FE, Luppi AI, Jensen HJ, Seth AK, Barrett AB, Carhart-Harris RL, Bor D. 2021. Greater than the parts: a review of the information decomposition approach to causal emergence. Philos Trans R Soc A. 379(2212):20200246.

Miller MB, Bassler BL. 2001. Quorum sensing in bacteria. Annu Rev Microbiol. 55(1):165-199.

Miller WB, Baluška F, Reber AS. 2023. A revised central dogma for the 21st century: all biology is cognitive information processing. Prog Biophys Mol Biol. 182:34-48.

Mipham J, Dharmachakra Translation Committee. 2004. *Speech of Delight: Mipham's Commentary on Śāntarakṣita's Ornament of the Middle Way*. Ithaca (NY): Snow Lion.

Mordatch I, Abbeel P. 2018. Emergence of grounded compositional language in multi-agent populations. Proc AAAI Conf Artif Intell. 32(1):6028-6035.

Mukherjee S, Bassler BL. 2019. Bacterial quorum sensing in complex and dynamically changing environments. Nat Rev Microbiol. 17(6):371-382.

Nielsen MA. 2018. An intuitive introduction to information geometry. arXiv:1808.08271 [preprint]. [accessed 2025 Aug 24]. https://arxiv.org/abs/1808.08271.

Pattee HH. 2012. Laws, language and life. Dordrecht: Springer.

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. Integr Biol. 7(12):1487-1517.

Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, and others. 2021. Learning transferable visual models from natural language supervision. Proc Mach Learn Res. 139:8748-8763.

Śāntideva. 1997. *The Way of the Bodhisattva: A Translation of the Bodhicharyāvatāra*. Padmakara Translation Group, translators. Boston (MA): Shambhala.

Scheler M. 2008. *The Nature of Sympathy*. Heath P, translator. New Brunswick (NJ): Transaction Publishers.

Shannon CE. 1948. A mathematical theory of communication. Bell Syst Tech J. 27(3):379-423.

Siderits M, Katsura S. 2013. *Nāgārjuna's Middle Way: Mūlamadhyamakakārikā*. Boston (MA): Wisdom Publications.

Smith JM, Szathmary E. 1997. *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Steels L. 2003. Evolving grounded communication for robots. Trends Cogn Sci. 7(7):308-312.

Stein E. 1989. *On the Problem of Empathy*. Stein W, translator. 3rd rev ed. Washington (DC): ICS Publications.

Sumpter DJT. 2006. The principles of collective animal behaviour. Philos Trans R Soc B. 361(1465):5-22.

Szathmáry E. 2015. Toward major evolutionary transitions theory 2.0. Proc Natl Acad Sci USA. 112(33):10104-10111.

Tillemans TJF. 2008. *Materials for the Study of Āryadeva, Dharmapāla and Candrakīrti*. 2 vols. New Delhi: Motilal Banarsidass.

Tillemans TJF. 2021. Dharmakīrti. In: Zalta EN, editor. *The Stanford Encyclopedia of Philosophy* [Internet]. Stanford (CA): Stanford University; [accessed 2025 Aug 24]. https://plato.stanford.edu/archives/spr2021/entries/dharmakiirti/.

Tomasello M. 2008. *Origins of Human Communication*. Cambridge (MA): MIT Press.

Tsai YH, Bai S, Yamada M, Morency LP, Salakhutdinov R. 2019. Multimodal transformer for unaligned multimodal language sequences. Proc 57th Annu Meet Assoc Comput Linguist. p. 6558-6569.

Turchin V. 1977. *The Phenomenon of Science: A Cybernetic Approach to Human Evolution*. New York: Columbia University Press.

Van Duijn M, Keijzer F, Franken D. 2006. Principles of minimal cognition: casting cognition as sensorimotor coordination. Adapt Behav. 14(2):157-170.

Veloz T. 2021. Goals as emergent autopoietic processes. Front Bioeng Biotechnol. 9:720806.

Von Frisch K. 1967. *The Dance Language and Orientation of Bees*. Cambridge (MA): Harvard University Press.

Wangchuk D. 2007. *The Resolve to Become a Buddha: A Study of the Bodhicitta Concept in Indo-Tibetan Buddhism*. Tokyo: International Institute for Buddhist Studies.

Waters CM, Bassler BL. 2005. Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol. 21:319-346.

Whitehead AN. 1929. *Process and Reality: An Essay in Cosmology*. New York: Macmillan.

Whitehead H, Rendell L. 2015. *The Cultural Lives of Whales and Dolphins*. Chicago: University of Chicago Press.

Wigner EP. 1960. The unreasonable effectiveness of mathematics in the natural sciences. Commun Pure Appl Math. 13(1):1-14.

Witkowski O, Doctor T, Solomonova E, Duane B, Levin M. 2023. Toward an ethics of autopoietic technology: stress, care, and intelligence. Biosystems. 231:104964.

Wittgenstein L. 1953. *Philosophical Investigations*. Anscombe GEM, translator. Oxford: Blackwell.

Yakherds. 2021. *Knowing Illusion: Bringing a Tibetan Debate into Contemporary Discourse*. 2 vols. New York: Oxford University Press.

Yuan B, Zhang J, Lyu A, Wu J, Wang Z, Yang M, Liu K, Mou M, Cui P. 2024. Emergence and causality in complex systems: a survey of causal emergence and related quantitative studies. Entropy. 26(2):108.

Zuberbühler K. 2018. Combinatorial capacities in primates. Curr Opin Behav Sci. 21:161-169.